Probabilidad y Estadística

Luis Bravo - Sofía Portillo Mongelli

COLEGIO
UNIVERSITARIO
IES

Editorial

Agradecimientos

Agradecemos a todos aquellos que han aportado su investigación, su experiencia y su tiempo para la elaboración de este material de estudio.
Especialmente a José CONFORTE, Mario MARÍN, Raúl MASSA, Carlos MONTEVERDE y Gonzalo MURILLO por haber ofrecido el texto que se utilizó como base para hacer este.
Agradecemos, finalmente, a los directores de carrera del Colegio Universitario IES.

Autores

Luis Bravo

- Ingeniero de Sistemas, IUA.
- Postgrado en Análisis Institucional de las Organizaciones Educativas, U.N.C.
- Postgrado en Sistema Ambiental, su Economía y Protección, IUA.
- Postgrado en Planeamiento Educativo, IESE.
- Doctorado en Educación, UCSF.
- Profesor adjunto, EAM.
- Asesor consultora CITERA.
- Seminario Estadística y Educación, UCSF.
- Jefe del Departamento de Control de Gestión, EAM.
- Jefe RRII y extensión, EAM.

Sofía Portillo Mongelli

- Licenciada en Matemática, egresada de la Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán.
- Premio "Mejor Egresada de la carrera Licenciatura en Matemática, año 2006", otorgado por la Federación Argentina de Mujeres Universitarias (FAMU), sede Tucumán.
- Auxiliar docente de 1° categoría en la materia "Elementos de Álgebra Lineal y Geometría Analítica FaCET, Universidad Nacional de Tucumán, 2002 a 2006.
- Profesora ayudante en la materia "Álgebra I", FaMAF, UNC, 2009 (suplencia de cargo concursado).
- Auxiliar docente, adscripta en la materia "Matemática I", Facultad de Ciencias Económicas, UNC, a partir de Marzo de 2009.
- Profesora ayudante en la materia "Análisis Matemático II", FaMAF, UNC, 2009 (suplencia de cargo concursado).
- Profesora de la materia "Probabilidad y Estadística en el Colegio Universitario IES, 2009.

Equipo de producción

Producción y dirección general
- Directora general: María Braganza
- Vicedirector general: Alberto Rabbat
- Directora Académica: María Fernanda Sin
- Vicedirectora Académica: Erica Bongiovanni

Planificación y coordinación general
- Coordinador de estudios a distancia: Erica Bongiovanni
- Coordinación de proceso editorial: Nicolás Irusta

Producción Editorial
- Sebastian Benito
- Diego Oliva

Producción Académica
- Ana Giró
- Telmo Torres

Coordinación de sistemas
- Marcela Giraudo

Uso de marcas

Cláusulas de uso de marcas y derechos de autor de terceros.

- Uso atípico de marca ajena: Exclusión de los usos no comerciales del Derecho Marcario. PERSPECTIVAS S.A. en su carácter de titular de los derechos intelectuales sobre la presente obra declara por esta vía que el uso que realiza de marcas comerciales de terceros lo es sólo a los fines informativos y didácticos, para mejor comprensión de los lectores y alumnos del contenido de la obra, siendo el mismo de carácter atípico (uso atípico de marca ajena) y lícito. Este uso, a tenor de la jurisprudencia vigente queda fuera del ius prohibendi, que detenta el titular de cada marca registrada, atento no ser el mismo de carácter comercial en relación al producto que distinguen las referidas marcas, y por ende de índole marcario.-
- Uso de derechos de autor en videos, diskettes, imágenes y audio: Libre utilización -Uso privado- de obras protegidas. PERSPECTIVAS S.A. en su carácter de titular de los derechos intelectuales sobre la presente obra declara por esta vía que el uso que realiza de determinadas grabaciones (audio y video), e imágenes (fotografías) de terceros, lo es a los fines informativos y didácticos, para mejor comprensión de los alumnos del contenido de la obra, siendo el mismo de carácter privado y no comercial, y desde ya respetando el derecho de cita, esto es declarando en toda ocasión la cita o fuente (obra y autor) de la cual se toman los fragmentos de obras de terceros para incorporarlos a la presente (Convenio de Berna, Acta de París, 1971 – Art. 10, § 2 y § 3).-
- Modificación de obras literarias por el titular de los derechos patrimoniales. PERSPECTIVAS S.A. en su carácter de titular de los derechos intelectuales (patrimoniales) sobre la presente obra, aclara que ha autorizado a Luis Bravo y a María Sofía Portillo Mongelli, a la modificación respecto de la obra original publicada con Nº de ISBN 978-987-600-093-2, cuyos autores son José Conforte, Mario Marín, Raúl Massa, Carlos Monteverde y Gonzalo Murillo, constituyéndose en autores morales de la obra referida y modificada. Se declara a todo efecto que, los derechos intelectuales se ceden y mantienen a favor de su titular, PERSPECTIVAS S.A.

Copyright

Bravo, Luis ; Portillo Mognelli, Sofía
Probabilidad y estadística / Bravo, Luis ; Portillo Mongelli, Sofía ; coordinado por María Teresa de las Casas; dirigido por José Alberto Rabbat. - 1a ed. - Córdoba : IES Siglo 21, 2010.

e-book

ISBN 978-987-600-139-7

1. Estadísticas. 2. Matemáticas. I. Bravo, Luis ; Portillo Mongelli, Sofía. II. De las Casas, Maria Teresa, coord. III. Rabbat, José Alberto, dir.
CDD 310.4

1ra Edición

Buenos Aires 563
TE: 54-351-4211717
5000 - Córdoba

Queda hecho el depósito que establece la Ley 11723

Libro de edición argentina

Se terminó de replicar durante el mes de Agosto de 2010 en el departamento de Logística en Editorial IES Siglo 21.

Índice

Cómo está organizado este texto

Usted está en presencia de este texto que los autores proponen para la comprensión y estudio de la asignatura. Ha sido preparado y diseñado para facilitar el acceso al conocimiento, a partir de una secuencia cuyo punto de partida es la práctica profesional cotidiana y no la teoría alejada de la realidad.

Está organizado de la siguiente manera:

- **Introducción**. Indica qué papel desempeña la asignatura dentro de la carrera y los conceptos básicos que usted conocerá.
- **Esquema**. Muestra los enlaces que unen los conceptos centrales de la asignatura entre sí.
- **Situación profesional.** Lo ubica frente a un problema de la práctica profesional cotidiana que puede ser resuelto, ya que existe al menos una solución para ello, a través de conocimientos específicos que en cada caso se aportan.
- **Herramientas**. Son los conocimientos necesarios para resolver la Situación profesional planteada.
- **Autoevaluación**. Para que usted compruebe si ha comprendido correctamente lo que se explicó en una herramienta, los autores proponen la resolución de actividades y le ofrecen las respuestas.
- **Ejercicio resuelto.** Bajo este título encontrará una manera de resolver los problemas de práctica profesional planteados, con la selección de las herramientas pertinentes.
- **Ejercicio por resolver.** Ahora le toca a usted. Es el momento de aplicar las herramientas a una situación profesional nueva o similar a la ya expuesta. Todas las dudas que le aparezcan podrán ser planteadas a su docente.
- **Evaluación de paso.** Para que usted compruebe si ha comprendido correctamente lo que se explicó en las distintas herramientas que hasta el momento se han presentado, los autores proponen la resolución de actividades y le ofrecen las respuestas.
- **Bibliografía.** Se indican los textos, revistas y links de consulta a los que podrá recurrir para complementar o ampliar algunos temas.

Introducción

En la actualidad, es común observar que en cada aspecto de la vida diaria, se realizan estudios para analizar, comprender, evaluar y predecir comportamientos. Y esto puede verse en cualquier sector de nuestra sociedad: la economía, la política, los recursos humanos, las relaciones públicas, como también en las ciencias sociales y naturales. El nombre de Estadística alude al enorme interés de esta rama matemática para los asuntos del Estado, y su introducción en el mundo científico se debe a la importancia indiscutible para el desarrollo de las ciencias sociales y humanas.

Este texto está dividido en 3 partes:

- Estadística Descriptiva
- Probabilidad
- Muestreo e Inferencia Estadística

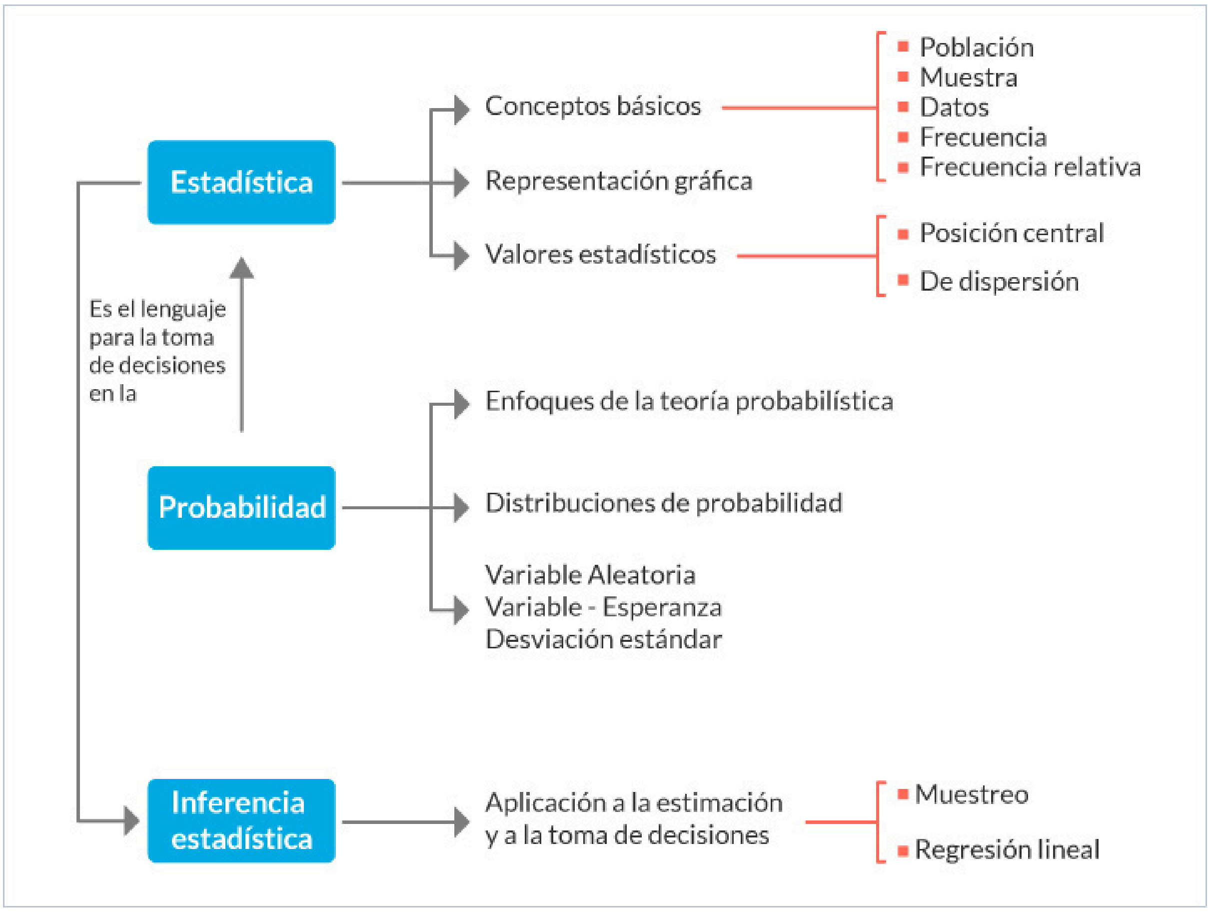
La tarea de describir y procesar de modo adecuado la masa de datos provenientes de las observaciones y experimentos, es el objeto de la estadística descriptiva, que es la parte de la estadística que conocemos desde los cursos de educación primaria donde se introducen nociones básicas, y que, por lo general, no pasa a ser un análisis más profundo de la información. Es un primer acercamiento y, por esa misma razón, es la manera de presentar ante cualquier lector, ya sea especialista o no, los datos en estudio. Esta parte se dedica única y exclusivamente al ordenamiento y tratamiento mecánico de la información para su presentación por medio de tablas y de representaciones gráficas, así como de la obtención de algunos parámetros útiles para la explicación de la situación involucrada.

El análisis de los datos se realiza mediante la teoría de la probabilidad, que es el lenguaje que permite expresar los fenómenos estadísticos, y que pueden variar en su descripción según el enfoque teórico que se adopte; a partir de este lenguaje de probabilidad se desarrolla toda la teoría estadística.

Finalmente, el arte de obtener con confianza conclusiones sobre el modo de proceder respecto del fenómeno que se estudia es el objeto de las diversas técnicas existentes de inferencia estadística.

Los autores

Esquema



Estadística
Probabilidad
Inferencia estadística
Es el lenguaje para la toma de decisiones en la
Conceptos básicos
Población
Muestra
Datos
Frecuencia
Frecuencia relativa
Representación gráfica
Valores estadísticos
Posición central
De dispersión
Enfoques de la teoría probabilística
Distribuciones de probabilidad
Variable Aleatoria
Variable - Esperanza
Desviación estándar
Aplicación a la estimación y a la toma de decisiones
Muestreo
Regresión lineal

Situación profesional 1: Estadística Descriptiva. Valores de Posición Central

Link del video: https://www.youtube.com/watch?v=4ewbPdeIka0

Texto del video: Usted se desempeña en la oficina de estrategias de comercialización de la importante cadena de supermercados “LA PAZ”, y el Jefe de Departamento le encomienda al equipo de trabajo que usted integra, que realice una clasificación de cinco categorías y una distribución de frecuencias relativas que se correspondan con las especificaciones mencionadas anteriormente, referido a las edades de los compradores de MP4, durante la semana pasada.

A la mañana siguiente se reúnen para organizar el trabajo y le solicitan al Departamento de Sistemas de Computación el registro de las edades de los compradores de MP4 de la semana anterior, quienes se lo entregan de manera inmediata por Intranet:

A continuación, usted con el equipo de trabajo, deberá analizar si esta clasificación le permite llegar a alguna conclusión sobre el mercado y la comercialización; además si es posible con la distribución de frecuencias aportar información adicional.

Ver tabla

25	19	24	36	31	28	39	43	17	40
49	61	46	44	54	16	33	45	20	26
34	67	57	33	62	28	29	24	43	36

SP1/H1: Estadística

La estadística comprende un conjunto de métodos y técnicas que permiten, a través del análisis de observaciones, estudios y las determinaciones de los valores estadísticos de una muestra debidamente representativa de una población, inferir sobre los parámetros de la misma con un cierto grado de bondad.

Ahora profundizaremos, analizando esta definición, con el objetivo de entender cada uno de los términos que utilizaremos:

1.1 Población
1.2 Muestra
1.3 Muestra debidamente representativa
1.4 Valores estadísticos
1.5 Parámetros
1.6 Bondad

1.1 Población

Constituye el todo, también puede ser llamada universo; es el conjunto de componentes o datos cuantificables pertenecientes a todo un sistema en estudio, comprende la población de referencia sobre la cual se realizan las observaciones.

Link del video: https://www.youtube.com/watch?v=HlFRS746p3o

Texto del video: Por ejemplo: si tomamos como sistema de estudio la edad promedio de los estudiantes del Colegio Universitario IES Siglo 21, la población estará formada por todos los estudiantes de esa institución educativa.

1.2 Muestra

La muestra está constituida por un subconjunto de la población.

Cada uno de los elementos que forman parte de la muestra se denomina observaciones, de esta manera también podemos definir a una muestra como: "Un conjunto de observaciones tomadas de una población determinada o en estudio".

Link del video: https://www.youtube.com/watch?v=yOu7g8MJ50Y

Texto del video: Para una mayor comprensión y volviendo al ejemplo anterior, podemos tomar una parte de esos alumnos del Colegio Universitario IES Siglo 21 como muestra representativa, ya que responden a un subconjunto de la población.

Diferencia entre Censo y Muestra

En el caso del estudio de las características de la población, a través del análisis de cada uno de los elementos que la caracterizan, se dice que se realiza un censo. Este tipo de estudio, para el caso de poblaciones grandes, se torna lento y costoso. De esta manera para facilitar el trabajo planteado conviene efectuar un estudio sobre una muestra de esa población.

1.3 Muestra debidamente representativa

Una muestra debidamente representativa debe poseer las características necesarias que permitan inferir sobre el comportamiento de la población, con un considerado margen de seguridad.

Pero profundicemos la definición con un ejemplo:

Link del video: https://www.youtube.com/watch?v=LQqNQgRpDBM

Texto del video: Cómo haríamos, si tuviésemos la intención de realizar un estudio estadístico en el Barrio Avenida de la ciudad de Córdoba Capital, referido a los jefes de familia que perciben un sueldo de $5000 y $10.000 y que han finalizado la escuela secundaria.

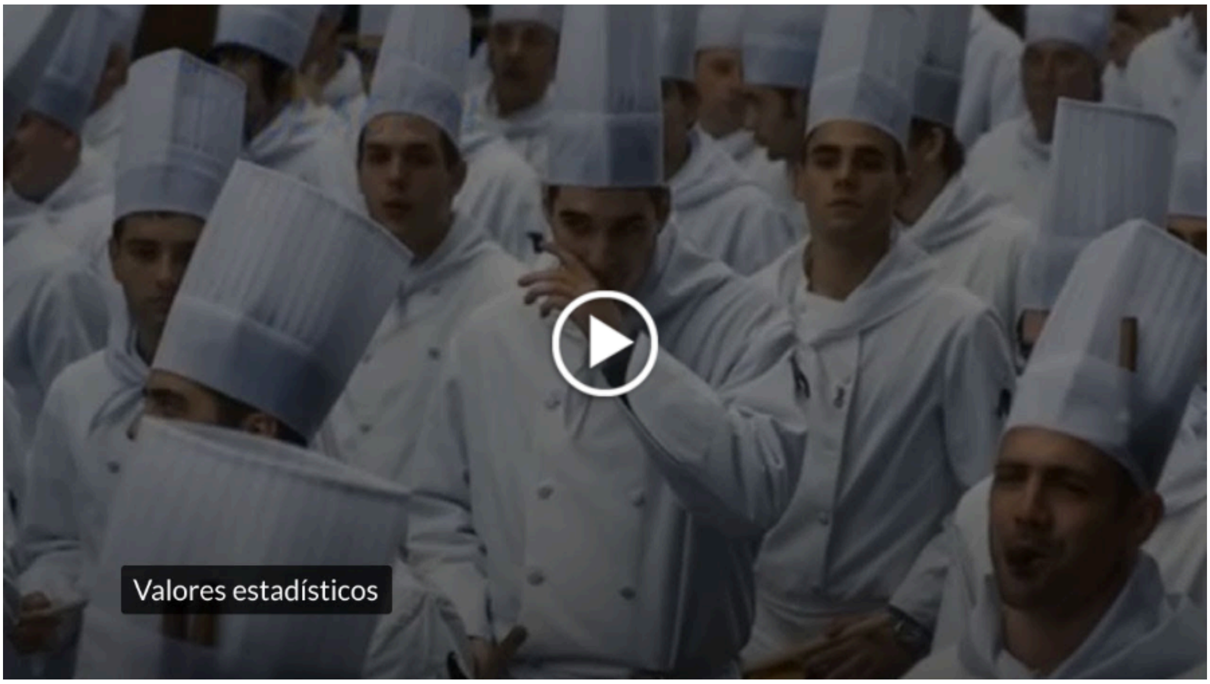
Deberíamos incorporar tanto a mujeres como a hombres en las proporciones que corresponden, de tal manera que la muestra contenga similares características de la población que se desea analizar.

1.4 Valores estadísticos

El estudio realizado sobre una muestra nos permite determinar valores a los cuales se los denomina estadísticos, esos mismos valores, pero pertenecientes a la población, toman el nombre de parámetros.

En el estudio de una población la media o promedio lleva el nombre de parámetro, mientras que en la muestra se denomina estadístico.

Entonces, podríamos decir a modo de ejemplo que:



Link del video: https://www.youtube.com/watch?v=KUdt0CX6qWk

Texto del video: El sueldo promedio de la población de los empleados gastronómicos de la ciudad de Córdoba se denomina Parámetro, mientras que el sueldo promedio de la muestra de los empleados gastronómicos se denomina Estadístico.

1.5 Parámetros

Parámetros son aquellos valores en estudio de una población, que en la muestra toman el nombre de Estadísticos.

1.6 Bondad

Se define como bondad al margen de seguridad con que se realiza la inferencia de acuerdo con los estudios efectuados sobre la muestra.

Indique la opción correcta

1- En un establecimiento dedicado a la cría de gallinas, se quiere comprobar la eficacia de un nuevo alimento balanceado. En el establecimiento tienen 200 gallinas y se las pesa antes y después de los 30 días que dura esta dieta con el nuevo alimento. Esto es _______.

- ☐ población
- ☐ muestra

Indique la opción correcta

2- Un fabricante de bulones desea un control de calidad de fabricación, para ello, selecciona 150 bulones del total de su producción, perteneciente a distintos lotes y se le mide su diámetro. Esto es _____.

- ☐ población
- ☐ muestra

Indique la opción correcta

3- En el club social "NUEVA VISTA" de 1500 socios, se registran las edades de todos ellos, con la finalidad de organizar nuevos grupos de gimnasia para diversos grupos etarios. Esto es ______.

- ☐ población
- ☐ muestra

Indique la opción correcta

4- La empresa de iluminación "LUCES" de la ciudad de Córdoba está interesada en realizar un estudio de mercado para determinar los diferentes tipos de artefactos de iluminación que se encuentran en el frente de los hogares de Córdoba. Para ello, se instalan puestos de observación en los barrios que componen la ciudad. La observación se efectúa sobre 1000 hogares. Esto es __________

- ☐ población
- ☐ muestra

Indique la opción correcta

5- Una muestra debidamente representativa debe poseer las características necesarias que permitan ________ sobre el comportamiento de la población, con un apreciable margen de seguridad.

- ☐ inferir datos
- ☐ resolver
- ☐ determinar

Indique la opción correcta

6- La media o promedio en la muestra se llama estadístico, en la población toma el nombre de _______.

- ☐ estadístico
- ☐ valor
- ☐ parámetro

Indique la opción correcta

7- El análisis de cada uno de los elementos de una _________ corresponden a un Censo.

- ☐ muestra
- ☐ población
- ☐ bondad

Indique la opción correcta

8- Se define como __________ al margen de seguridad con que se realiza la inferencia de acuerdo con los estudios realizados sobre la muestra.

- ☐ parámetro
- ☐ valor estadístico
- ☐ bondad

Respuestas correctas[1]

---

[1] 1) población. 2) muestra. 3) población. 4) muestra. 5) inferir datos. 6) parámetro. 7) población. 8) bondad.

Los estudios estadísticos se basan en datos correspondientes a la variable en estudio, y cada uno de los datos se denomina observación.

Como ejemplos de variables podríamos tomar:

- Las temperaturas registradas cada una hora en un laboratorio, durante dos semanas.
- El número de materias aprobadas con 7 puntos por alumnos del Colegio Universitario IES durante el período 2010-2012, etc.

2.1 Tipos de datos

En el estudio de las ventas de calculadoras, la cantidad de unidades vendidas es un número entero finito, no es así en el caso de los promedios de calificaciones de los alumnos del último curso. Es por ello que, el tipo de dato de cada una de las incógnitas en estudio puede tener características diferentes.

¿Qué podemos concluir?

Definida la variable, su contenido o valor puede ser de distinto tipo y, por lo tanto, es necesario clasificarla:

Datos discretos

Los datos discretos, definidos generalmente por valores enteros, son el resultado de contar un número de conceptos u objetos. Por ejemplo:

- Número de computadoras en un hogar (1, 2 o 3).
- Cantidad de viviendas en un barrio (200, 201, etc.)

Datos continuos

Las variables del tipo continuo pueden asumir cualquier valor dentro de un intervalo de números reales. Es decir, cuando el dato puede asumir virtualmente cualquier valor dentro de un determinado intervalo, se dice que es del tipo continuo. Así entonces, podemos decir como ejemplos que son variables continuas: 1° El período de tiempo desde el acopio y hasta que se produce la disminución del 25% de la fruta guardada. 2° Los milímetros de lluvia caída en un pueblo cualquiera durante un año.

Datos nominales

Sus categorías son nóminas o etiquetas y no se puede establecer ninguna secuencia lógica de ordenamiento. En este caso, medir representa simplemente la asignación de un atributo a una unidad de análisis.

Por ejemplo podríamos decir:

- Juan es panadero.
- Rodolfo es rubio, etc.
- Código: 1; categoría: cadetes; cantidad: 25.

Datos jerarquizados

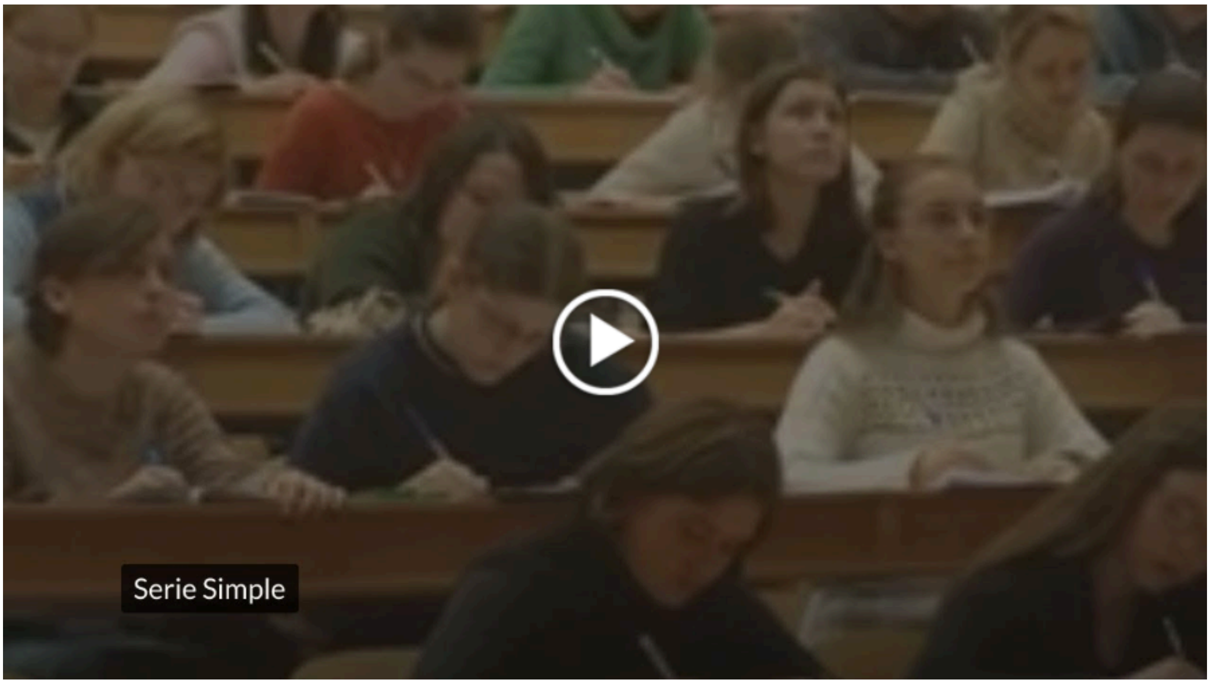
Las variables de datos jerarquizados están constituidas por valores relativos asignados para denotar un cierto orden (primero, segundo, tercero, etc.).

En el caso de tener que estudiar el grado de calidad de las pinturas expuestas por pintores importantes de la ciudad de Córdoba para premiarlos, es necesario acudir al juicio de expertos que permita definir el grado de categoría de los artistas. Esto implica que, en el proceso de jerarquizar las pinturas, influirá el criterio de quien realiza el estudio, ya que de ser otro el que efectúe el análisis podría modificar dicho orden a su criterio.

2.2 Serie simple

Definiremos como serie simple al conjunto de observaciones de una muestra ordenadas de menor a mayor.

Veamos cómo podemos ejemplificarlo.



Link del video: https://www.youtube.com/watch?v=kaLAIRUEelA

Texto del video: Obtenidas las siguientes calificaciones por un curso:

10, 4, 8, 4, 6, 9, 5, 8, 5, 7.

Si quisiéramos descubrir rápidamente los valores mínimos y máximos en los datos o darnos cuenta de los valores que aparecen más de una vez, esto se hace dificultoso.

Para ello es más conveniente presentar los datos ordenados de la siguiente manera:

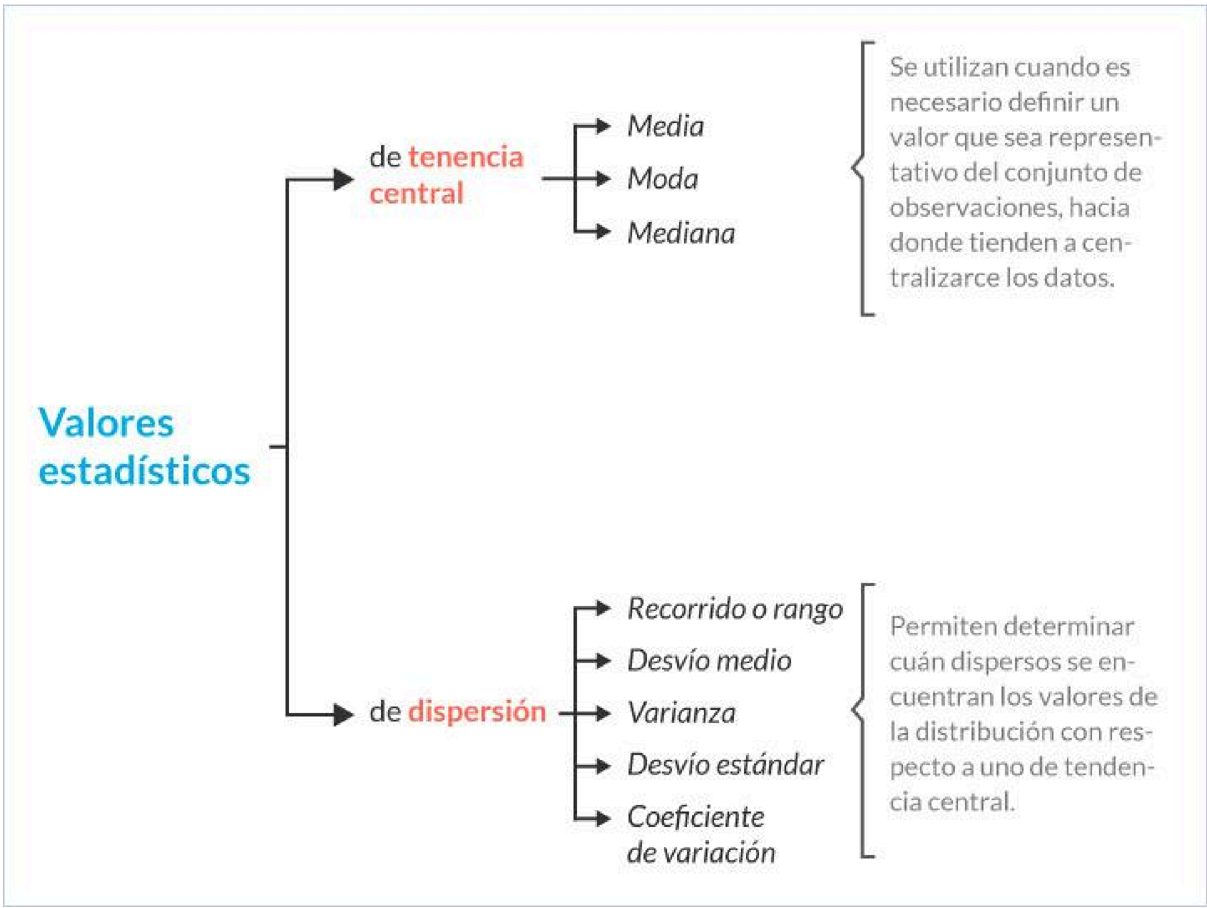
4, 4, 5, 5, 6, 7, 8, 8, 9, 10

Tenemos en este caso una distribución ordenada llamada serie simple. Una vez generada la serie simple, estamos en condiciones de efectuar el análisis y estudio correspondiente a esa distribución, determinando lo que hemos definido como valores estadísticos.

2.2.1 Valores estadísticos

En Estadística surge la necesidad de inferir, sobre las características de la población, a través del análisis y estudio de una muestra. Por lo tanto, debemos definir cuáles son los valores estadísticos que se hace necesario determinar y estudiar en una muestra. Dichos valores se pueden clasificar de la siguiente manera:

Deslice el cursor sobre el texto para ampliar la información.



2.3 Valores de tendencia central: Media

La Central Hidroeléctrica cuenta con 10 generadores que han sufrido fallas por mantenimiento de acuerdo a los siguientes registros:

Nos interesa saber el promedio de días que estuvo fuera de servicio, para ello sumamos los valores y los dividimos después entre el número de observaciones:

$$\bar{X} = \frac{4+8+12+6+12+8+2+4+22+7}{10} = \frac{85}{10} = 8{,}5 \text{ días}$$

Sabemos entonces que la definición de la media aritmética es: un indicador dado por el promedio aritmético de los datos.

$$\bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + \dots\dots\dots\dots + X_{n-2} + X_{n-1} + X_n}{n}$$

n = número de observaciones

Teniendo en cuenta el concepto de sumatoria, la media se expresa como:

Valores Estadísticos

Valores Estadísticos

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

En caso de una muestra

$$\mu = \frac{\sum_{i=1}^{n} X_i}{N}$$

si se trata de una población

Recuerde que:

dada una serie de valores $X_1, X_2, X_3, \dots, X_i, \dots, X_n$ la suma de todos los valores desde X_1 hasta X_n, se puede expresar como:

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + X_4 + \dots\dots\dots\dots + X_{n-2} + X_{n-1} + X_n$$

Tenga en cuenta que la media es un valor que podríamos decir representativo de aquellos obtenidos, pero a veces no es la mejor para una buena toma de decisiones.

Es como si todos los datos obtenidos en las mediciones tuviesen un valor similar, igual a la media.

Por ejemplo: si tomamos la edad promedio de un grupo de alumnos en un aula universitaria, y su media es 22,8 años, podríamos entender que todos los participantes en el curso tienen esa edad.

Obviamente que esto no es así, no obstante, para los efectos del estudio, para la toma de decisiones, funcionaría muy bien porque se puede comparar con un grupo similar que anteriormente recibió este mismo curso y con edad promedio de 21,3 años. El mismo, es un grupo más joven y eso debe ser tenido en cuenta, para dar una perspectiva diferente a este nuevo curso.

Recapitulemos lo expuesto:

	Media	Nº de elementos	Denominación
Población	μ	N	Parámetro
Muestra	$\bar{x}$	n	Estadístico

2.4 Mediana

En una división se toman 7 alumnos para competir en una carrera. Los tiempos de cada uno de los corredores son:

Corredor	1	2	3	4	5	6	7
Tiempo	4,2	4,3	4,7	4,8	5,0	5,1	9

Para dicha distribución, la mediana está dada por el valor de la misma que ocupa su punto medio, en este caso es 4,8.

Se define, entonces, como mediana de una distribución, al valor que deja a su izquierda y a su derecha el mismo número de elementos, es decir, de todos los valores de la

distribución o serie simple[2] la mitad de ellos son menores o iguales a la mediana y la otra mitad son mayores o iguales a ella.

Si representamos a una distribución de n=9 valores como:

$$X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7 \quad X_8 \quad X_9$$

$$Me = X_5$$

Observe que a la izquierda de x5 se ubican cuatro elementos de la distribución que son, además, menores o eventualmente iguales a él. Igual cantidad se ubican a su derecha.

Cuando el número de elementos es par, el valor de la mediana está dado por el promedio de los dos elementos centrales. Supongamos tener una distribución de n = 10 elementos, si los representamos en forma general:

$$X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7 \quad X_8 \quad X_9 \quad X_{10}$$

Se tendría en este caso que los elementos centrales son X_5 y X_6:

$$Me = \frac{X_5 + X_6}{2}$$

Entonces debemos tener en cuenta que la expresión de la mediana varía según el número de elementos de la distribución fuere par o impar.

En el caso de ser *n impar*, entonces $Me = X_i$ el subíndice i toma el valor de:

$$i = \frac{n+1}{2}$$

[2] Distribución o Serie Simple: Es el conjunto de observaciones ordenados de menor a mayor.

Pero en el caso de ser n par, la mediana surge como el promedio de los dos elementos centrales, si n = 10, observamos que los elementos centrales eran X_5 y X_6, es decir, el primer subíndice surge de dividir n/2 y el otro es el mismo subíndice, pero incrementado en una unidad i+1, con lo cual si:

$$i = \frac{n}{2} \qquad \text{entonces} \qquad i + 1 = \frac{n}{2} + 1$$

Podemos ahora definir el valor de la Me para cada uno de los casos.

Me (Mediana)

Si n es impar	***Si n es par***
$Me = X_{(n+1)/2}$	$Me = \dfrac{X_{(n/2)} + X_{(n/2+1)}}{2}$

¿Cuándo es beneficioso utilizar la mediana como referente de una distribución?

La mediana, contrario al caso de la media es insensible a los valores extremos; por lo tanto, cuando la media se vea distorsionada por los extremos, es conveniente la utilización, como elemento de referencia de la distribución a la mediana.

En resumen, los pasos para la determinación de la mediana son:

1. Establecer si el número de elementos es par o impar.
2. Generar la serie simple.
3. Si n es impar, la mediana es un valor perteneciente a la distribución.
4. Si n es par, la Me está dada por el promedio de los dos elementos centrales y, por lo tanto, bien puede no pertenecer a la distribución.

2.5 Moda

Si tenemos registrado la cantidad de viajes que hizo el repartidor de la leche por DÍA en un período de 20 días, podemos observar que el valor 15 (viajes) se repite 3 veces, lo que nos indica que 15 es el número más frecuente de viajes.

12	8	7	7	6
4	4	5	5	6
2	2	1	1	0
0	19	15	15	15

Podemos definir como Moda de una distribución al valor que más veces se repite.
La Moda como valor de posición central es un referente importante de la distribución, sobre todo en aquellas donde la variable en estudio es del tipo nominal.

Si una distribución posee más de un valor con el mismo máximo de repeticiones, cada uno de ellos se constituirá en una nueva moda, es decir que una distribución puede tener más de una moda. En caso de tener dos modas se la denomina bimodal y en caso de tres, trimodal, etc. Cuando en una distribución todos los valores tienen el mismo número de repeticiones, diremos que dicha distribución no tiene moda.

Es necesario resaltar que este valor, de tendencia central, es un indicador de la distribución, realmente significativo en distribuciones de variable de tipo nominal.

2.6 Serie de frecuencias

La Cooperativa Agrícola "Pucará" estudia la posibilidad de posicionar adecuadamente la actividad que ellos realizan en el mercado, para lo cual han decidido destinar los campos de la siguiente manera: 70 % para siembra de soja y el 30 % para engorde de ganado. Se propone realizar un estudio previo sobre 100 vacas, y para llevarlo a cabo, consideran que sería conveniente seleccionar 30 animales y pesarlos en este momento y luego a los 30 días, además sugieren que los animales elegidos al "azar" sean uno de cada 3 vacunos para luego ser numerados y registrados.

A los 30 días, los peones pesan nuevamente a cada vacuno de los seleccionados y el equipo registra la ganancia de peso de acuerdo a como fueron numerados, obteniendo los siguientes resultados:

Vacuno	Peso	Vacuno	Peso	Vacuno	Peso
1	16	11	15	21	16
2	15	12	16	22	12
3	14	13	15	23	15
4	12	14	14	24	13
5	16	15	15	25	13
6	16	16	14	26	11
7	14	17	17	27	15
8	15	18	17	28	17
9	14	19	13	29	18
10	17	20	15	30	13

Teniendo en cuenta el concepto de frecuencia, número de repeticiones, podemos simplificar la tabla presentada y, en lugar de sumar todos los valores de engorde, sumamos cada valor de engorde multiplicado por el número de veces que se repite y, a esa suma, la dividimos por el número total de repeticiones, es decir:

$$\bar{X} = \frac{\sum_{i=1}^{8} X_i \,.\, fi}{\sum_{i=1}^{8} fi}$$

A cada uno de los ocho valores considerados se los denomina clase y se los representa gráficamente con x, y el número de veces que cada uno de ellos se repite, frecuencia, y se la representa también genéricamente con fi.

Con estos nuevos conceptos podemos organizar una nueva tabla, denominada serie de frecuencias, con cuatro columnas:

Número de clase	Clase Xi	Frecuencia fi	Xi . fi
1	11	1	11
2	12	2	24
3	13	4	52
4	14	5	70
5	15	8	120
6	16	5	80
7	17	4	68
8	18	1	18
Totales		$\sum_{i=1}^{8} fi = n = 30$	$\sum_{i=1}^{8} Xi \cdot fi = 443$

El valor de la media se determina como el cociente entre la suma de la última columna y la suma de la anteúltima.

$$\overline{X} = \frac{\sum_{i=1}^{8} X_i \cdot f_i}{\sum_{i=1}^{8} f_i} = \frac{443}{30} = 14{,}76$$

2.7 Frecuencia relativa (FR)

Se define como frecuencia relativa de un determinado valor o clase, fri, a la relación entre su frecuencia y la suma de todas las frecuencias (la suma de todas las frecuencias es igual al número de elementos de la distribución).

Por el momento contamos con tres valores de tendencia central, la media, la mediana y moda, de las cuales en realidad se hace uso sólo de la primera, demostrando que el valor que la misma arroja para su distribución satisface las condiciones de exigencia.

Sin embargo, ahora podemos manejar mayor información, como por ejemplo, averiguar qué preeminencia tienen en la distribución cada una de las clases, es decir qué porcentaje de los vacunos aumentó 12 Kg de peso, 11 Kg, etc.

Por lo tanto, surge un nuevo concepto para tener en cuenta: Frecuencia Relativa.

$$f_{ri} = \frac{f_i}{\sum f_i}$$

Como Σ fi = n entonces la frecuencia relativa también se puede expresar como:

$$f_{ri} = \frac{f_i}{n}$$

Propiedad:

La suma de todas las frecuencias relativas correspondientes a los valores clases de una distribución es igual a 1.

Frecuencia acumulada

Continuamos con la tabla de frecuencia de clase, correspondiente a la ganancia de peso de la muestra de 30 vacas, a fin de definir un nuevo tipo de frecuencia.

Para cada una de las clases podemos definir una frecuencia acumulada (fa), cuyo valor estará dado por la suma de su frecuencia más las frecuencias de las clases que le anteceden.

De esta manera podemos completar la tabla de frecuencias, agregando una columna más que contiene a fa.

Nº de clase	Clase Xi	Frecuencia fi	Xi . fi	F. relativa fri	F. acumulada fai
1	11	1	11	0,03	1
2	12	2	24	0,07	3
3	13	4	52	0,13	7
4	14	5	70	0,17	12
5	15	8	120	0,27	20
6	16	5	80	0,17	25
7	17	4	68	0,13	29
8	18	1	18	0,03	30

¿Qué ventaja nos brinda la frecuencia acumulada?

Analicemos la clase correspondiente a un engorde de 14 kg., su frecuencia acumulada es de 12, lo cual implica que doce vacunos han tenido un engorde de 14 o menos kilogramos. Podemos agregar que sólo 7 (frecuencia acumulada de la clase anterior) de los vacunos tuvieron un engorde menor a los 14 kg.

La columna de frecuencias acumuladas nos permite observar, para una clase, cuántos animales han tenido un aumento de peso menor a dicha clase, evitándonos tener que analizar toda la distribución.

Indique la opción correcta

1- Al conjunto de observaciones de una muestra ordenadas de menor a mayor se la define como _________.

- ☐ serie simple
- ☐ serie nominal
- ☐ serie jerarquizada

Indique la opción correcta

2- Los valores de dispersión permiten determinar qué tan __________ se encuentran los valores de la distribución con respecto a una tendencia central.

- ☐ coherentes
- ☐ dispersos
- ☐ unidos

Indique la opción correcta

3- La/el ________ es una medida de tendencia central

- ☐ desvío estándar
- ☐ coeficiente de variación
- ☐ media

Indique la opción correcta

4- Se define como ___________ de una distribución, al valor que ocupa el punto medio cuando los datos han sido ordenados de menor a mayor.

- ☐ media
- ☐ mediana
- ☐ moda

Indique la opción correcta

5- La suma de todas las frecuencias relativas correspondientes a los valores clases de una distribución es igual a _______.

- ☐ uno
- ☐ cero
- ☐ infinito

Indique la opción correcta

6- Para cada una de las clases podemos definir la __________ cuyo valor estará dado por la suma de su frecuencia más las frecuencias de las clases que le anteceden.

- ☐ frecuencia acumulada
- ☐ frecuencia relativa
- ☐ serie de frecuencias

Indique la opción correcta

7- Los estudios estadísticos se basan en datos correspondientes a la variable en estudio y cada uno de los datos se denomina _________.

- ☐ observación
- ☐ discreto
- ☐ continuo

Indique la opción correcta

8- Las variables del tipo _________ pueden asumir cualquier valor dentro de un intervalo de números reales.

- ☐ continuos
- ☐ jerarquizados
- ☐ nominales

Indique la opción correcta

9- Las variables de datos ___________ están constituidos por valores relativos asignados para denotar un cierto orden.

- ☐ continuas
- ☐ nominales
- ☐ jerarquizadas

Indique la opción correcta

10- Se define como ________ aquellos que se utilizan cuando es necesario definir un valor que sea representativo del conjunto de observaciones, hacia donde tienden a centralizarse los datos.

- ☐ valor de dispersión
- ☐ valor de tendencia central
- ☐ serie simple

Indique la opción correcta

11- Se define como __________ aquellos que permiten determinar cuán dispersos se encuentran los valores de la distribución con respecto a una de tendencia central.

- ☐ valor de dispersión
- ☐ valor de tendencia central
- ☐ serie simple

Indique la opción correcta

12- Los datos ________, definidos generalmente por valores enteros, son el resultado de contar un número de conceptos u objetos.

- ☐ continuos
- ☐ discretos
- ☐ nominales

Respuestas correctas[3]

[3] 1) Serie simple. 2) Dispersos. 3) Media. 4) Mediana. 5) Uno. 6) Frecuencia acumulada. 7) Observación. 8) Continuos. 9) Jerarquizadas. 10) Valor de tendencia central. 11) Valor de dispersión. 12) Discretos.

Tabla 1: Tabla de frecuencias para las edades de compradores de MP4

Variable	Clase	LI	LS	MC	FA	FR	FAA
Edades	1	16,00	26,20	21,10	8	0,27	8
Edades	2	26,20	36,40	31,30	9	0,30	17
Edades	3	36,40	46,60	41,50	7	0,23	24
Edades	4	46,60	56,80	51,70	2	0,07	26
Edades	5	56,80	67,00	61,90	4	0,13	30

Tabla 2: Medidas de resumen para las edades de compradores de MP-4

Variable	n	Media	Mín	Máx	Mediana
Edades	30	36,97	16,00	67,00	35,00

Si bien la edad promedio del comprador de MP4 es de 36,97 años, el 50% de los compradores de MP4 son menores de 35 años. De la Tabla 1 se observa que las personas mayores son menos frecuentes en la compra del producto, y dichas compras se concentran en un mercado más joven.

SP1/Ejercicio por resolver

La empresa "LOS ALERCES" tiene un contrato con el banco "MICRO" referido a créditos blandos, y el resultado de los saldos mensuales del año anterior fueron los siguientes:

Enero	121.200	Abril	72.900	Julio	58.600	Octubre	52.900
Febrero	112.400	Mayo	72.700	Agosto	61.200	Noviembre	49.100
Marzo	72.700	Junio	57.400	Septiembre	50.300	Diciembre	46.200

La empresa puede ser beneficiada con una tasa de interés baja si su saldo mensual promedio es mayor a 65.000.

A usted, como pasante de la empresa, le solicitan:

a. Calcular las medidas de tendencia central.

b. Confeccionar el diagrama de barras para los saldos mensuales y comentar cualquier característica de los datos.

c. En función del criterio del banco, ¿la empresa será beneficiaria del crédito?

SP1/Evaluación de paso

Indique la opción correcta

1- Una muestra está constituida por el conjunto de la población.

☐ Verdadero

☐ Falso

Indique la opción correcta

2- Una muestra debidamente representativa debe poseer las características necesarias que permitan poder inferir sobre el comportamiento de la población.

☐ Verdadero

☐ Falso

Indique la opción correcta

3- Los valores en estudio, que en la muestra toman el nombre de estadísticos, en la población se denominan parámetros.

☐ Verdadero

☐ Falso

Indique la opción correcta

4- Cuando el dato puede asumir cualquier valor dentro de un determinado intervalo, se dice que es del tipo discreto.

☐ Verdadero

☐ Falso

Indique la opción correcta

5- En el caso de una muestra, la media se expresa como:

$$\overline{x} = \frac{\sum_{i=1}^{n} X_i}{n}$$

☐ Verdadero

☐ Falso

Indique la opción correcta

6- Una serie simple es un conjunto de observaciones desordenadas.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

7- La frecuencia relativa puede ser expresada de la siguiente manera:

$$Fri = \frac{fi}{n}$$

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[4]

[4] 1) Falso. 2) Verdadero. 3) Verdadero. 4) Falso. 5) Verdadero. 6) Falso. 7) Verdadero.

Situación profesional 2: Estadística descriptiva

Link del video: https://www.youtube.com/watch?v=I2tX39MuoIs

Texto del video: El laboratorio de medicamentos "MEDICINAL S.A", contrata un equipo de profesionales que usted integra, y que se encargará de evaluar la efectividad de los medicamentos próximos a lanzarse al mercado. Dicho laboratorio, necesita determinar la efectividad de un nuevo producto para dejar de fumar. Para ello selecciona a un grupo de 30 personas que son fumadores, y que usarán este producto a lo largo de un mes, registrando la cantidad de cigarrillos fumados por día, antes y después del uso del medicamento. Los datos relevados son:

Como parte de este equipo de trabajo, se le pide analizar la situación, a fin de determinar la efectividad de este nuevo producto.

Número de personas	Cantidad de cigarrillos antes	Cantidad de cigarrillos después	Número de personas	Cantidad de cigarrillos antes	Cantidad de cigarrillos después
1	19	6	16	22	9
2	20	9	17	21	8
3	23	8	18	23	8
4	21	9	19	21	9
5	22	8	20	22	10
6	24	9	21	20	9
7	21	8	22	21	7
8	20	9	23	21	9
9	21	7	24	22	12
10	20	9	25	22	9
11	20	8	26	21	8
12	24	11	27	20	9
13	22	9	28	23	8
14	19	8	29	21	8
15	25	10	30	22	9

La utilidad de la representación gráfica de los datos es que, al poder graficar una distribución, esto nos permite realizar una lectura e interpretación rápida de las características de la misma, y sacar conclusiones inmediatas a partir de ello.

Además, el representar de manera gráfica los datos, permite resaltar y aclarar los patrones que no son posibles de distinguir fácilmente en las tablas. Permiten también estimar valores con una rápida mirada y proporcionar una verificación visual sobre las precisiones de los cálculos.

Consideremos primero la distribución de frecuencia que se genera con el número de cigarrillos fumados por día, antes del consumo del medicamento. Recuerde que para generar la distribución de frecuencias es necesario clasificar los valores de la distribución en clases, y determinar la frecuencia en cada una de las mismas.

Confeccionaremos la tabla de frecuencia que corresponde a la distribución correspondiente al momento previo del tratamiento con este medicamento, que contiene en este caso, además, la frecuencia relativa de cada clase:

Tabla 2.1

Número de clase	Cantidad de cigarrillos	fi	xi.fi	fri
1	19	2	38	0,066667
2	20	6	120	0,2
3	21	9	189	0,3
4	22	7	154	0,233333
5	23	3	69	0,1
6	24	2	48	0,066667
7	25	1	25	0,033333
		30		1

Recuerde siempre que:
a) La suma de las frecuencias es igual al número de observaciones.
b) La suma de las frecuencias relativas es igual a 1.

Ahora usted podrá preguntarse, una vez dada la distribución, ¿qué debemos graficar?

Debe, antes que nada, tener en cuenta que se habla de graficar, es decir, se tiene como objetivo plasmar de manera gráfica la situación en estudio, dada por una determinada relación.

En este caso, es importante poder representar el número de cigarrillos y su frecuencia, referidos al momento previo del uso del medicamento (tabla 2-1), a fin de poder comparar con la misma relación, pero referida al período posterior del uso del medicamento.

La segunda pregunta que podría plantearse es ¿con qué tipo de gráfico representaremos la relación? Esta herramienta del texto responderá a este interrogante, presentándole distintas alternativas:

a. Diagramas de sectores
b. Diagrama de bastones
c. Histogramas de frecuencias y frecuencias relativas
d. Polígonos de frecuencias y frecuencias relativas.

a. Diagramas de sectores

Es común encontrar en un periódico o una revista, o en un informativo de la televisión, un diagrama de sectores.

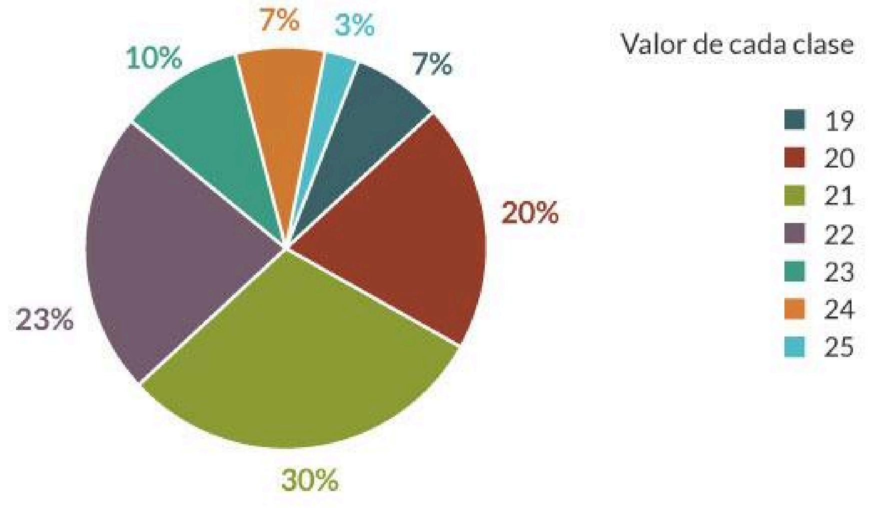
Un diagrama de sectores es un gráfico circular dividido en tantos sectores como clases se tenga, y a cada clase le corresponde un tamaño proporcional a la frecuencia (o frecuencia relativa).

¿Cómo determinar la amplitud de un sector? Notemos que el total de clases corresponde a una amplitud de 360 grados, por tanto, con un razonamiento sencillo basado en la regla de tres simple, observamos que:

Dada una clase de frecuencia fi, y siendo n el total de observaciones, n corresponde a 360 grados. Luego, si x es la amplitud de dicha clase, x=360 grados.fi/n=360 grados.fri

A continuación, presentamos el gráfico que corresponde a la distribución del número de cigarrillos fumados antes del tratamiento (correspondiente a los datos de la tabla 2-1).

Gráfico 2.2



b. Diagrama de bastones

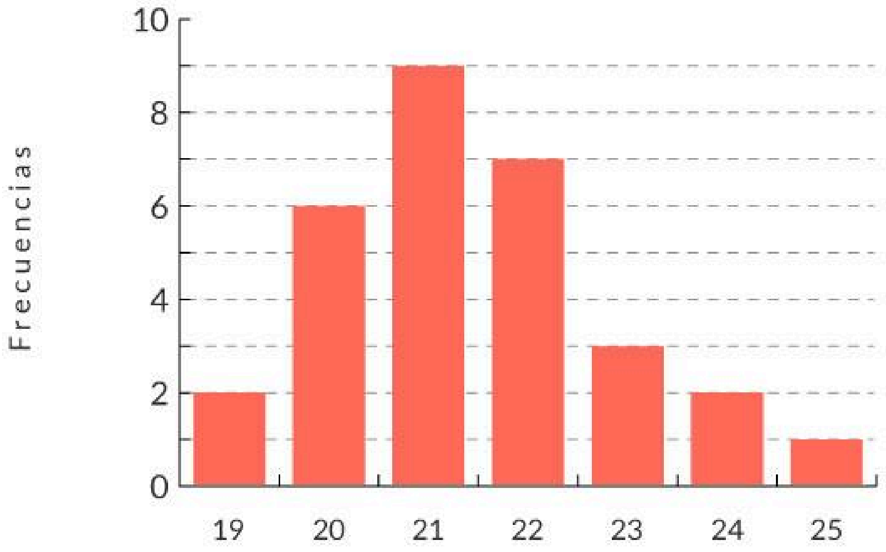
Los diagramas que veremos a continuación, se efectúan en un sistema de coordenadas de ejes ortogonales x-y.

¿Cuál es la manera de construirlo? Sobre el eje de las abscisas (el eje de las x) representaremos las clases, en este caso el número de cigarrillos fumados por día, y en el eje de las ordenadas (el eje de las y) indicaremos su frecuencia.

Si levantamos para cada valor de clase un segmento o "bastón" que represente la frecuencia, se obtiene el diagrama de bastones.

A continuación se muestra el diagrama de bastones para la distribución antes analizada.

Gráfico 2.3



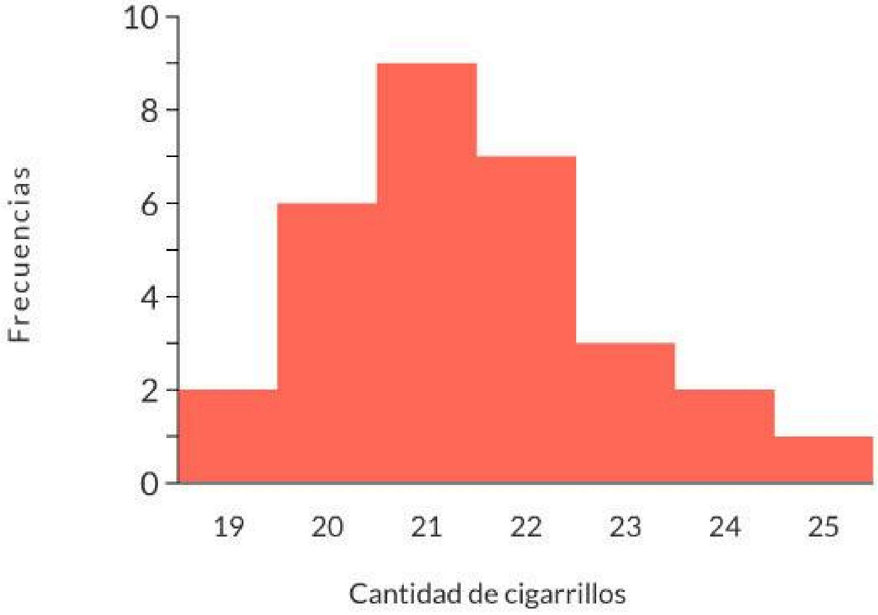
c. Histogramas de frecuencias y frecuencias relativas

Según vimos en el diagrama de bastones, en el eje de las x ubicamos las clases. La diferencia entre dos clases consecutivas se denomina intervalo de clase y lo representamos con Δx; en el caso que estamos analizando, la diferencia entre 2 clases consecutivas es la unidad, es decir que $\Delta x=1$.

Un histograma se define como un diagrama de barras sin discontinuidades; de esta manera podemos así operar con superficies. Para su construcción recurrimos a un procedimiento análogo al del diagrama de bastones; si llevamos a izquierda y derecha de cada bastón la mitad de la longitud del intervalo (es decir, $\Delta x/2$), y levantamos segmentos normales al eje x e iguales en longitud a cada bastón, se tendrá como resultado un histograma de frecuencias.

El gráfico 2-4 que se presenta a continuación, también corresponde a los datos de la tabla 2-1.

Gráfico 2.4



Observe que la superficie encerrada por cada una de las barras, corresponde a la de un rectángulo, y la misma se calcula como el producto de la base y la altura. Luego:

$$S1 = \Delta x.f$$

que correspondería al valor de la frecuencia de la primera clase.

Como ocurrirá lo mismo para cada barra, podemos concluir que:

"El área encerrada por el Histograma de frecuencias es igual a la suma de todas las frecuencias y, por lo tanto, igual al número de observaciones".

Recuerde que: la suma de las frecuencias es igual al número de observaciones.

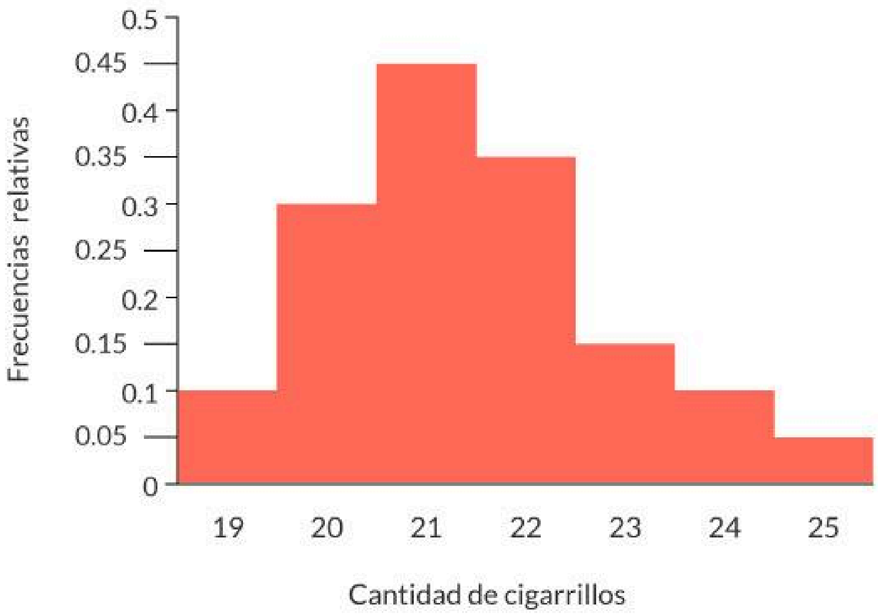
$$\sum f_i = n$$

Histogramas de frecuencias relativas

El procedimiento para construir este gráfico, es exactamente igual al del histograma de frecuencias, con la diferencia de que en el eje de las ordenadas, colocaremos las frecuencias relativas.

Así, para los datos que corresponden a la tabla 2-1, el histograma de frecuencias relativas resulta:

Gráfico 2.5



Al igual que en el histograma de frecuencias, la superficie encerrada por una barra corresponde al área de un rectángulo y se calcula:

$$S = \Delta x . f_r \text{ donde, al considerar } \Delta x = 1,$$

$$\text{tendremos que } S = f_r$$

Por lo tanto:

El área encerrada bajo el histograma de frecuencias relativas es:

$$\sum f_{ri} = 1$$

Este último concepto nos permite independizarnos del tipo de distribución. Para cualesquiera de ellas, independientemente de sus frecuencias y del número de observaciones, la superficie encerrada por el histograma de frecuencias relativas es 1.

d. Polígonos de frecuencias y frecuencias relativas.

Cuando se necesita hacer comparaciones entre distintas distribuciones, en general es conveniente realizarlo a partir de polígonos de frecuencias o frecuencias relativas.

¿Cómo se construyen los polígonos de frecuencias o frecuencias relativas?

Consideramos los ejes como en los casos anteriores, y en el eje de las ordenadas tendremos frecuencias o frecuencias relativas, según corresponda. Uniremos los puntos medios superiores de cada una de las barras de los histogramas, considerando nulas las frecuencias de los valores adyacentes a los valores de la distribución.

Luego, si consideramos los datos de la tabla 2-1, el gráfico será:

Gráfico 2.4

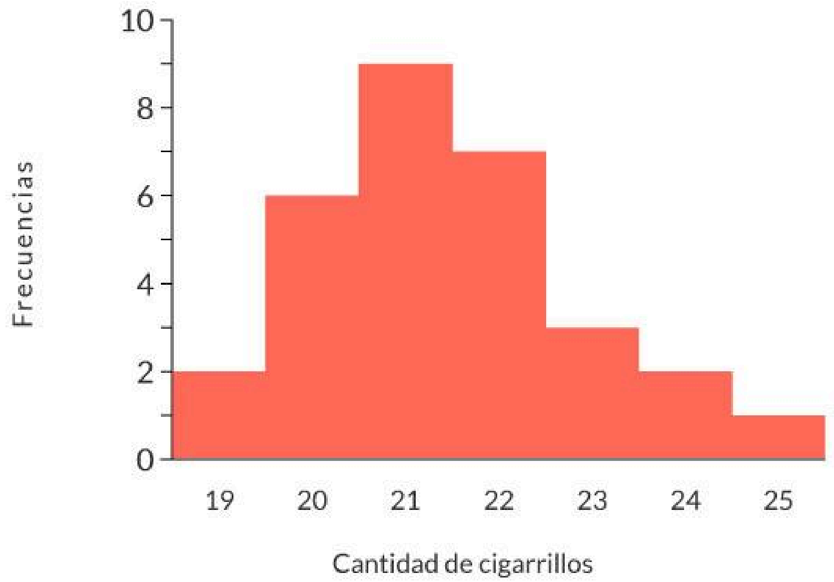
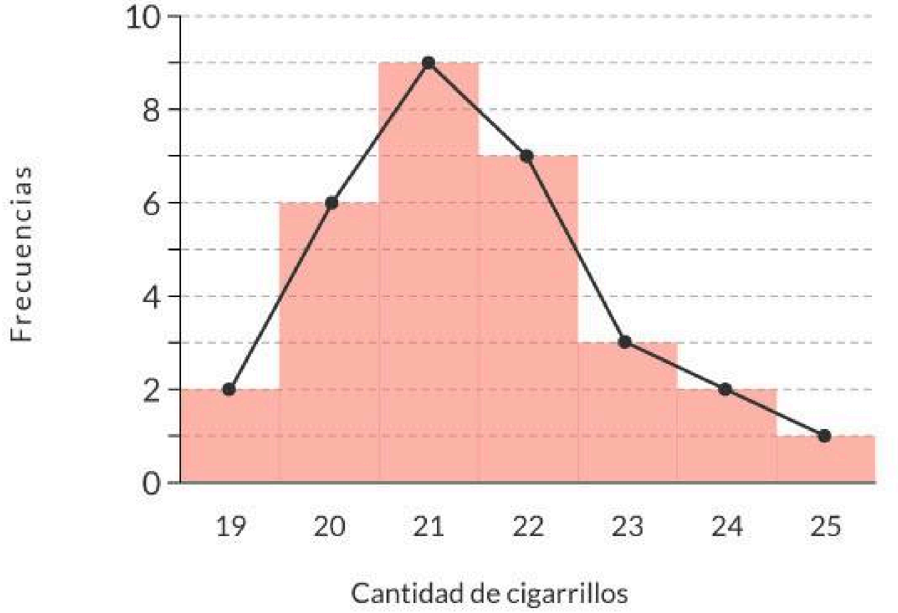


Gráfico 2.6



¿Cómo comparar y sacar conclusiones de este gráfico? Esto será analizado en detalle cuando desarrollemos el ejercicio resuelto.

Indique la opción correcta

1- Un comerciante estudia la venta de cierta marca de fideos. Por ello, analiza una muestra de 50 clientes, registrando la cantidad de paquetes vendidos en la última quincena. En base a los resultados obtenidos que se muestran en el cuadro, indique la moda de la distribución.

Cantidad de paquetes (x)	Frecuencia
1	3
2	7
3	9
4	15
5	8
6	6
7	2

- ☐ Mo=8
- ☐ Mo=4
- ☐ Mo=2

Indique la opción correcta

2- La diferencia entre dos clases consecutivas se denomina intervalo de clase.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- Un histograma se define como un diagrama de barras con discontinuidades.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- El área encerrada por el Histograma de frecuencias es igual a la suma de todas las observaciones.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- Cuando se necesita hacer comparaciones entre similares distribuciones, en general es conveniente realizarlo a partir de polígonos de frecuencias o frecuencias relativas.

- ☐ Verdadero
- ☐ Falso

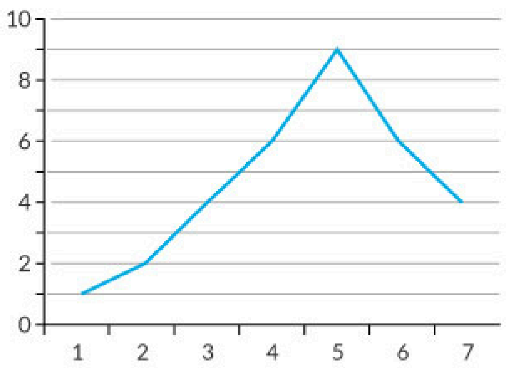
Respuestas correctas[5]

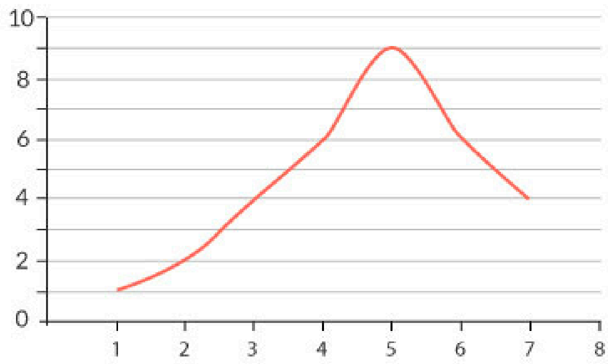
[5] 1) Mo=4. 2) Verdadero. 3) Falso. 4) Verdadero. 5) Falso.

SP2/H2: Relación entre la media, mediana y moda

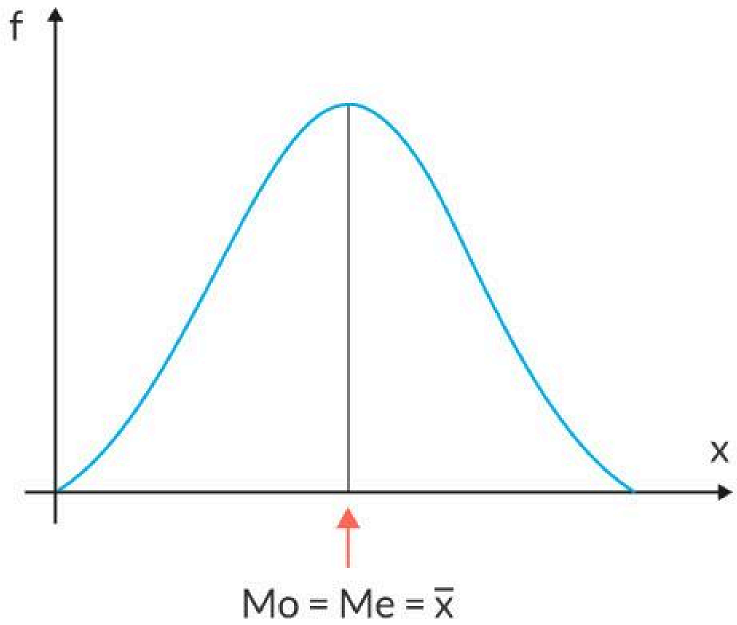
En todo polígono de frecuencias en donde sus lados fueron "**suavizados**" convenientemente, podemos sacar las siguientes conclusiones:

Suavizados



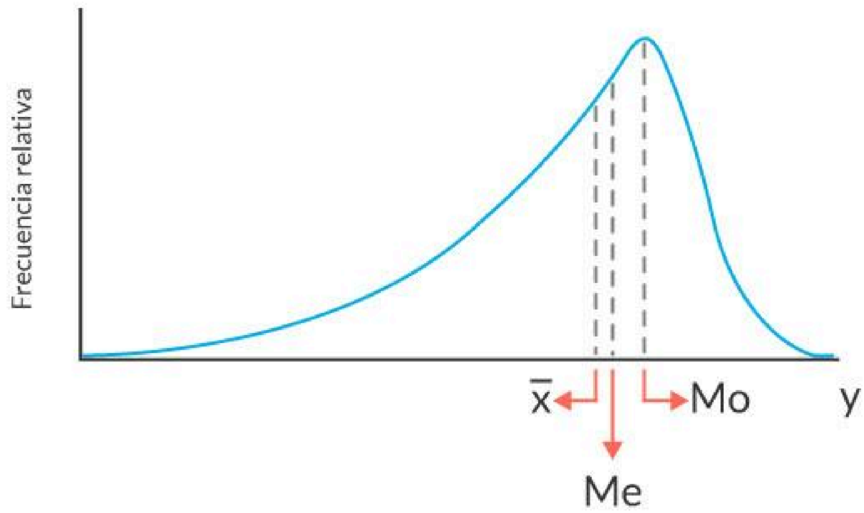


a. El valor de la abscisa que corresponde a la mayor ordenada, es la Moda.
b. El polígono puede presentarse en forma simétrica respecto de un eje vertical que pasa por la Moda.
c. El polígono puede presentar un sesgo o inclinación con respecto al eje vertical que pasa por la Moda Mo.
d. La mediana deja a su izquierda tantos valores como los que deja a la derecha.
e. La media es el punto de equilibrio de la distribución.



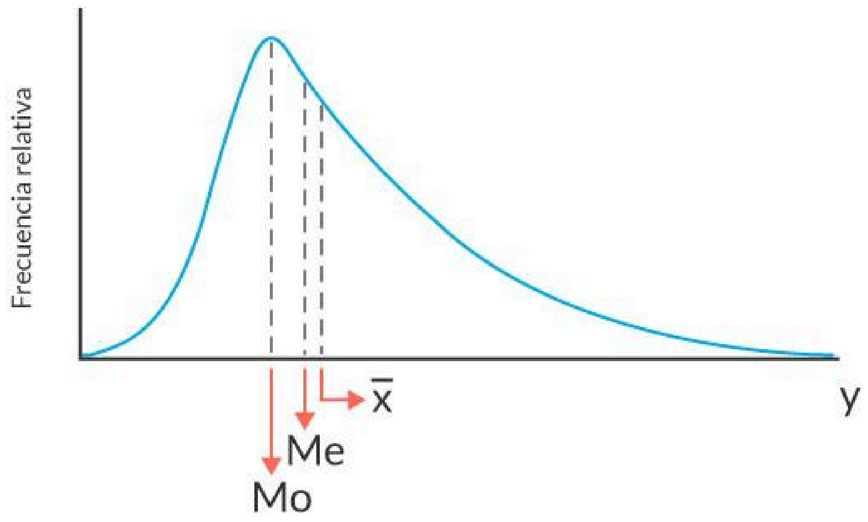
1. Cuando la curva es simétrica, los valores de Media, Mediana y Moda coinciden.
2. Cuando la curva es asimétrica, ésta puede ser:

2a. Asimétrica izquierda, se le asigna signo negativo.



(b) Distribución desplazada a la izquierda

2b. Asimétrica derecha, con signo positivo



(c) Distribución desplazada a la derecha

Para esquematizar, diremos que una distribución puede ser:

$$\text{Simétrica} \longrightarrow \left\{ M_o = M_e = \bar{x} \right.$$

$$\text{Asimétrica} \longrightarrow \begin{cases} \text{Derecha} & M_o < M_e < \bar{x} \\ \text{Izquierda} & M_o > M_e > \bar{x} \end{cases}$$

Una empresa de software e insumos de computación analiza el funcionamiento de un nuevo software desarrollado para resolver problemas matemáticos. Para ello, se ingresa cierto problema elegido con un grado alto de dificultad y se registra el tiempo que demora en resolverlo. Los tiempos registrados, medidos en segundos son:

6 9 5 7 5 8 3 5 4 2 5 4 4
4 4 2 5 6 3 7 3 6 8 4 4 4

Indique la opción correcta

1- La característica que tiene la distribución es...

- ☐ Una asimetría derecha
- ☐ Una asimetría izquierda
- ☐ Una asimetría centrada

Indique la opción correcta

2- El valor de la moda es...

- ☐ Mo=6 segundos
- ☐ Mo=3 segundos
- ☐ Mo=4 segundos

Indique la opción correcta

3- Teniendo en cuenta que un tiempo de 6 segundos o más se considera muy ineficiente, indique qué porcentaje de resultados está en esta categoría...

- ☐ El 40%
- ☐ El 30%
- ☐ El 10%

Indique la opción correcta

4- Los valores de Media, Mediana y Moda coinciden cuando la curva es:

simétrica

- ☐ asimétrica
- ☐ asimétrica izquierda

Indique la opción correcta

5- El valor de la media es:

- ☐ $\bar{x}$=4,88 segundos
- ☐ $\bar{x}$=4,38 segundos
- ☐ $\bar{x}$=3,88 segundos

Respuestas correctas[6]

[6] 1) Una asimetría derecha. 2) Mo=4 segundos. 3) El 30%. 4) asimétrica. 5) $\bar{x}$=4,88 segundos.

Su equipo de trabajo decide que los ítems a desarrollar para resolver la situación son:

a. Generar dos posibles distribuciones de frecuencia: una referida al "antes" del uso del producto, y otra referida al "después" del uso del mismo. De esta manera, se tiene una noción más clara de la situación en estudio.
b. Considerar una nueva distribución que resulta de la diferencia entre el número de cigarrillos fumados por día antes del uso del medicamento, y el número de cigarrillos fumados por día después.
c. Confeccionar histogramas de frecuencia y diagrama de sectores para la distribución que corresponde al consumo de cigarrillos antes del medicamento y después del mismo; y presentar los resultados que corresponden a la tercera distribución a través de un polígono de frecuencias relativas.
d. ¿Qué conclusiones puede sacar del polígono de frecuencias relativas para la distribución del "antes menos el después" del tratamiento?
e. Si el objetivo es determinar la bondad de este producto, y considerando el análisis gráfico antes realizado ¿qué se puede concluir?

Luego, analizando los ítems propuestos:

a. Recordemos que la distribución correspondiente al momento previo del tratamiento con este medicamento, que contiene en este caso, además, la frecuencia relativa de cada clase, está expresado en la tabla a continuación:

Ejercicio resuelto SP2 - tabla 2.1

Número de clase	Cantidad de cigarrilos	fi	xi . fi	fri
1	19	2	38	0.066667
2	20	6	120	0.2
3	21	9	189	0.3
4	22	7	154	0.233333
5	23	3	69	0.1
6	24	2	48	0.066667
7	25	1	25	0.033333

Además, si consideramos la distribución dada por la etapa posterior al tratamiento, es decir "el después" del tratamiento, la distribución correspondiente estará dada por la siguiente tabla:

Ejercicio resuelto SP2 - tabla 2.7

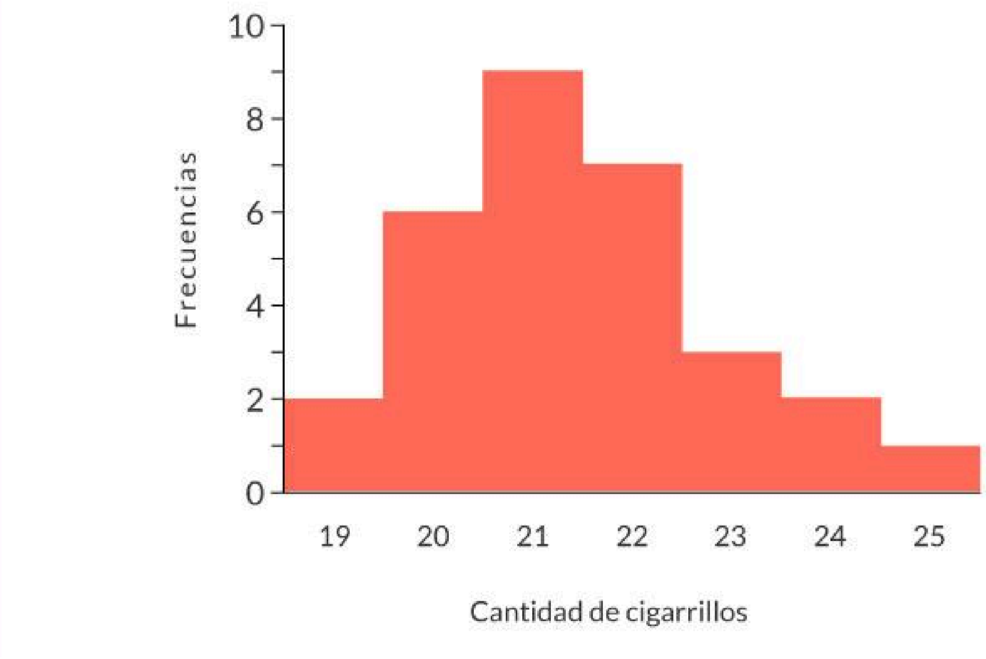
Número de clase	Cantidad de cigarrilos	fi	xi . fi	fri
1	6	1	6	0.033333
2	7	2	14	0.066667
3	8	10	80	0.333333
4	9	13	117	0.433333
5	10	2	20	0.066667
6	11	1	11	0.033333
7	12	1	12	0.033333

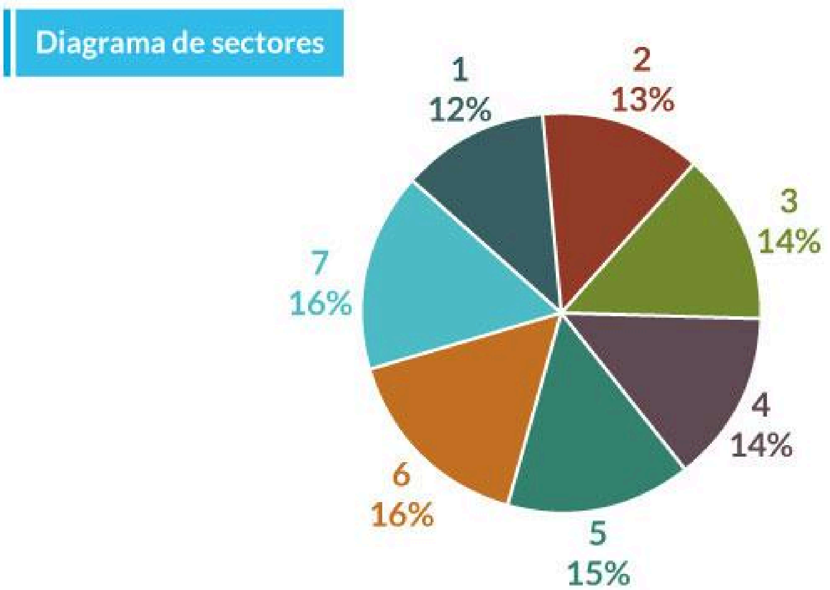
b. Existe una tercera distribución implícitamente planteada en esta situación, que resulta de la diferencia entre los cigarrillos fumados antes y después del tratamiento. A continuación construimos la tabla de frecuencias que corresponde a la misma, que será útil para el análisis posterior:

Ejercicio resuelto SP2 - tabla 2.8

Número de clase	Cantidad de cigarrilos	Frecuencia	xi . fi	fri
1	10	1	10	0.033333
2	11	6	66	0.2
3	12	5	50	0.166667
4	13	10	130	0.333333
5	14	3	42	0.1
6	15	5	75	0.166667

c. Distribución "antes" del uso del medicamento





Distribución "después" del uso del medicamento

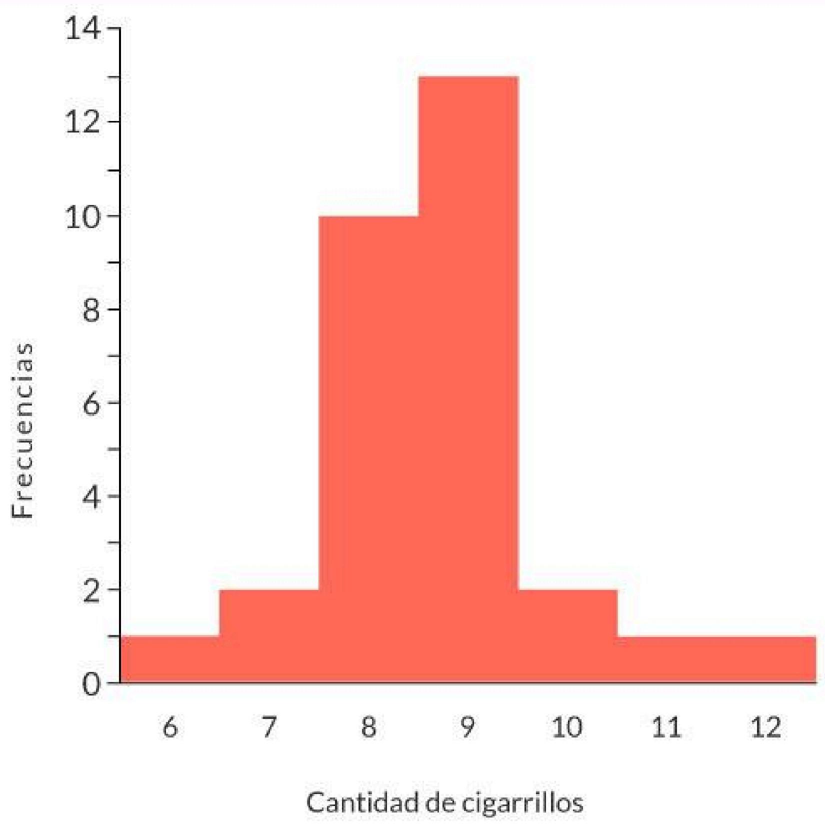
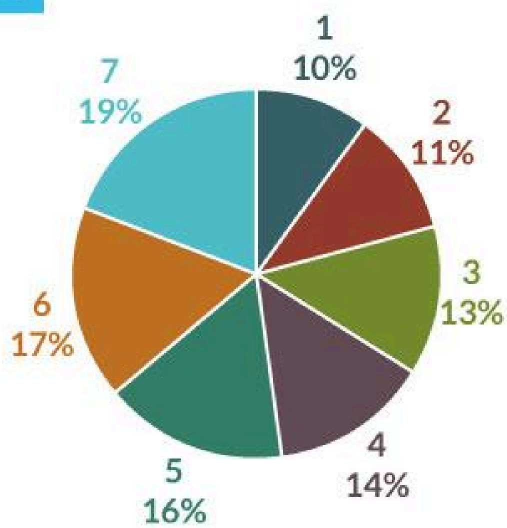
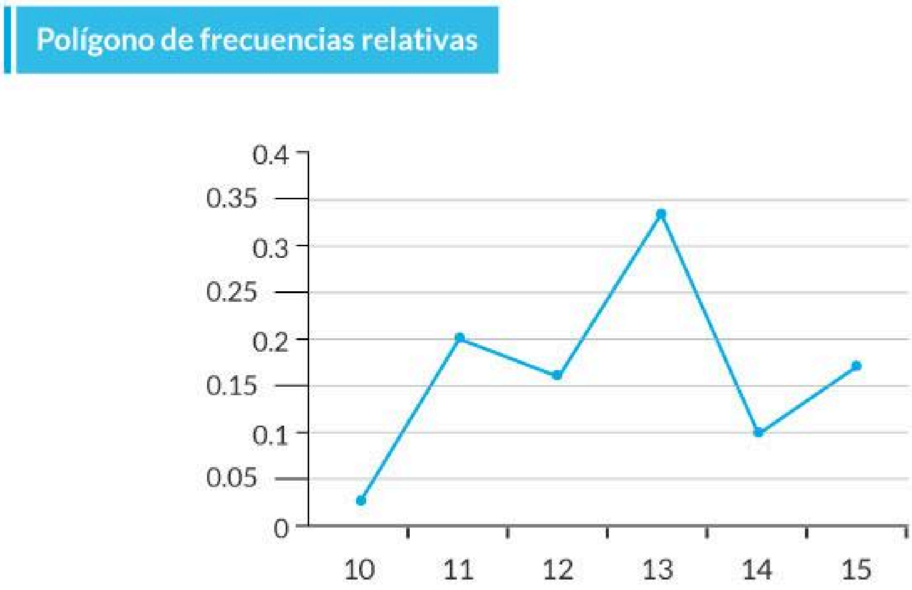


Diagrama de sectores



Distribución "antes menos después" del uso del medicamento



d. Si observa el polígono, podrá observar que la gráfica no es simétrica, sino que presenta cierta inclinación hacia la derecha; recuerde que este gráfico muestra la diferencia entre el consumo de cigarrillos antes y después del tratamiento. Considerando que, del gráfico, tenemos que: un grupo menor se encuentra en la zona tal que la diferencia está entre 10 y 12, y gran parte de las observaciones se encuentran entre 12 y 15, y esto quiere decir que este medicamento ha tenido gran impacto sobre el consumo de cigarrillos de los pacientes.

e. Podemos concluir, a partir del apartado anterior, que el medicamento tuvo un impacto positivo entre los pacientes. No obstante, chequearemos analíticamente la situación.

Observemos que si calculamos para cada una de las distribuciones, obtendremos lo siguiente:

1. En el caso de la distribución "antes" del medicamento, se tiene que $\bar{x}=643/30=21,43$ cigarrillos.
2. En el caso de la distribución "después" del medicamento, se tiene que $\bar{x}=260/30=8,66$ cigarrillos.

3. Por último, en el caso de la distribución de la diferencia, tendremos que $\bar{x}=373/30=12,43$ cigarrillos.

Nótese que estos resultados avalan nuestra conclusión anterior.

SP2/Ejercicio por resolver

Su equipo de trabajo decide que los ítems a desarrollar para resolver la situación son:
Una fábrica realiza una prueba a sus operarios para verificar el desempeño de los mismos ante posibles situaciones problemáticas en la producción, y plantear, sobre la base de los resultados, un curso de capacitación y perfeccionamiento. Para ello, selecciona a 20 operarios y toma una evaluación cuyo puntaje máximo es 20 puntos.

Las puntuaciones obtenidas han sido:

15, 20, 15, 18, 20, 13, 13, 16, 15, 19, 18, 15, 16, 20, 16, 15, 18, 16, 14, 13.

A partir de estos datos:

a. Construya la tabla de distribución de frecuencias.

b. Construya el correspondiente polígono de frecuencias.

SP2/Evaluación de paso

Indique la opción correcta

1- Cuando la curva es asimétrica, esta puede ser izquierda (positivo) o derecha (negativo).

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Cuando la curva es simétrica, los valores de Media, Mediana y Moda coinciden.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- La mediana es el punto de equilibrio de la distribución.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Un diagrama de sectores es un gráfico circular dividido en tantos sectores como clases tenga.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- La suma de las frecuencias relativas es igual a 0.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[7]

---

[7] 1) Falso. 2) Verdadero. 3) Falso. 4) Verdadero. 5) Falso.

Situación profesional 3: Distribución de intervalos de clases

Link del video: https://www.youtube.com/watch?v=wdzFEgV7ZMI

Texto del video: La nueva tarjeta de crédito "Credi Todo" decidió implementar una línea de atención al cliente para consultar sobre los planes de financiación, comercios adheridos y promociones vigentes para pagos con la tarjeta. Los directivos de la empresa desean determinar si este nuevo medio es útil para sus clientes, y en ese caso ampliar este nuevo sistema, ya sea en cuanto a personal, servicios y productos ofrecidos. Usted trabaja en el área de base de datos de la empresa, y decide registrar llamadas de clientes a lo largo de 60 días. Los resultados obtenidos se expresan en la siguiente tabla:

Sobre la base de esta información, se le pide al equipo de trabajo del cual usted forma parte, que determine si este servicio es realmente útil, y si se justifica la ampliación del área en estudio.

Además, si se tiene en cuenta el comportamiento de los datos, conforme pasaron los días, es necesario saber si con el correr de los mismos el número de llamadas aumentó o no, y de haber aumentado, poder sugerir una solución inmediata para mejorar el servicio.

Número de cliente	Cantidad de llamadas	Número de cliente	Cantidad de llamadas	Número de cliente	Cantidad de llamadas
1	31	15	40	29	31
2	7	16	28	30	44
3	21	17	17	31	33
4	44	18	52	32	33
5	30	19	21	33	44
6	40	20	52	34	48
7	13	21	32	35	58
8	27	22	40	36	51
9	11	23	32	37	60
10	7	24	21	38	37
11	21	25	33	39	44
12	37	26	30	40	30
13	33	27	27		
14	31	28	32		

SP3/H1: Distribución de Intervalos de Clases

Muchas veces, en situaciones concretas de trabajo, y ante un estudio estadístico a realizar, la cantidad de datos de los que se dispone para el estudio es bastante considerable, lo que hace dificultoso y lento, en muchas ocasiones, el manejo eficiente y correcto de la información. Por ello, bajo ciertas condiciones, existe una forma de reorganizar los datos, de manera tal que se torne más eficiente el manejo de los mismos, que se llama distribución de intervalos de clases. Esto asegurará que la información original, reorganizada de esta nueva forma, no perderá sus características principales, es decir, que no tendremos pérdida considerable de información, por lo cual, las características de la población en estudio resultarán las mismas, con la ventaja de la organización por intervalos de clases de los datos.

La razón fundamental para utilizar la distribución de intervalos de clases es proporcionar mejor comunicación acerca del patrón establecido en los datos, y facilitar la manipulación de los mismos. Los datos se agrupan en clases con el fin de sintetizar, resumir, condensar o hacer que la información obtenida de una investigación sea manejable con mayor facilidad.

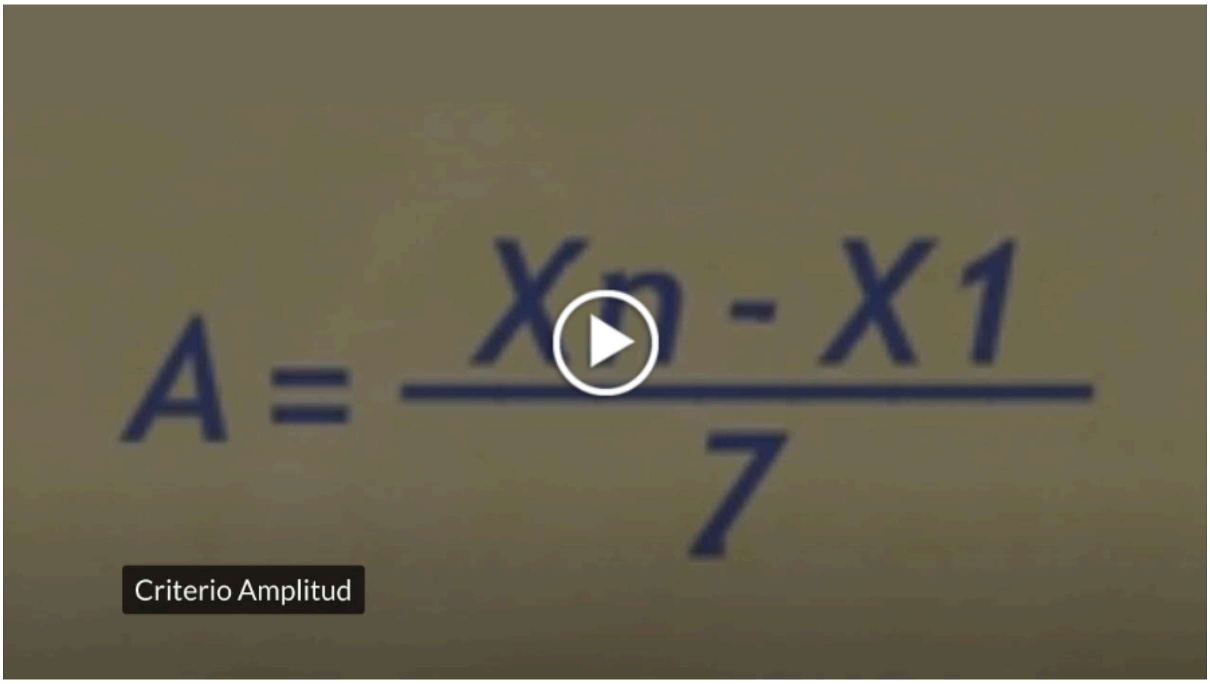
En aquellos casos que las observaciones sean mayores a 20, no se definen clases sino que tomamos intervalos de clase y seleccionamos el número de valores que caen en cada intervalo, considerando los siguientes criterios:

a.

Link del video: https://www.youtube.com/watch?v=C8DzvMBK3t8

Texto del vídeo: Determinar el número total de intervalos (z), que por regla general debe ser mayor a 5 y menor a 16. Se toma el valor aproximado de la raíz cuadrada de n (número total de observaciones).

b.



Link del video: https://www.youtube.com/watch?v=GZORXbiddDU

Texto del video: Los intervalos de clase deben ser de igual tamaño y el número de clases determina la amplitud del intervalo que se obtiene como el cociente entre, la diferencia entre el mayor y menor valor de la distribución, dividido el número de intervalos (z).

c. Los intervalos están definidos por su límite inferior y límite superior.

Tipos de intervalo:

a-[Li –Ls]

b-[Li- Ls)

c-(Li –Ls]

d-(Li-Ls)

Donde el corchete indica que el extremo del intervalo corresponde al mismo, y el paréntesis indica que ese extremo no pertenece al intervalo.

d. No debe haber ningún intervalo con frecuencia cero (excepto los intervalos extremos).

Además, dada esta nueva distribución, podríamos preguntarnos el valor que representa cada intervalo, ya que en cada uno pueden existir varias observaciones; para ello definiremos la marca de cada clase o valor representativo de la misma, que será un valor que represente cada intervalo de clase. La utilidad de estas marcas tomará protagonismo al momento de buscar valores de posición o de dispersión. Este concepto de marca de clase es análogo al valor de observación, al considerar una serie simple, estudiada anteriormente.

SP3/Autoevaluación 1

Indique la opción correcta

1- Cuando el número de observaciones es mayor a 20, no es conveniente tomar intervalos de clase.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Los intervalos están definidos únicamente por su límite superior.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- La raíz cuadrada de n proporciona un valor aproximado para el número apropiado de intervalos a utilizar.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- En las distribuciones de frecuencias:

- ☐ Los datos no entran en más de una clase.
- ☐ Generalmente existen más clases que datos.
- ☐ Los datos se reorganizan y se encuentran en alguna de las clases.
- ☐ Las afirmaciones anteriores son correctas.
- ☐ Las primeras dos afirmaciones son correctas.

Indique la opción correcta

5- En la elaboración de la distribución de clases se debe:

- ☐ Separar los datos en 5 clases.

- ☐ Dividir los datos en clases y contar la cantidad de observaciones que pertenecen a cada una de ellas.
- ☐ Los datos deben estar divididos en tipo y número de clases.
- ☐ Ninguno de los anteriores.

Indique la opción correcta

6- Como criterio general:

- ☐ Se utilizan entre 20 y 30 clases.
- ☐ Las clases muestran una característica de los datos para ser organizados.
- ☐ Se utilizan entre 6 y 15 clases.
- ☐ Ninguna de las anteriores.

Indique la opción correcta

7- La cantidad de clases que se utilizan para organizar los datos son:

- ☐ Menos de 6 clases.
- ☐ Entre 1 y 6 clases.
- ☐ Más de 20 clases.
- ☐ Entre 20 y 40 clases.
- ☐ Ninguna de las anteriores.

Indique la opción correcta

8- En la tabla de distribución de frecuencias cada dato pertenece a una única _______.

- ☐ serie de frecuencias
- ☐ frecuencia relativa
- ☐ clase

Indique la opción correcta

9- El número total de intervalos debe ser mayor a 5 y menor a ____.

- ☐ 16
- ☐ 60
- ☐ 100

Indique la opción correcta

10- Se adopta como amplitud del intervalo ______ derivado de la diferencia entre el mayor y menor valor de la distribución, dividido el número de intervalos.

- ☐ el producto
- ☐ el cociente
- ☐ la suma

Indique la opción correcta

11- Cuando se diseña una distribución de frecuencias para una muestra, el número de clases depende de:

- ☐ Cantidad de datos.
- ☐ Alcance de los datos recolectados.
- ☐ La población.

Respuestas correctas[8]

[8] 1) Falso. 2) Falso. 3) Verdadero. 4) Los datos se reorganizan y se encuentran en alguna de las clases. 5) Dividir los datos en clases y contar la cantidad de observaciones que pertenecen a cada una de ellas. 6) Se utilizan entre 6 y 15 clases. 7) Ninguna de las anteriores. 8) clase. 9) 16. 10) el cociente. 11) Cantidad de datos.

De acuerdo a las herramientas definidas, la situación planteada se resuelve de la siguiente manera:

Primer Paso: ordenamos las observaciones de menor a mayor valor.

Segundo Paso: de acuerdo a lo desarrollado, el número de intervalos z es $z=\sqrt{40}$, que es aproximadamente igual a 6.

Tercer Paso: definimos la amplitud de los intervalos, que denotamos Δx, y obtenemos que $\Delta x= (60-7)/6 = 53/6$, que es aproximadamente igual a 9.

Cuarto Paso: como hicimos notar anteriormente, el primer intervalo debe contener al límite inferior y el último intervalo al límite superior; por tanto el primer intervalo debe contener al valor 7, y el último debe contener al valor 60.

Tomaremos para el primer intervalo Li=6,5 y Ls=15,5.

Para el segundo intervalo tomamos Li=15,5 y Ls=24,5. Observemos que, como el límite superior de cada intervalo coincide con el límite inferior del intervalo que le sucede, tomaremos intervalos cerrados a izquierda y abiertos a derecha, es decir intervalos de la forma [Li, Ls).

Quinto Paso: debemos ahora determinar el número de valores que pertenecen a cada intervalo, dado que estos valores nos darán la frecuencia de observaciones, cuyos valores se encuentran en el intervalo indicado. Así, por ejemplo, en el primer intervalo considerado que es [6,5; 15,5) la frecuencia será 4.

Con estos valores es posible confeccionar una tabla con las frecuencias. Observemos que para poder determinar el valor de la media, y al contar con intervalos de clase, tenemos que determinar un valor que representa a cada intervalo.

Tomaremos el valor medio de cada intervalo como representante del mismo; de esta manera estaríamos suponiendo que los valores en cada intervalo se distribuyen de forma simétrica con respecto a este valor que llamaremos marca de la clase o valor testigo, que denotaremos **Xmi**. Como en muchos casos no se tiene certeza de la distribución simétrica de valores a lo largo del intervalo, este método genera una cierta pérdida mínima de información.

Para cada intervalo, entonces, el valor de la marca de la clase se obtiene:

$$Xmi = (Ls + Li) / 2$$

Así, la tabla que presenta las clases con sus frecuencias correspondientes y los valores de las marcas de las clases será:

Número de clase	Intervalo	fi	Xmi	Xmifi
1	[6,5; 15,5)	4	11	44
2	[15,5; 24,5)	5	20	100
3	[24,5; 33,5)	16	29	464
4	[33,5; 42,5)	5	38	190
5	[42,5; 51,5)	6	47	282
6	[51,5; 60,5)	4	56	224

Luego

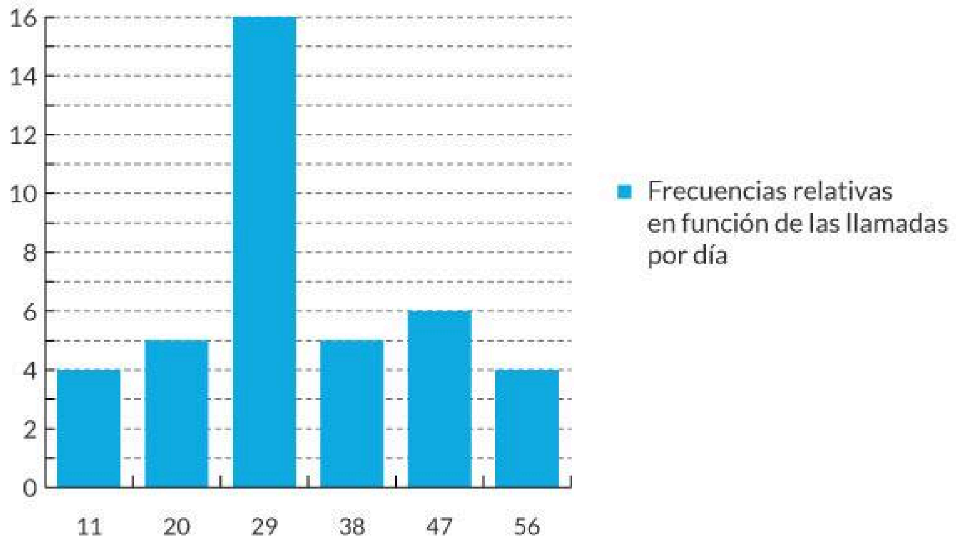
$$\sum fi = 40 \qquad \sum Xm_i f_i = 1304$$

Así, el valor de la media será:

$$\bar{x} = \frac{(\sum Xm_i f_i)}{\sum fi} = \frac{1304}{40} = 32{,}6$$

Es decir que, en promedio, la cantidad de llamadas que se reciben por día es de 32,6 llamadas.

Para obtener el histograma de frecuencias y/o de frecuencias relativas, representamos en el eje de las abscisas x las marcas de las clases y, a partir de cada una de ellas, se toma de derecha e izquierda el valor de Δx/2, levantando a partir de esos puntos los extremos de cada barra.



Como se observa en el histograma, en las tres primeras clases se concentra gran parte de las llamadas registradas. Esto indica que, según los registros, los clientes hacen uso de este servicio telefónico, y aunque en promedio la mayor cantidad de llamadas no se registran en los intervalos de mayores valores de llamadas, se tiene en promedio más de 30 llamadas al día. Si se sabe que las personas que trabajan ofreciendo estos servicios son, por turno, alrededor de 4, esto hace concluir que, de seguir ofreciendo el mismo servicio, no sería necesario agregar más personal, pero si se agregarán más servicios por esta vía, sería recomendable ampliar este sector de la empresa.

El Planificador Gerencial de la empresa "SUPERSÓNICOS" de la ciudad de Córdoba Capital, especializada en soldaduras de punto ultrasónicas que se realizan sobre un tipo especial de láminas de aluminio, ha recibido el informe de las observaciones que se han realizado con respecto a la resistencia del corte (en libras).

5435	4947	4522	4571	4989	5201	5240	5110
5014	4660	4805	4635	5669	4380	4819	5044
4885	4598	5287	5298	4845	5376	5054	5259
6000	5217	4860	4779	5026	5007	4608	4771
5132	5094	4617	4847	5088	5517	5332	5165
5341	5068	4754	4926	5005	4804	4952	5678
5255	5206	5622	4917	5137	4785	4501	5462
5048	4975	4591	4000	5295	4966	5169	4739
5172	4565	5652	5077	4890	4697	5249	5244
4722	5276	5420	5206	4450	5226	5554	5390
5499	4682	5077	4775	4932	4494	5310	5580
4307	4822	4418	5365	5639	5070	5189	5765
5274	5041	5190	4985				

A continuación le solicita a usted, como uno de los integrantes del equipo, que reorganice los datos y, dado que es una muestra de tamaño razonable, usted decide analizar el problema por distribución de intervalos de clases.

1. ¿Cuántos intervalos considera que debe utilizar?
2. ¿Qué amplitud tendrá cada intervalo?
3. Determine las frecuencias de cada clase, las marcas de las clases y obtenga la tabla de frecuencias.
4. Con lo obtenido, construya un histograma de frecuencias de las resistencias de las láminas. ¿Qué características de la distribución puede observar, que de la tabla original no se infiere?

Indique la opción correcta

1- Cuando el número de observaciones es mayor a 20, no es conveniente tomar intervalos de clase.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Los intervalos están definidos únicamente por su límite superior.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- La raíz cuadrada de n proporciona un valor aproximado para el número apropiado de intervalos a utilizar.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- El valor que representa a cada clase es el límite inferior.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- Los intervalos pueden ir variando en su forma.

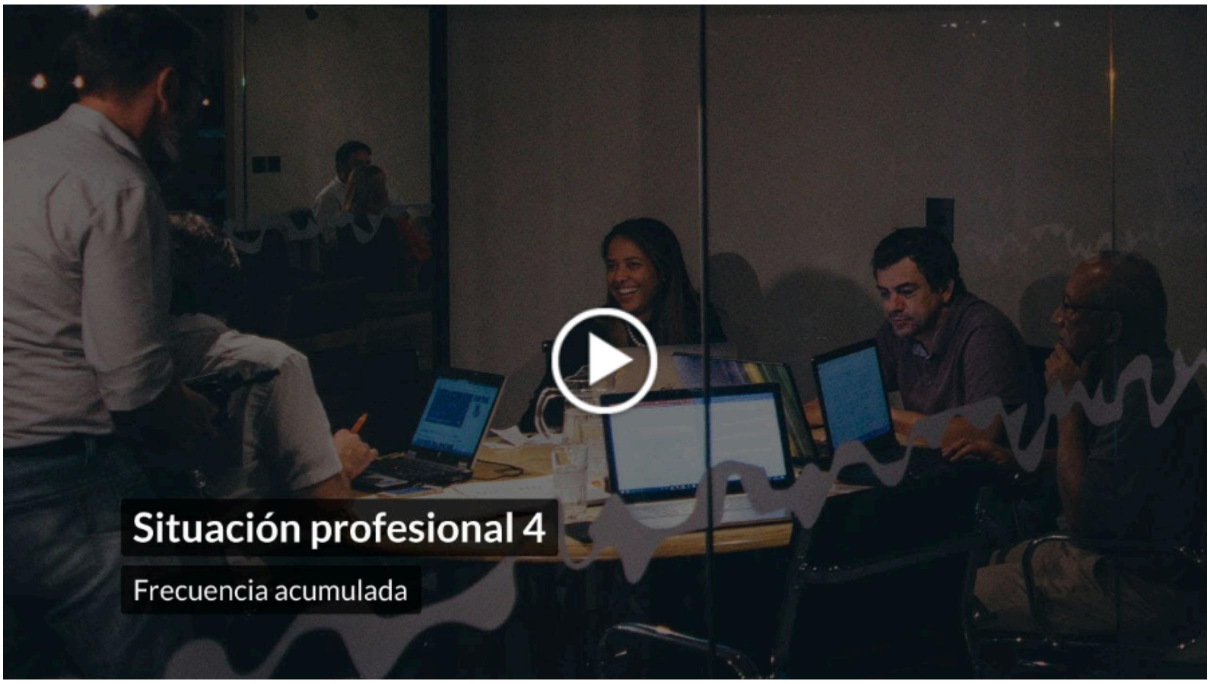
- ☐ Verdadero
- ☐ Falso

Respuestas correctas[9]

---

[9] 1) Falso. 2) Falso. 3) Verdadero. 4) Falso. 5) Falso.

Situación profesional 4: Frecuencia acumulada



Link del video: https://www.youtube.com/watch?v=O_IRmc6g4HM

Texto del video: La Compañía Agrícola "Prosperidad S.A.", está formada por un grupo de empresarios que invierten en la producción y comercialización de productos agrícolas en la provincia de Córdoba, y exportación de los mismos a mercados internos e incluso a mercados europeos. Esta compañía investiga sobre la posibilidad de producir frutilla, ya que diversos estudios indicarían la posibilidad de que esta producción sea exitosa en suelos de las sierras cordobesas.

Se conoce que el tener un clima libre de heladas y pocas precipitaciones significa una potencialidad importante, ya que en el caso de la frutilla no se necesitaría proteger el cultivo con túneles plásticos para evitar que las heladas quemen el cultivo o controlar excesos de humedad por lluvias, tarea que representa casi un 30% del costo de producción de la frutilla. Por ello se estudian los registros de la distribución de lluvias por mes en la época de mayo-septiembre (5 meses) en los últimos 20 años en la zona (es decir que se tiene registro de 100 meses en total), ya que por las demás condiciones (humedad relativa que no supera el 70% y que existe la disponibilidad de agua para riego por canales) se tiene una situación favorable.

Hay registros de la frecuencia de cantidad de lluvia caída en los últimos 20 años.

Los registros de lluvia en la zona estudiada son:

Esta empresa estaría dispuesta a invertir en el cultivo de la frutilla, en caso de que las frecuencias que corresponden a condiciones de déficit moderado, condiciones favorables o muy favorables sean mayores al 60%.

En su carácter de empleado del área de desarrollo de producción, y a través del análisis de la situación, deberá determinar si este será un negocio rentable para la empresa.

Milímetros de agua	Frecuencias (en meses)
Entre 60 Mm. y 120 Mm. *(déficit moderado)*	20
Entre 120 Mm. y 180 Mm. *(favorable)*	40
Entre 180 Mm. y 240 Mm. *(muy favorable)*	30
Entre 240 Mm y 300 Mm. *(excesivo)*	10

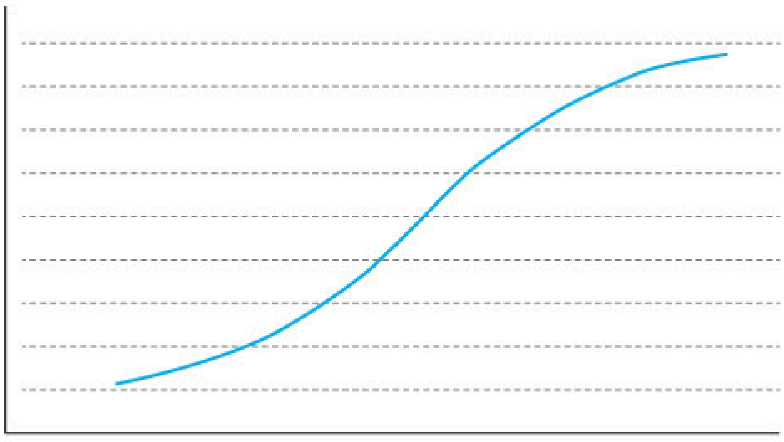
SP4/H1: Gráfica de frecuencias acumuladas. Ojiva

Hasta ahora, al trabajar con distribución de Intervalos de clases, construimos tablas que contienen los intervalos en estudio, las marcas de las clases y sus correspondientes frecuencias y/o frecuencias relativas. Ahora agregaremos a las tablas antes construidas, dos nuevas columnas: una con las frecuencias acumuladas de cada clase, y otra con las frecuencias relativas acumuladas de cada clase.

Definiremos como ojiva al gráfico correspondiente a las frecuencias acumuladas. Es decir que en el eje de las abscisas se tienen las observaciones (o marcas de clases), y en las ordenadas se indican las frecuencias acumuladas. La ojiva, en la estadística descriptiva, tiene interesantes aplicaciones.

La **ojiva** es una gráfica asociada a la distribución de frecuencias (o distribución por intervalos de clases); en ella se permite ver cuántas observaciones se encuentran por encima o debajo de ciertos valores. La ojiva es una gráfica similar al polígono de frecuencias, pero ésta se obtiene de aplicar parcialmente la misma técnica a una distribución acumulativa.

Ojiva



Para efectuar la gráfica de una ojiva nos apoyaremos en la técnica para construir histogramas de frecuencias y/o frecuencias relativas. En el eje de las abscisas representaremos los intervalos de clase, y en el de las ordenadas la frecuencia acumulada.

Determinación de la mediana

En una distribución de intervalos de clase, en general no se dispone de la serie simple; por lo tanto, presuponer que la mediana es la marca de la clase, del intervalo para el cual la frecuencia acumulada contiene al elemento (n+1)/2 es totalmente aproximado (recuerde que la mediana de una distribución, determinada a partir de una serie simple, es el valor que deja el mismo número de valores tanto a su izquierda como a su derecha, es decir que sería:

$$X_{(n+1)/2} \longrightarrow \text{si } n \text{ es } \textit{impar}.$$

$$\frac{X_{n/2} + X_{n/2+1}}{2} \longrightarrow \text{si } n \text{ es } \textit{par}.$$

A los fines de simplificar el cálculo de la mediana en el caso de tener una distribución por intervalos de clase, esta medida de tendencia central se calcula de la siguiente manera:

$$M_e = \left(\frac{n/2 - f_a}{f_m} \right) . \Delta x + Li$$

SP4/Autoevaluación 1

Indique la opción correcta

1- Por medio de la ojiva relativa el porcentaje del flujo que ocurre en menos de 1300 litros por minuto es:

- ☐ 88,2%
- ☐ 64,2%
- ☐ 84,2%

Indique la opción correcta

2- La ojiva es el gráfico correspondiente a las frecuencias acumuladas.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- A través de la ojiva se permite ver cuántas observaciones se encuentran por encima o debajo de ciertos valores.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Una gráfica diferente al polígono de frecuencias es la ojiva, ya que se obtiene de aplicar parcialmente la técnica a una distribución acumulativa.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- Para elaborar la gráfica de una ojiva representaremos en el eje de las abscisas los intervalos de frecuencia acumulada, y en el de las ordenadas la clase.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[10]

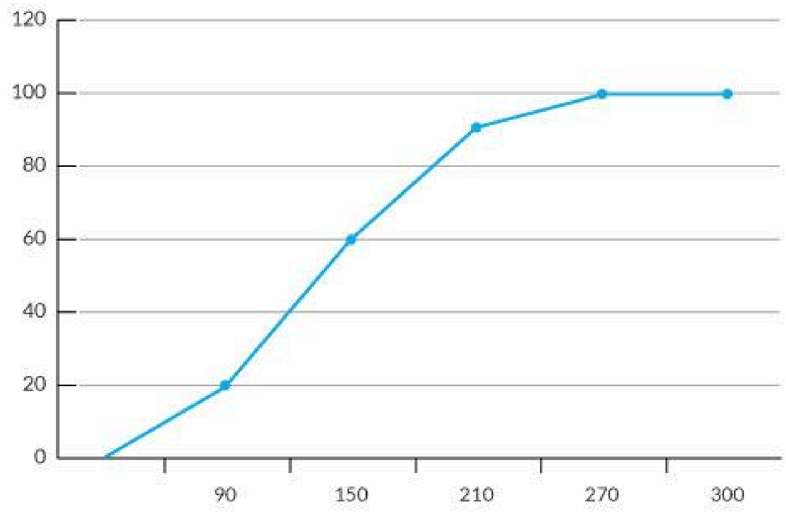
[10] 1) 84,2%. 2) Verdadero. 3) Verdadero. 4) Falso. 5) Falso.

Para responder al problema, confeccionaremos una tabla con las frecuencias acumuladas relativas (la que nos entregará porcentajes al multiplicar las mismas por 100%) y la correspondiente ojiva. Observemos que los datos del problema ya están expresados como una distribución de intervalos de clase, por tanto sólo falta agregar las columnas de frecuencias acumuladas, frecuencias relativas y frecuencias relativas acumuladas de las respectivas clases.

Obtenemos la siguiente tabla:

Clase	Frecuencias (f_i)	Frecuencias relativas (fr)	Frecuencias acumuladas (fa)	Frecuencias acumuladas relativas (fra)
60 - 120 Mm.	20	0,2	20	0,2
120 - 180 Mm.	40	0,4	60	0,6
180 - 240 Mm.	30	0,3	90	0,9
240 - 300 Mm.	10	0,1	100	1

Gráfico de frecuencias acumuladas



Fuente: Elaboración propia - DEPROE

De acuerdo a la tabla y la ojiva, el 90% de los registros resultan condiciones rentables para los empresarios, ya que la frecuencia relativa acumulada en dichas condiciones es de 0,9, por tanto la frecuencia relativa acumulada porcentual será de un 90%. Concluimos que se recomienda iniciar este nuevo emprendimiento ya que, de acuerdo al análisis anterior, resultará un negocio muy rentable.

SP4/Ejercicio por resolver

Usted trabaja en el área de estadística de cierta línea aérea, y el presidente del Consejo de administración le ha pedido que recoja y organice datos relativos a las operaciones de vuelo. Su interés principal a partir de los valores diarios se centra en la variable del número de pasajeros.

Ha obtenido estos datos de la cantidad de pasajeros por día de los diarios de vuelo de los últimos 50 días y ha reflejado esta información:

68	72	50	70	65	83	77	78	80	93
71	74	60	84	72	84	73	81	84	92
77	57	70	59	85	74	78	79	91	102
83	67	66	75	79	82	93	90	101	80
69	76	94	71	97	95	83	86	69	85

a. Construya la tabla de distribución de intervalos de clases.
b. Construya un histograma y un polígono de frecuencias.
c. Construya una ojiva, y sobre la base de ella responda si la empresa considera que si por lo menos en el 70% de los viajes se tienen más de 90 pasajeros, la empresa estaría dispuesta a agregar más vuelos diarios. Sobre la base de los datos del problema ¿qué recomendaría a la línea aérea?

Indique la opción correcta

1- Se define a la ojiva como un gráfico de frecuencias relativas.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Al construir una ojiva, observaremos en las abscisas los intervalos de clase, y en las ordenadas las frecuencias acumuladas.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- La utilidad de la ojiva es que permite ver cuántas observaciones se encuentran por encima o debajo de ciertos valores.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- El procedimiento para construir una ojiva es completamente diferente al que se utiliza para construir un histograma.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

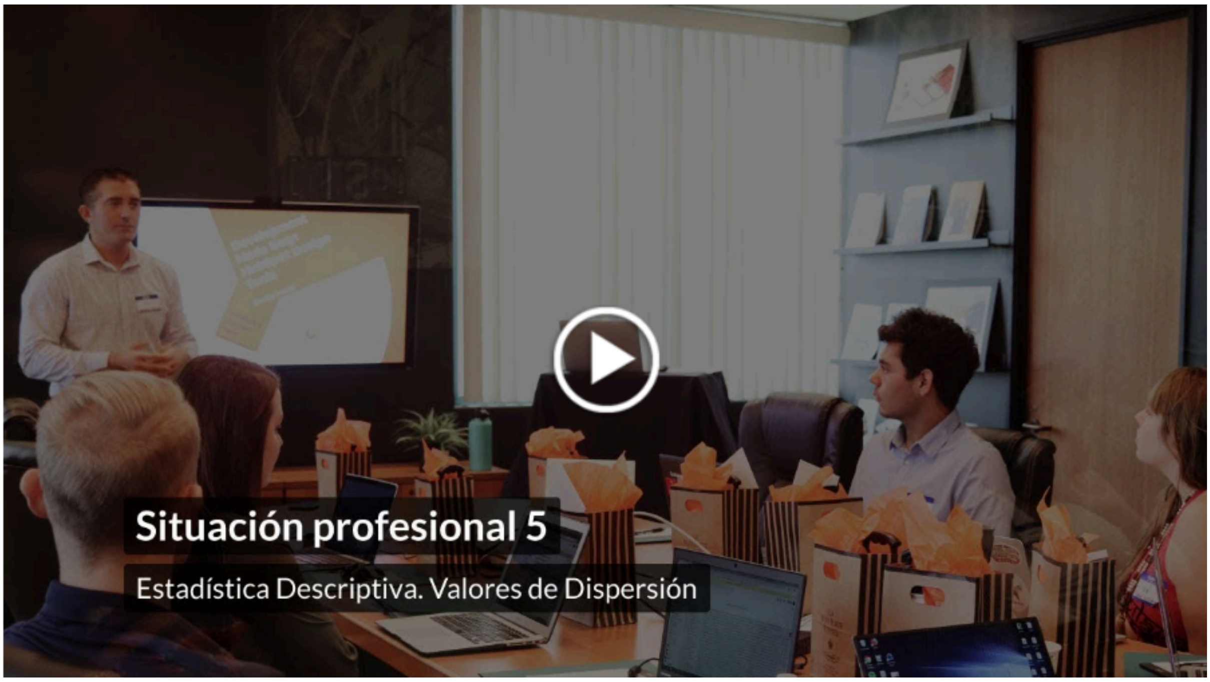
5- La clase mediana es la que tiene mayor frecuencia.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[11]

[11] 1) Falso. 2) Verdadero. 3) Verdadero. 4) Falso. 5) Falso.

Situación profesional 5: Estadística Descriptiva. Valores de Dispersión



Link del video: https://www.youtube.com/watch?v=ySK5yZhnKkE

Texto del video: La Empresa “Lechita” ubicada en la ciudad de Alta Gracia, se encuentra colaborando con un grupo de investigadores, pertenecientes a una revista especializada en nutrición, que desea obtener conclusiones sobre el volumen de leche que puede generar una madre “fumadora” versus una madre “no fumadora”, durante su etapa de lactancia.

Las muestras obtenidas son las siguientes (gramos/día):

Madres fumadoras

621 753 767 714 545 655

693 895 793 593 598 S/D

Madres no fumadoras

899 1086 745 981 961 947

1202 903 945 973 930 S/D

Fuente: DEVORE, Jay: Revista de Nutrición.

Usted como pasante debe comparar las dos muestras en función de la media y desviación estándar, para luego presentarle el informe al equipo de especialistas.

Madres fumadoras					
621	753	767	714	545	655
693	895	793	593	598	S/D

Madres no fumadoras					
899	1086	745	981	961	947
1202	903	945	973	930	S/D

Fuente: DEVORE, Jay. Revista de nutrición.

SP5/H1: Valores de dispersión

Los valores de posición estudiados hasta el momento nos muestran un cuadro, en cierta forma satisfactorio en cuanto a las características de la muestra; pero nada agrega con respecto a la dispersión que tienen los valores de la distribución con respecto a un valor fijo.

Dos distribuciones que tienen la misma cantidad de valores, y poseen la misma media y mediana pueden tener características totalmente distintas en lo que se refiere a la distribución de sus valores.

A	*********	********	*********
B	**		

Si tomamos dos distribuciones a las que designaremos A y B, y representamos a cada valor de ambas distribuciones con una notación (asteriscos) observamos que, tanto la distribución A como la B, tienen la misma media y no obstante ello, son distintas.

“C”	3	-1	2	2	2	6	6	6
“D”	4	-1	4	4	4	4	5	5

En el caso de C y D distribuciones de igual número de valores tienen la misma media 2, no obstante son distintas las distribuciones.

Por lo tanto, es conveniente efectuar (conjuntamente con el estudio de los valores de posición) un análisis que permita establecer cuál es el grado de dispersión que presentan las observaciones de una distribución, con respecto a un valor fijo.

Los valores de dispersión a estudiar son:

a. Alcance o Rango

b. Desvío medio

c. Varianza

d. Desvío estándar

a. Alcance o Rango

La cadena de supermercados "La Paz", que se menciona en la situación profesional 1 de la ciudad de Córdoba se encuentra realizando un estudio estadístico de los talles de sus empleados, debido a un reclamo que se ha producido en la Gerencia General, por el aspecto personal que presentan algunos de ellos, lo que incide negativamente en la imagen de la organización. Con la finalidad de revertir dicha situación, la división abastecimiento ha efectuado un muestreo de 49 empleados sobre un total de 500, que es el plantel permanente, registrando la altura de cada uno de ellos.

El registro de las alturas se vuelca en la siguiente planilla:

Empleados	Altura	Empleados	Altura	Empleados	Altura
1	172	18	183	35	169
2	177	19	167	36	172
3	161	20	181	37	183
4	171	21	156	38	175
5	170	22	182	39	178
6	161	23	179	40	176
7	160	24	173	41	187
8	173	25	179	42	186
9	179	26	177	43	169
10	173	27	170	44	183
11	164	28	173	45	190
12	167	29	191	46	161
13	168	30	184	47	184
14	166	31	182	48	175
15	177	32	173	49	167
16	177	33	178		
17	176	34	174		

Link del video: https://www.youtube.com/watch?v=-mBNyiLuYg0

Texto del video: Podemos observar que la mayor altura registrada corresponde a Xn = 191 cm y la menor a X1= 156 cm, por lo tanto la amplitud de la distribución generada por las 49 observaciones es de: A = 191 – 156 = 35

La diferencia entre el mayor y menor valor observado de una distribución, se conoce como alcance o rango.

Hemos definido a Xn como el mayor valor observado y a X1 como el más pequeño, entonces: A = Xn – X1

Aunque, una vez ordenados todos los valores, el cálculo del alcance o rango es simple e inmediato, debemos tener en cuenta que no nos brinda información sobre las características de los valores intermedios. Veamos las siguientes distribuciones:

"A"	0	1	1	2	4	6	8	10	12	14
"B"	50	50,1	50,2	50,3	50,5	50,6	50,7	50,8	60	64

Ambas tienen el mismo número de elementos, el mismo alcance y, sin embargo, son distintas.

$$\text{En "A"} \longrightarrow A = 14 - 0 = 14$$

$$\text{En "B"} \longrightarrow B = 64 - 50 = 14$$

Por lo tanto, su aplicación es limitada, ya que la información otorgada por el alcance o rango puede en muchos casos, ser equívoca (falaz), debido a que sólo se apoya en dos valores e ignora el resto.

b. Desvío medio

Continuando con el ejemplo de la cadena de supermercados "La Paz", podemos confeccionar una distribución de intervalos de clase, tomando un número de intervalos igual a z = 7, con lo cual de acuerdo con lo analizado anteriormente, nos entrega una amplitud de intervalo de:

$$\Delta x = \frac{191 - 156}{7} = 5$$

El límite inferior del primer intervalo será de 155,5 de forma tal que el límite superior tomará el valor de 160,5; de esta manera nos aseguramos que ninguna observación coincida con los límites de los intervalos.

L_i	L_s	f_i	$X_{mi} = (Ls+Li)/2$
155,5	160,5	2	158
160,5	165,5	4	163
165,5	170,5	9	168
170,5	175,5	11	173
175,5	180,5	11	178
180,5	185,5	8	183
185,5	190,5	3	188
190,5	195,5	1	193

Completemos la tabla obteniendo los productos de las marcas de clase de cada intervalo por su correspondiente frecuencia, y determinemos, además, la frecuencia relativa de cada intervalo.

Li	Ls	F	Xmi=(Ls+Li)/2	Xmi. fi	Fri
155,5	160,5	2	158	316	0,04
160,5	165,5	4	163	652	0,08
165,5	170,5	9	168	1512	0,18
170,5	175,5	11	173	1903	0,22
175,5	180,5	11	178	1958	0,22
180,5	185,5	8	183	1464	0,16
185,5	190,5	3	188	564	0,06
190,5	195,5	1	193	193	0,02
		49		8562	1

Entonces la media de la distribución es:

$$\overline{x} = \frac{\sum Xmi\,.\,fi}{\sum fi} = \frac{8562}{49}$$

$$\overline{x} = 174{,}7 \text{ cm.}$$

Ahora tenemos las 49 observaciones distribuidas en intervalos (es lo que se define como frecuencia de cada intervalo) y el valor que representa a cada intervalo es la marca de clase xmi.

Por lo tanto, estamos en condiciones de generar una nueva columna que contenga los desvíos, en valor absoluto, ya que siendo la media el valor donde se mueve la distribución, nos encontraremos con diferencias positivas y negativas, con respecto a la ubicación de cada observación, conforme a la media y la suma de todas ellas, nos dará cero.

Si lo que buscamos es establecer cuánto se alejan las observaciones de un valor tomado como base y fijo, tal como lo sería la media, en verdad, entonces el valor más apropiado sería el promedio de todas las diferencias:

$$(X_{mi} - \overline{X})$$

Se define como desvío medio en una distribución, al promedio de las desviaciones absolutas de las observaciones con respecto a la media.

Entonces la tabla queda de la siguiente forma:

L_i	L_s	F	Xmi=(Ls+Li)/2	Xmi.fi	Fri	$\lvert X_{mi} - \overline{X} \rvert .fi$
155,5	160,5	2	158	316	0,04	33,35
160,5	165,5	4	163	652	0,08	46,69
165,5	170,5	9	168	1512	0,18	60,06
170,5	175,5	11	173	1903	0,22	18,41
175,5	180,5	11	178	1958	0,22	36,59
180,5	185,5	8	183	1464	0,16	66,61
185,5	190,5	3	188	564	0,06	39,90
190,5	195,5	1	193	193	0,02	18,30
		49		8562	1	319,91

De acuerdo con la definición de desvío medio:

$$DM = \frac{|X_{mi} - \overline{X}|\, fi}{\Sigma\, fi} = \frac{319{,}91}{49}$$

$$DM = 6,53\ \text{cm}$$

Debemos aclarar que si bien el desvío medio no es de gran utilización, usted debe conocerlo y relacionarlo con el desvío estándar, valor que sí es de gran aplicación práctica.

c. Varianza

Por el momento tenemos un valor que nos indica un grado de dispersión de las observaciones con respecto a la media, nos referimos al desvío medio. Busquemos obtener otro valor que cumpla una función similar. Por lo tanto, en lugar de tomar los valores absolutos de los desvíos medios, elevemos al cuadrado cada uno de ellos, así nos queda positivo.

Se define como varianza de una distribución, al promedio de los desvíos cuadráticos medios.

La varianza se enuncia como Var(x) o también como σ^2.

De acuerdo con su definición es:

$$Var_{(x)} = \frac{\sum (X_{mi} - \bar{x})^2 \; fi}{\sum fi}$$

Es conveniente tener en consideración lo siguiente: la sumatoria de todas las frecuencias es igual al número total de observaciones, N para cuando nos referimos a una población.

En el caso de manejarnos con una muestra, la sumatoria de las fi conciernen al número de observaciones y se las denomina n.

Cuando se refiere a una muestra pequeña, menor de 30 observaciones, al denominador de la expresión de la varianza se le resta una unidad, mientras que si el número de elementos que contiene la muestra es grande, la diferencia de una unidad en el denominador no varía mayormente el valor de la varianza.

Así, debemos tener en cuenta dos expresiones para la varianza, según se trate de una población o una muestra.

Población $Var(x) = \dfrac{\sum (Xi - \mu)^2 . fi}{N}$

Muestra pequeña $Var_{(x)} = \dfrac{\sum (X_{mi} - \overline{X})^2 . fi}{n - 1}$

Recuerde que la expresión que nos define la media en una población es:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

Aunque la varianza nos da un grado de dispersión de los valores de una distribución, presenta la restricción de que sus unidades no coinciden con la de los valores, por estar elevadas al cuadrado, por ejemplo: pesos cuadrados. Este problema se soluciona con la aplicación del desvío estándar.

d. Desvío estándar

Se define como desvío estándar en una distribución, a la raíz cuadrada de la varianza.
La ventaja que presenta el desvío estándar respecto de la varianza, es que sus unidades coinciden con la de los valores de la distribución.

Población $$\sigma = \sqrt{\frac{\sum (xi - \mu)^2 . f_i}{N}}$$

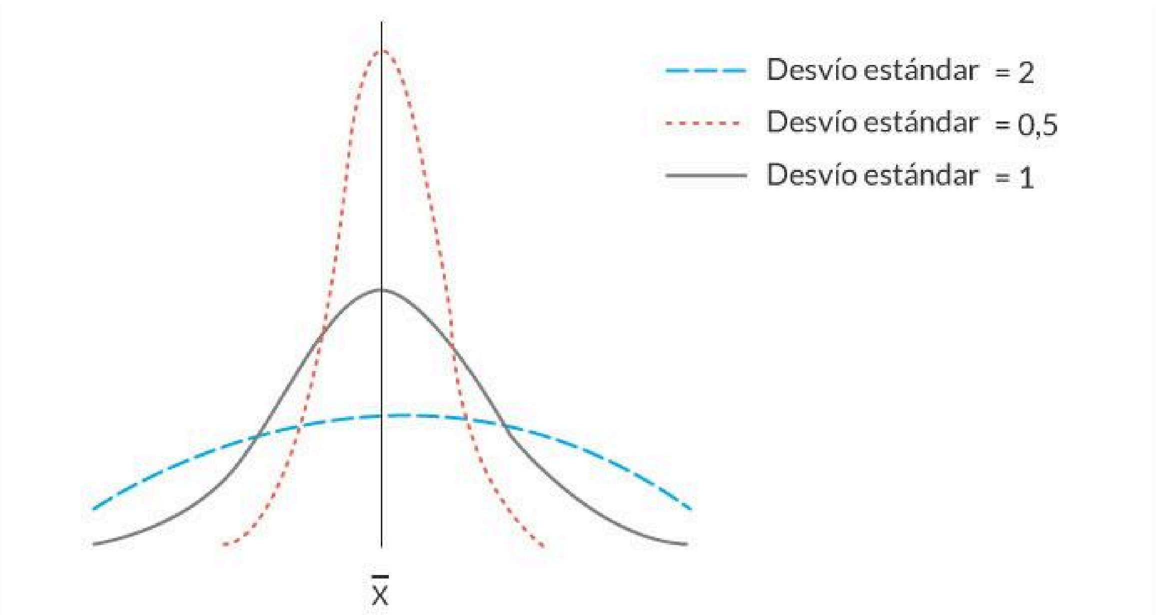
Muestra $$\sigma = \sqrt{\frac{(Xmi - \overline{x})^2 \, f_i}{n - 1}}$$

El desvío estándar se convierte en un valor de gran importancia para definir en una distribución el grado de dispersión de sus valores.

Un valor pequeño de sigma con respecto a sus observaciones, implica que los valores se concentran próximos a su media, por el contrario, un valor elevado de sigma mostraría que los valores se encuentran dispersos y alejados de su media.

Si graficamos el polígono de frecuencias relativas de una distribución y afinamos convenientemente sus lados, obtendremos una gráfica que tiende a ser acampanada, esto nos mostrará la dispersión de esa distribución.

Si tomamos tres distribuciones que posean la misma media, pero con valores de sigma distintos, tal como lo presenta el siguiente gráfico:



Las tres curvas toman la misma superficie, 1, tienen la misma media, pero los valores que forman cada una de sus distribuciones están colocados de manera distinta.

Si la campana se cierra, nos muestra que los valores tienden a agruparse alrededor de la media y, por consiguiente, tendrán una menor dispersión, por el contrario, si la curva se achata, nos muestra claramente una mayor dispersión de los valores que forman dicha distribución.

En consecuencia, ahora podemos obtener, para la distribución de los valores pertenecientes a las alturas de los empleados, el valor de la varianza y del desvío estándar.

Entonces confeccionemos una última columna que quedará de la siguiente forma:

L_i	L_s	F	X_{mi}=(Ls+Li)/2	X_{mi} . fi	Fri	(X_{mi} - $\bar{X}$). fi	(X_{mi} - $\bar{X}$)^2 . fi
155,5	160,5	2	158	316	0,04	33,35	557,78
160,5	165,5	4	163	652	0,08	46,69	547,56
165,5	170,5	9	168	1512	0,18	60,06	404,01
170,5	175,5	11	173	1903	0,22	18,41	31,79
175,5	180,5	11	178	1958	0,22	36,59	119,79
180,5	185,5	8	183	1464	0,16	66,61	551,12
185,5	190,5	3	188	564	0,06	39,90	530,67
190,5	195,5	1	193	193	0,02	18,30	334,89
		49	8562	1		319,91	3077,61

Por lo tanto, surge lo siguiente:

$$\text{Var (x)} = \frac{3077{,}61}{49 - 1} = 64{,}11 \text{ cm}^2$$

Y el desvío estándar:

$$\sigma = \sqrt{Var(x)} = \sqrt{64,11}$$

$$\sigma = 8\,cm$$

Debemos observar que el desvío medio que obtuvimos anteriormente fue de 6,53 cm., un valor no tan alejado del desvío estándar.

El desvío estándar puede adquirir el valor de 0, cuando todos los valores son iguales y, por lo tanto, todas las diferencias serán nulas.

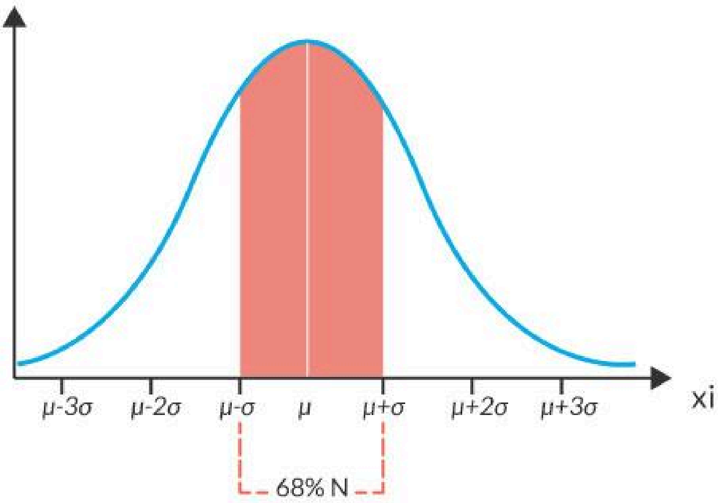
La varianza no puede adquirir un valor negativo, ya que la misma está dada por el promedio del cuadrado de las diferencias y, por lo tanto, no podrá tomar en ningún momento valores negativos.

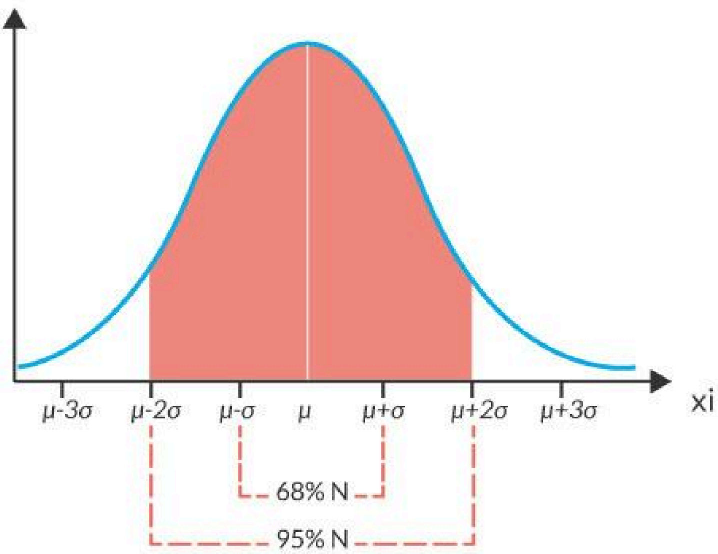
Regla Empírica

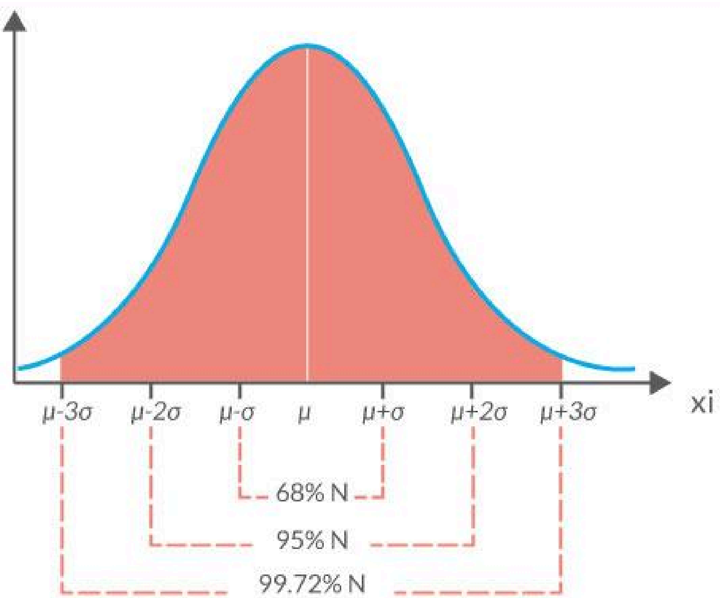
Según lo visto anteriormente, dada una distribución, el valor de su desvío estándar nos muestra el grado de dispersión de sus valores con respecto a la media, pero es en verdad la regla empírica la que relaciona a los dos parámetros: x o μ, y σ con el siguiente enunciado.

En el intervalo centrado en la media y tal que su origen izquierdo esté dado por la media menos un desvío estándar, y el derecho por la media más un desvío estándar, se agrupa el 68 % de los valores de la distribución, en el intervalo comprendido por la media menos dos desvíos estándar y más dos desvíos se ubica el 95 % de los valores de la distribución. Por último, en el intervalo comprendido por la media menos tres desvíos estándar y la media más tres desvíos estándar, se tendrá el 99,7 % de N (se considera en numerosas oportunidades en este intervalo el 100 % de N).

Debe considerar que, entre el inicio de la gráfica y su fin, se encuentran ubicados todos los valores de la distribución N.


μ-3σ
μ-2σ
μ-σ
μ
μ+σ
μ+2σ
μ+3σ
xi
68% N


μ-3σ
μ-2σ
μ-σ
μ
μ+σ
μ+2σ
μ+3σ
xi
68% N
95% N


μ-3σ
μ-2σ
μ-σ
μ
μ+σ
μ+2σ
μ+3σ
xi
68% N
95% N
99.72% N

RESUMEN: REGLA EMPÍRICA

$(\mu \pm 1 \ \sigma)$ se tiene el 68 % de N

$(\mu \pm 2 \sigma)$ se tiene el 95 % de N

$(\mu \pm 3 \ \sigma)$ se concentra el 100 % de N

SP5/Autoevaluación 1

Indique la opción correcta

1- La ________ entre el mayor y menor valor observado de una distribución, es la definida como Alcance o Rango.

- ☐ suma
- ☐ diferencia
- ☐ cociente

Indique la opción correcta

2- Se define como desvío medio en una distribución al promedio de las desviaciones absolutas de las observaciones con respecto a la _______.

- ☐ media
- ☐ moda
- ☐ mediana

Indique la opción correcta

3- Se define como ________ de una distribución, al promedio de los desvíos cuadráticos medios.

- ☐ rango
- ☐ varianza
- ☐ desviación estándar

Indique la opción correcta

4- En el intervalo centrado en la media y tal que su límite inferior está dado por la media menos un desvío estándar, y el derecho por la media más un desvío estándar, se agrupa el __________ de los valores de la distribución.

- ☐ 50 %
- ☐ 45 %
- ☐ 68 %

Indique la opción correcta

5- En el intervalo comprendido por la media menos dos desvíos estándar y más dos desvíos se ubica el ___________ de los valores de la distribución.

- ☐ 99 %
- ☐ 95 %
- ☐ 70 %

Indique la opción correcta

6- En el intervalo comprendido por la media menos tres desvíos estándar y la media más tres desvíos estándar, se tendrá el ______ de N.

- ☐ 99,7 %
- ☐ 95 %
- ☐ 68 %

Indique la opción correcta

7- Se define como desvío estándar en una distribución a la raíz cuadrada de la __________.

- ☐ media
- ☐ varianza
- ☐ frecuencia relativa

Respuestas correctas[12]

[12] 1) Diferencia. 2) Media. 3) Varianza. 4) 68%. 5) 95%. 6) 99,7%. 7) Varianza.

Según la situación profesional planteada observamos que:

Para el caso de madres fumadoras:

$$\text{Media}\ (\bar{x}) = \frac{7627}{11} = 693{,}36$$

$$\text{Desviación estándar}\ (\sigma) = \sqrt{10787{,}1} = 103{,}86$$

Para el caso de no fumadoras:

$$\text{Media}\ (\bar{x}) = \frac{10572}{11} = 961{,}09$$

$$\text{Desviación estándar}\ (\sigma) = \sqrt{12952{,}5} = 113{,}81$$

Volumen de leche	Media	DE
Fumadoras	693,36	103,86
No fumadoras	961,09	113,81

Podemos concluir que, las madres no fumadoras generan en promedio un volumen mayor de leche respecto a las no fumadoras, aunque la dispersión del volumen de leche producido por las no fumadoras es mayor alrededor de su promedio que en el caso de las fumadoras.

SP5/Ejercicio por resolver

La compañía "TODO MODA" cuyo rubro es la comercialización de ropa para adolescentes, tiene a su cargo 40 sucursales. El quinto día de cada mes envía a la casa central los gastos producidos por el consumo de energía eléctrica.

A continuación se presenta la siguiente tabla correspondiente a los gastos mencionados en cada una de las sucursales, multiplicadas por $ 1000.

12	10,2	7,8	6,2
8,5	12,6	14,8	9,5
11,4	9	11,6	7,9
7,6	14,2	12,4	6
9,4	13,4	8,9	8,2
11,2	9,8	13,2	9,2
12,8	10,4	8,6	10
9,6	11,8	11,5	6,8
12,7	11,5	8,5	14
9,5	7,3	7,7	13,5

Determine:

- La media
- La mediana
- La moda
- Desvío Medio
- Varianza
- Desviación estándar

Muestre en qué proporción las sucursales se ubican con sus gastos para:

$(\mu \pm 1 \ \sigma)$ se tiene el 68 % de N

$(\mu \pm 2 \sigma)$ se tiene el 95 % de N

$(\mu \pm 3 \ \sigma)$ se concentra el 100 % de N

Indique la opción correcta

1- Dos distribuciones que tengan la misma cantidad de valores, la misma media y mediana, pueden tener características totalmente distintas en lo que se refiere a la distribución de sus valores.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- El alcance o rango puede ser expresado de la siguiente manera:

$$A = \frac{X_n - X_1}{2}$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- El desvío medio, en algunos casos, toma valores negativos.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- La varianza para una muestra pequeña puede ser expresada de la siguiente manera:

$$Var_{(x)} = \frac{\Sigma\,(X_i - \mu)^2 \quad f_i}{N}$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- La varianza para una población puede ser expresada de la siguiente manera:

$$Var(x) = \frac{\Sigma\,(\overline{X}mi - X)^2 \; f_i}{n - 1}$$

☐ Verdadero

☐ Falso

Indique la opción correcta

6- La desviación estándar para una muestra pequeña, puede ser expresada de la siguiente manera:

$$\sigma = \sqrt{\frac{\left(x_{mi} - \bar{x}\right)^2 f_i}{n-1}}$$

☐ Verdadero

☐ Falso

Indique la opción correcta

7- La varianza puede adquirir un valor negativo.

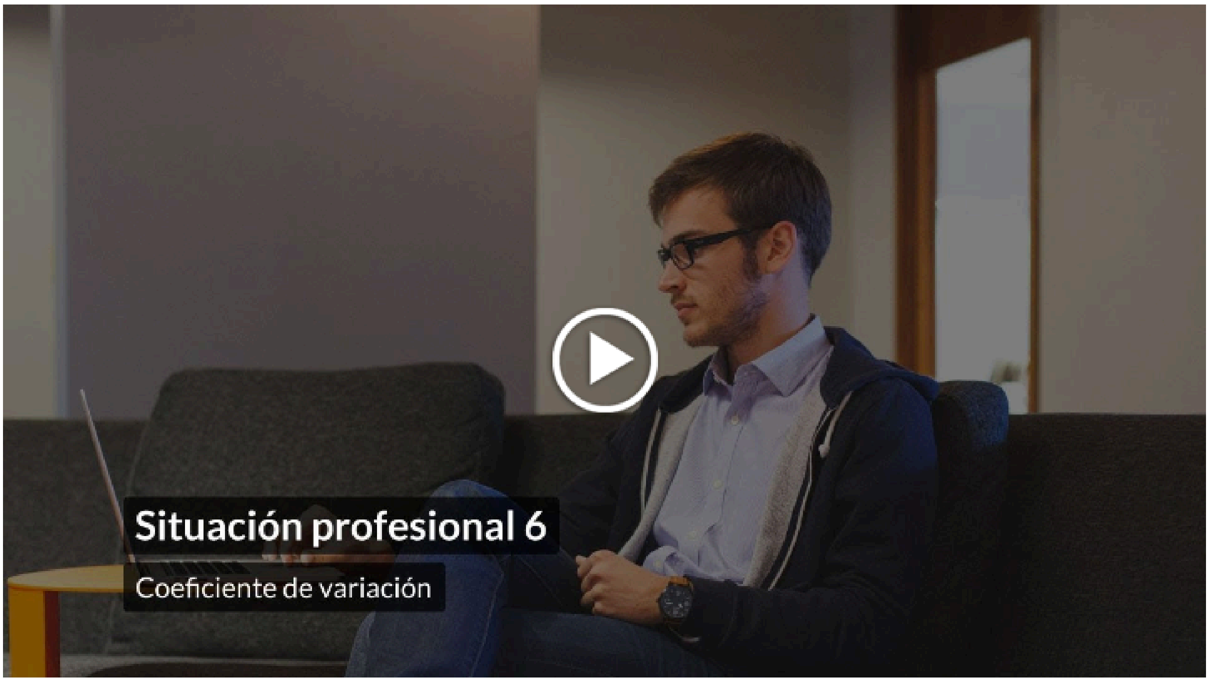
☐ Verdadero

☐ Falso

Respuestas correctas[13]

[13] 1) Verdadero. 2) Falso. 3) Falso. 4) Falso. 5) Falso. 6) Verdadero. 7) Falso.

Situación profesional 6: Coeficiente de variación



Link del video: https://www.youtube.com/watch?v=7v9OOIOHrZ4

Texto del video: Usted se desempeña como ayudante en la Empresa “Nueva Era” y el comité administrativo, con el asesoramiento de los analistas de mercado y gestión contable, le exponen al Gerente General la posibilidad de comprar una o dos empresas familiares. Debido a ello, están investigando y estudiando la administración de las mismas con la finalidad de realizar un negocio con las mínimas posibilidades de riesgo.

Durante los últimos cuatro años, sobre la base de la información obtenida por los analistas, la empresa familiar “La Libertad” tuvo en función de lo invertido, en promedio, una recuperación del 29%, con una desviación estándar del 6,3%.

La empresa familiar “El Vector” tuvo una desviación estándar del 5,8% y una recuperación promedio de lo invertido del 38,8 %.

El Gerente General considera que es riesgoso, en función de la situación económica de Córdoba, comprar una empresa familiar que tenga una elevada dispersión en la recuperación, por lo que le encarga a usted que lo asesore personalmente sobre cuál de las dos empresas familiares en cuestión han estado realizando una estrategia riesgosa.

SP6/H1: Coeficiente de variación

Se define como coeficiente de variación en una distribución al cociente entre su desvío estándar sobre el valor de su media.

$$CV = \frac{\sigma}{\mu}$$

Observación: el coeficiente de variación es adminesional.

De acuerdo a lo analizado anteriormente, el desvío estándar de una distribución se convertía en un coeficiente que nos daba el grado de dispersión de los datos u observaciones respecto de la media. Además, se observó que al comparar dos distribuciones del mismo tipo de variable y con la misma media, la distribución de menor desvío estándar mostraría una menor dispersión con respecto a la de mayor dispersión.

A menor coeficiente de variación consideraremos que la distribución de la variable medida es más homogénea.

Pero frente a la necesidad de analizar y comparar la variabilidad en distribuciones de distintas medias, del mismo modo que cuando las distribuciones correspondan a distintas variables, por ejemplo, una de ellas referidas a notas de un curso y la otra al diámetro de los bujes producidos por una fábrica; aparece la obligación de definir un coeficiente relativo y no absoluto como el desvío estándar, un coeficiente que sea adimensional, es decir, sin dimensiones y, por tal motivo, acceder a la comparación en cuanto a la variabilidad de los datos en distribuciones de distintas variables.

Es frecuente que el coeficiente de variación se exprese en porcentajes, de forma tal que su valor esté dado por:

$$CV\% = \frac{\sigma}{\mu} \cdot 100\%$$

Link del video: https://www.youtube.com/watch?v=TD1doGoL53E

Texto del video: tenemos el especialista de un laboratorio “A”, quien realiza 40 análisis diarios, con una desviación estándar de 5. El especialista “B” realiza 160 análisis diarios, con una desviación estándar de 15. Obteniendo los siguientes resultados:

Laboratorio “A” = 5/40 . 100 = 12,5 %

Laboratorio “B” = 15/160 . 100 = 9,4 %

Con lo cual la variabilidad de la primera distribución es mayor que la segunda, aunque en forma leve, la segunda tiene menos dispersión que la primera.

Indique la opción correcta

1- Se define como ______________ en una distribución al cociente entre su desvío estándar sobre el valor de su media.

- ☐ coeficiente de variación
- ☐ varianza
- ☐ desvío medio

Indique la opción correcta

2- El coeficiente de variación que presenta mayor homogeneidad es:

- ☐ CV= 0,1671
- ☐ CV= 0,1646

Indique la opción correcta

3- El coeficiente de variación que presenta mayor homogeneidad es:

- ☐ CV= 3,01%
- ☐ CV= 2,12%

Indique la opción correcta

4- El coeficiente de variación que presenta mayor homogeneidad es:

- ☐ CV= 0,1684
- ☐ CV= 0,1708

Indique la opción correcta

5- El coeficiente de variación que presenta mayor homogeneidad es:

- ☐ $\mu = 6{,}65$ $\sigma = 1{,}11$
- ☐ $\mu = 6{,}34$ $\sigma = 1{,}12$

Indique la opción correcta

6- El coeficiente de variación que presenta mayor homogeneidad es:

- ☐ $\mu = 6{,}21$ $\sigma = 1{,}02$
- ☐ $\mu = 6{,}40$ $\sigma = 1{,}10$

Respuestas correctas[14]

[14] 1) Coeficiente de variación. 2) CV= 0,1646. 3) CV= 2,12%. 4) CV= 0,1684. 5) $\mu = 6{,}65$ $\sigma = 1{,}11$. 6) $\mu = 6{,}21$ $\sigma = 1{,}02$.

Según la situación profesional planteada, observamos lo siguiente:

	La Libertad	El Vector
Promedio	29%	38,8%
Desviación estándar (D.E)	6,3%	5,8%
CV% = CV x 100%	21,72%	14,95%

En conclusión podemos afirmar que de acuerdo a los CV para ambas empresas, la compra de la empresa La Libertad es una inversión de mayor riesgo (debido a su alta dispersión en función de su promedio en recuperación) que la empresa El Vector.

SP6/ Ejercicio por resolver

La muestra de edades de los alumnos que concurren a una de las aulas del Colegio Universitario IES en la carrera Planeamiento Gerencial, se indican con los siguientes datos:

Turno 1	22	30	26	21	23	22	24	27	26	25
Turno 2	26	35	29	30	27	31	33	36	27	30

Si la homogeneidad es un factor pedagógico positivo en el aprendizaje, utilice la medida de variabilidad relativa que le permita determinar cuál de los dos grupos cumple con dicho factor pedagógico.

Indique la opción correcta

1- El desvío estándar de una distribución se constituye en un coeficiente que otorga el grado de dispersión de los datos u observaciones respecto de la media

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Al comparar dos distribuciones del mismo tipo de variable y con la misma medida, la distribución de menor desvío estándar presentaría una mayor dispersión con respecto a la otra distribución.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- El coeficiente de variación puede ser expresado de la siguiente manera:

$$CV = \frac{\sigma}{\mu}$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- El coeficiente de variación puede ser expresado en porcentajes de la siguiente manera:

$$CV\% = \frac{\sigma}{\mu} \cdot 100\%$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

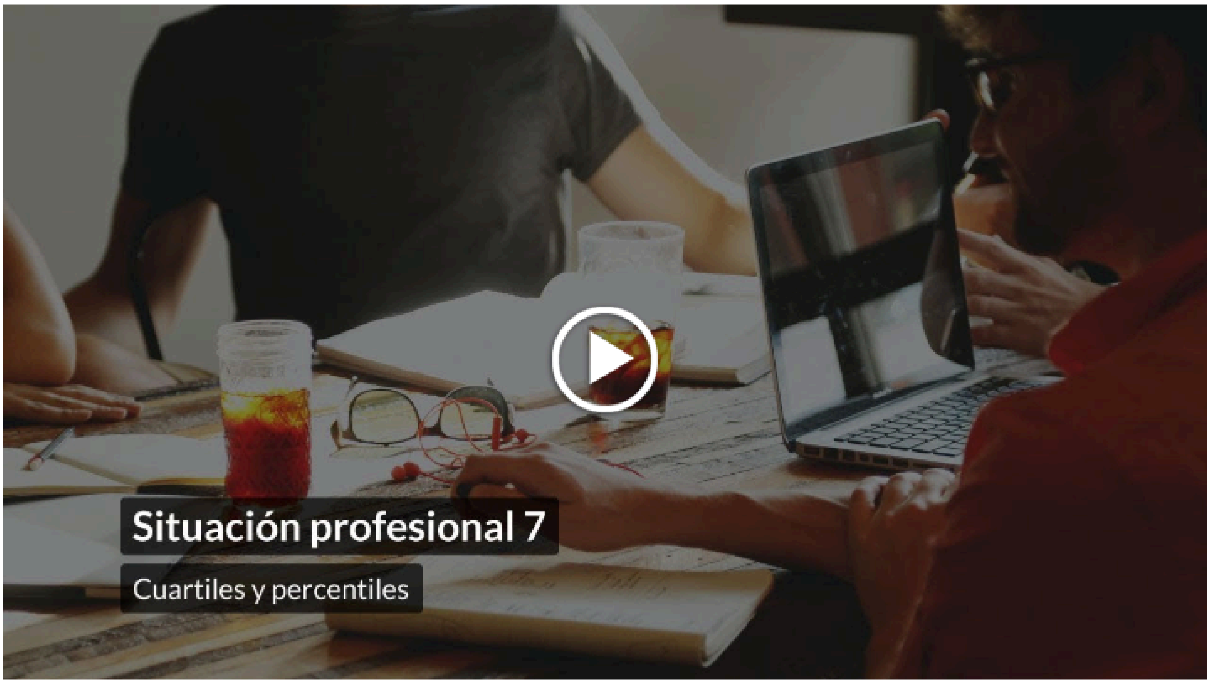
5- El coeficiente de variación es una medida absoluta de la dispersión.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[15]

[15] 1)Verdadero. 2) Falso. 3) Verdadero. 4) Verdadero. 5)Falso.

Situación profesional 7: Cuartiles y percentiles



Link del video: https://www.youtube.com/watch?v=KJwcUpTJdI8

Texto del video: Usted se desempeña como pasante en la empresa "El Ancla" de la ciudad de Córdoba, como asesor directo de la comisión administrativa y, junto con su equipo, diseñó el slogan publicitario para la empresa: "La iniciativa es la madre de las investigaciones ¡SUMATE!".

Entre los informes que le han llegado, existe uno que le provoca una gran preocupación, la cantidad mensual en pesos que la empresa invierte en publicidad debería ser bastante estable, ya que un desequilibrio en la misma produciría fluctuaciones en el presupuesto.

Los datos en miles de pesos son:

200 221 499 163 56 468

632 379 105 42 254 302

Es por ello que decide junto con su equipo de trabajo averiguar el alcance intercuartil y la mediana.

Datos en miles de pesos					
200	221	499	163	56	468
632	379	105	42	254	302

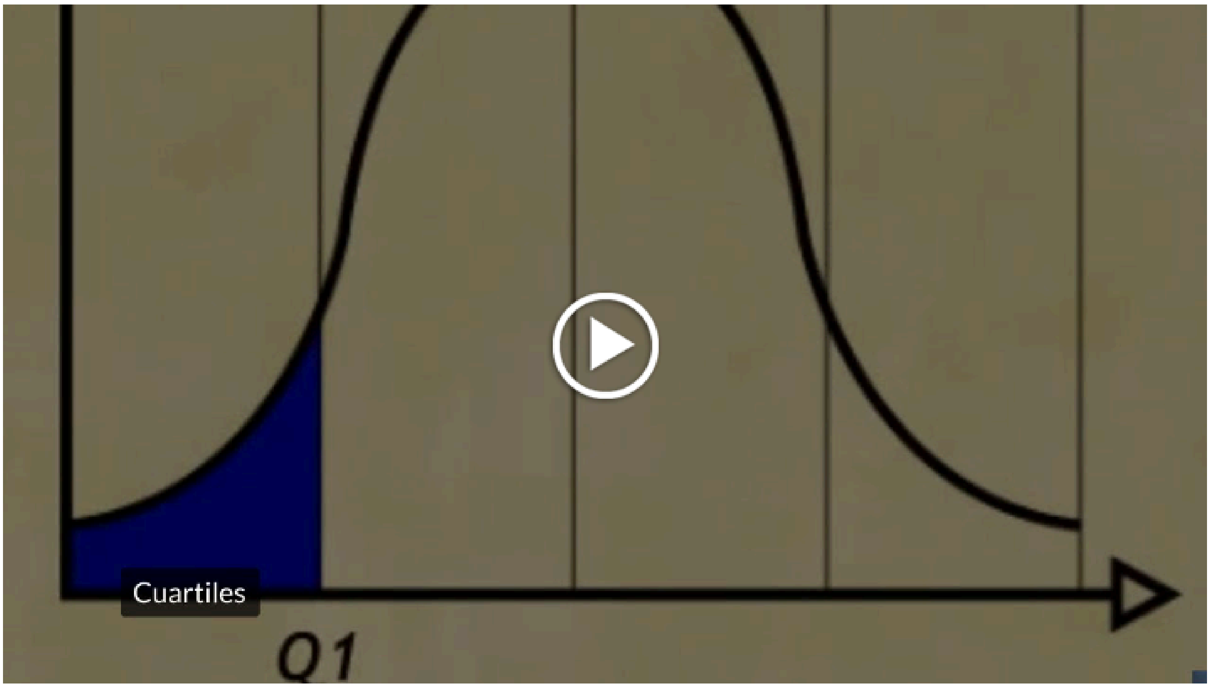
Teniendo en cuenta que el desvío estándar y la varianza son valores que nos muestran el grado de dispersión de una distribución, se tiene, además, dos valores de media que permiten mejorar la información sobre el comportamiento de una serie de valores.

Imaginemos analizar las características de los alumnos de una división a la cual, se la ha evaluado con un determinado examen. Posiblemente no es suficiente el conocer la media de la distribución de notas, sino que se desea saber qué cantidad de alumnos están por debajo del 25 % o, por el contrario, qué cantidad de alumnos de la división tienen calificaciones comprendidas en el 25 % superior de notas, lo cual nos lleva a definir un nuevo tipo de valores denominado CUARTILES.

Cuando nos referimos al histograma de frecuencias, demostramos que el área encerrada por el histograma era igual al número de observaciones. De acuerdo a dicho concepto, entonces, si tomamos un intervalo cualquiera sobre el eje horizontal, el área encerrada por el histograma en ese intervalo será igual al porcentaje de valores que caen en él.

Cuartiles

Los cuartiles dividen al área del histograma en cuatro partes iguales, de forma tal que:



Link del video: https://www.youtube.com/watch?v=UB_yK-CWJwk

Texto del video: 1º) El primer cuartil deja a su izquierda el 25 % de las observaciones y, por lo tanto, a su derecha deja el 75 % de los valores de la distribución.

2º) El segundo cuartil es coincidente con la mediana, ya que deja a la izquierda la misma cantidad de valores que a la derecha.

3º) Por último, el tercer cuartil deja a su izquierda el 75 %.

En el caso que los números de valores no sean lo suficientemente grandes, el establecimiento exacto de los cuartiles puede ser dificultoso, ya que es posible que el valor del cuartil quede entre observaciones; de cualquier forma, por lo general, podemos decir que el establecimiento de cada cuartil queda satisfecho con las expresiones:

$$\text{Primer cuartil } Q1 = 1/4 . N + 3/4$$

$$\text{Segundo cuartil } Q2 = 1/2\, N + 1/2$$

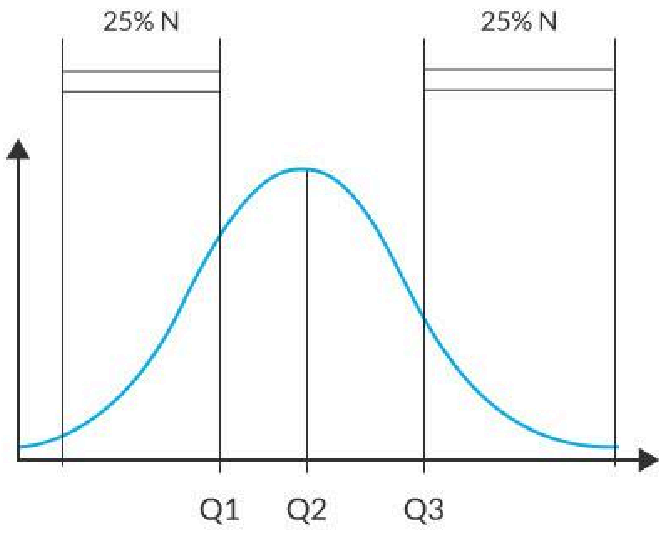
$$\text{Tercer cuartil } Q3 = 3/4\, N + 1/4$$

ALCANCE INTERCUARTIL

Podemos definir como alcance intercuartil a la diferencia entre el tercero y primer cuartil.

$$\text{Alcance intercuartil} = Q3 - Q1$$

Dicho valor nos permite establecer los valores extremos entre los cuales se encuentran incluidos la mitad de los datos observados.



Percentiles

A la distribución en estudio la podemos dividir en cuatro partes; en el caso de hacerlo en 100, cada una de ellas se llamará percentil. Un percentil z en general es un valor tal que z por ciento de las observaciones quedan a su izquierda, mientras que (100 – z) porcientos de los valores quedan a su derecha.

El percentil 70 está dejando a su izquierda el 70 % de los valores de la distribución, se refiere, a todos aquellos menores a él, y, por consiguiente, el 30 % de la distribución que los superan, están a su derecha.

SP7/Autoevaluación 1

Indique la opción correcta

1- Se define como alcance intercuartil a la diferencia entre el __________ y el primer cuartil.

- ☐ cuarto
- ☐ segundo
- ☐ tercero

Indique la opción correcta

2- Los cuartiles dividen el área del histograma en _________ partes iguales.

- ☐ dos
- ☐ cuatro
- ☐ seis

Indique la opción correcta

3- El primer cuartil deja a su izquierda el ______ % de las observaciones y, por lo tanto, a su derecha deja el ______ .% de los valores de la distribución.

- ☐ 25% a su izquierda y 75% a su derecha
- ☐ 75% a su izquierda y 25% a su derecha
- ☐ 80% a su izquierda y 20% a su derecha

Indique la opción correcta

4- El segundo cuartil es coincidente con el/la ______, ya que deja a su izquierda la misma cantidad de valores que a la derecha.

- ☐ moda
- ☐ mediana
- ☐ cuartil

Indique la opción correcta

5- El tercer cuartil deja a su izquierda el ______.

- ☐ 25%
- ☐ 80%
- ☐ 75%

Indique la opción correcta

6- A la distribución en estudio se la puede dividir en _________ partes, en ese caso se lo llama percentil.

- ☐ cuatro
- ☐ cincuenta
- ☐ cien

Respuestas correctas[16]

[16] 1) Tercero. 2) Cuatro. 3) 25% a su izquierda y 75% a su derecha. 4) Mediana. 5) 75%. 6) Cien.

Según la situación profesional planteada observamos lo siguiente:

Tabla 1: Medidas de Resumen para las inversiones mensuales

0,94	0,58	1,12	0,68	0,88	0,14	0,70	0,79	0,24	0,53
0,52	0,34	0,90	1,25	0,56	0,69	0,56	0,46	0,49	0,12

De la Tabla 1 se puede observar que las inversiones no son muy estables durante el año: el 50% de las inversiones son inferiores a 237, 50, mientras que el 50% central de las inversiones tiene un recorrido de 274 miles de pesos.

SP7/ Ejercicio por resolver

La empresa "TODO VIAJE" que se dedica a la fabricación de ladrillos vistos de distinto tamaño, ha desarrollado un sistema de "Justo a tiempo y servicio al cliente" para lo cual ha elaborado el registro de tiempos, en minutos, para medir lo que tardan sus camiones en descargar los ladrillos en el lugar designado.

0,94	0,58	1,12	0,68	0,88	0,14	0,70	0,79	0,24	0,53
0,52	0,34	0,90	1,25	0,56	0,69	0,56	0,46	0,49	0,12

Calcule el alcance intercuartil y el alcance.

Indique la opción correcta

1- El primer cuartil deja a su izquierda el 25 % de las observaciones.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Los cuartiles dividen al área del histograma en cinco partes iguales.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- El segundo cuartil es coincidente con la mediana, ya que deja a la izquierda la misma cantidad de valores que a la derecha.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- El tercer cuartil deja a su izquierda el 25 % de las observaciones.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- El alcance intercuartil se define como la diferencia entre el tercero y primer cuartil (Q3 – Q1).

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

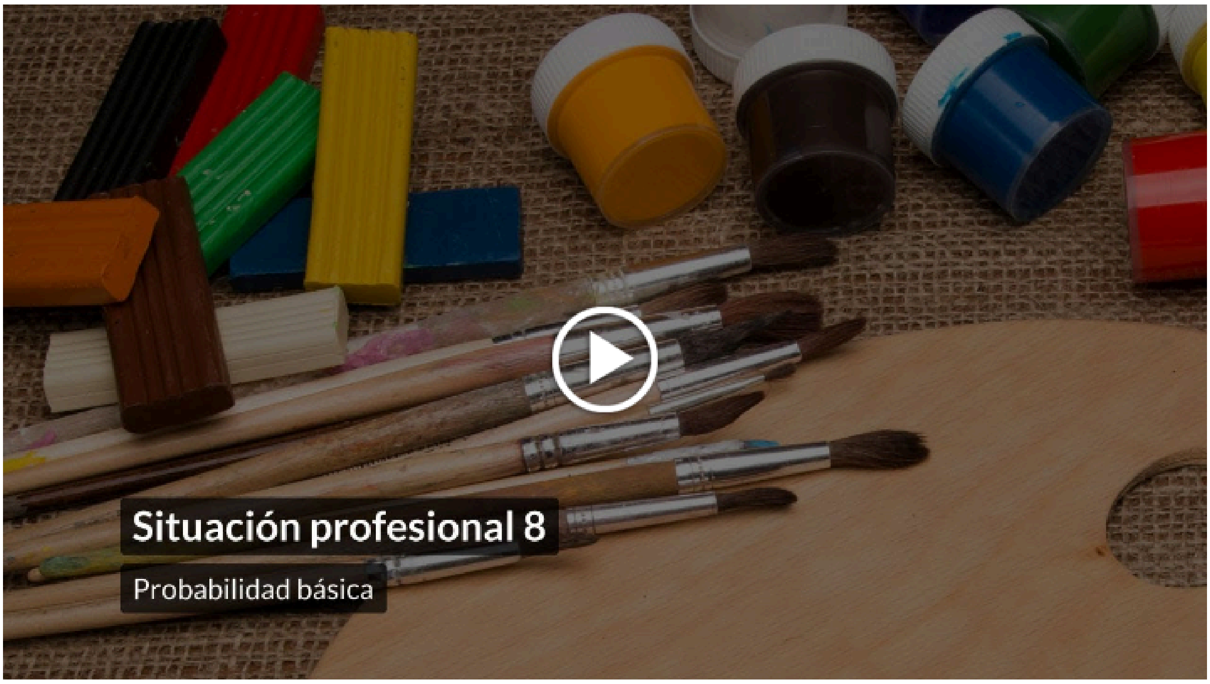
6- Un percentil Z cualquiera es un valor tal que Z por ciento de las observaciones quedan a su izquierda, mientras que (100 – Z) porcientos de los valores quedan a su derecha.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[17]

[17] 1) Verdadero. 2) Falso. 3) Verdadero. 4) Falso. 5) Verdadero. 6) Verdadero.

Situación profesional 8: Probabilidad básica



Link del video: https://www.youtube.com/watch?v=a8hbegTGyBM

Texto del video: La pinturería "Córdoba Color" realiza todos los años un mega evento en la zona del Cerro de las Rosas, donde se invita a personalidades cordobesas, empresarios del rubro, clientes preferenciales y personas relacionadas con la construcción, arte y diseño; allí se hace un festejo por cada aniversario, se anuncian los lanzamientos destacados del año y se realizan importantes sorteos.

Usted integra el grupo de relaciones públicas que se encargará de la organización del evento, y desean estimar la probabilidad de que 2000 o más personas asistan este año.

SP8/H1: Probabilidad objetiva y subjetiva

Estudiaremos varios enfoques de probabilidad básica, que se pueden usar para evaluar la posibilidad de que ocurran diferentes fenómenos.

En las situaciones anteriores, usted adquirió herramientas para organizar, presentar y describir datos de muestras univariables.

Los datos de la muestra generalmente no son usados para estudiar la muestra en sí, sino que son empleados para revisar conocimientos anteriores por parte de la persona que toma decisiones, o bien para generalizar acerca de valores estadísticos de la población.

Los datos de la muestra pueden ayudarnos a reducir la incertidumbre, aunque rara vez pueden eliminarla completamente. Por tanto, la toma de decisiones, con o sin datos de muestra, se asocia con cierto grado de incertidumbre.

La estadística, vista como un método de toma de decisiones, y ante la incertidumbre, se basa en la teoría de las probabilidades, ya que ésta es el lenguaje y la medida de la incertidumbre y los riesgos. Así, el desarrollo de la teoría probabilística llevó consigo como consecuencia la evolución de la estadística, que resultó de utilidad a diversas disciplinas: desde la física hasta las ciencias sociales, desde las ciencias de la salud hasta el control de calidad, y es usada para la toma de decisiones en áreas de negocios e incluso instituciones gubernamentales.

Las probabilidades son muy útiles, ya que pueden servir para desarrollar estrategias. Por ejemplo, algunos automovilistas parecen mostrar una mayor tendencia a aumentar la velocidad si creen que existe un riesgo pequeño de ser multados; los inversionistas estarán más interesados en invertir dinero si las posibilidades de ganar son buenas. El punto central en todos estos casos es la capacidad de cuantificar cuán probable es determinado evento. En concreto decimos que, la probabilidad se utiliza para expresar la posibilidad de ocurrencia de un determinado evento.

Antes de aprender procedimientos estadísticos relacionados a la toma de decisiones, usted debe adquirir conceptos básicos de la teoría de las probabilidades.

Enfoque de probabilidad clásica a priori

Esta teoría es la más antigua y se origina en los juegos de azar. Se basa en el supuesto de que todos los resultados posibles para un experimento aleatorio (actividad que se diseña e implementa de manera que sea imposible predecir con certeza un resultado) son igualmente probables.

Así, empleando el enfoque clásico, la probabilidad de ocurrencia de un evento se calcula dividiendo el número de resultados favorables, entre el número de resultados posibles.

Link del video: https://www.youtube.com/watch?v=RoB271rIITA

Texto del video: Ejemplo 1

Suponga que se lanza un dado perfectamente equilibrado, y se quiere conocer la probabilidad de que salga un dos hacia arriba, llamemos a este evento "hecho A".

En este caso la posibilidad de que ocurra el hecho A se puede definir de la siguiente manera:

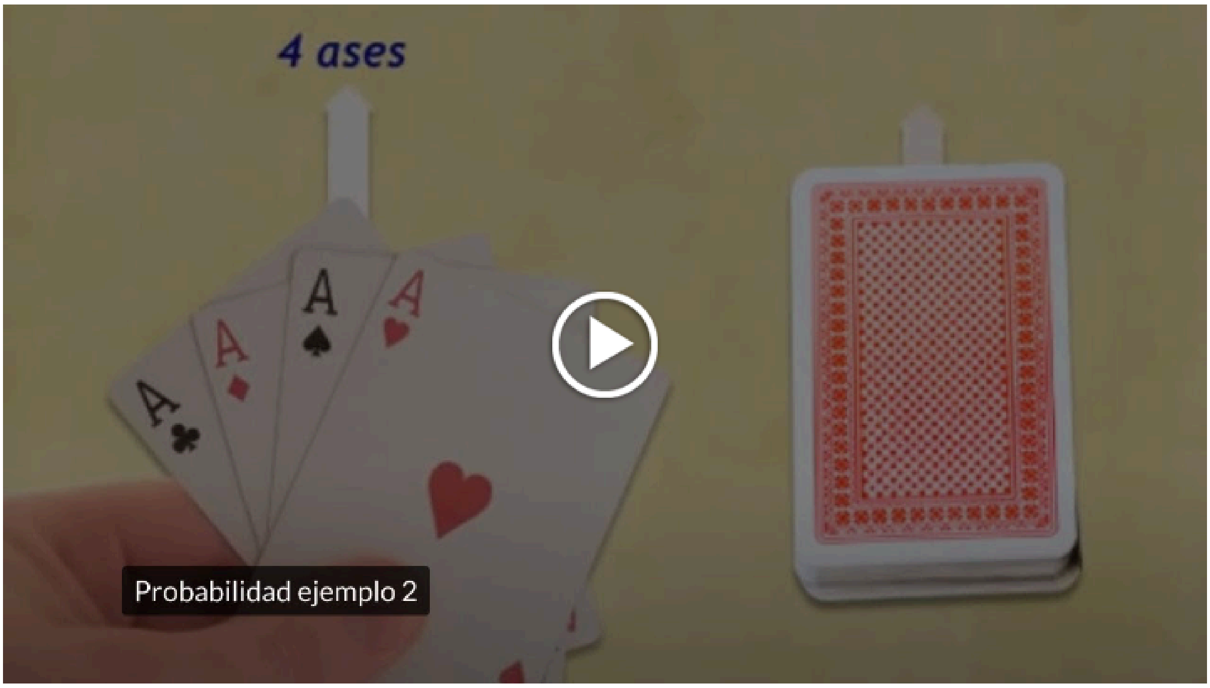
P(A)=n(A)/N(S) Ecuación E-1

Donde:

N(A) es el número de veces que ocurre el hecho que se está observando, y N(S) es el número total de resultados posibles.

Reemplazando los datos del ejemplo 1 en la ecuación E-1:

P(salga un 2)=(número de caras "2")/(número total de caras)=1/6=0,166.



Link del video: https://www.youtube.com/watch?v=DbyPPsIZi8M

Texto del video: Ejemplo 2

En un mazo de cartas bien barajadas donde se tienen 4 ases y 48 cartas de otro tipo (es decir un total de 52 cartas), se quiere calcular la probabilidad de que salga un as. Dicha probabilidad será:

P(salga un as)=4/52= 0,076.

Enfoque de probabilidad clásica empírica o de frecuencia relativa

Según esta teoría, el único procedimiento válido para determinar probabilidades es a partir de la información obtenida realizando repeticiones del experimento de la situación estudiada. No implica ningún supuesto previo de igualdad de probabilidades. A este enfoque se le denomina también enfoque empírico debido a que para determinar los valores de probabilidad se requiere de la observación y de la recopilación de datos. También se le denomina a posteriori, ya que el resultado se obtiene después de realizar el experimento un cierto número de veces.

Volviendo al ejemplo 1, sería necesario realizar lanzamientos del mismo un número de veces (por ejemplo 100 veces) en las mismas condiciones y calcular, después, la proporción de veces que salió la cara seleccionada.

Supongamos que se realiza este experimento de lanzar el dado 100 veces y arroja los siguientes resultados:

Número	Cantidad de observaciones
1	20
2	18
3	15
4	17
5	14
6	16
Total	100

Entonces la proporción 18/100, nos da una estimación de la probabilidad de que salga un 2 cada vez que realizamos el experimento de lanzar un dado. Un momento de reflexión nos mostrará que, aún si el dado fuera idealmente perfecto, puede que no obtuviésemos una probabilidad de 1/6. En otras palabras, no podemos obtener la probabilidad verdadera de experimentos repetidos. Sin embargo, si el dado estuviese perfectamente equilibrado, la estimación de la probabilidad de que salga un 2 se acercaría a 1/6 cuando el número de pruebas aumenta.

Este razonamiento conduce a la interpretación en términos de frecuencias relativas: si un experimento se ejecuta n veces en las mismas condiciones y hay x resultados favorables, con $x \leq n$, una estimación de la probabilidad de ese hecho es la razón x/n. Además, la estimación de la probabilidad de un hecho x/n se acerca a un límite (la verdadera probabilidad del hecho), cuando n aumenta hasta aproximarse a infinito, es decir:

$$P(A) = \lim_{n \to \infty} \frac{x}{n} \qquad \text{Ecuación E-2}$$

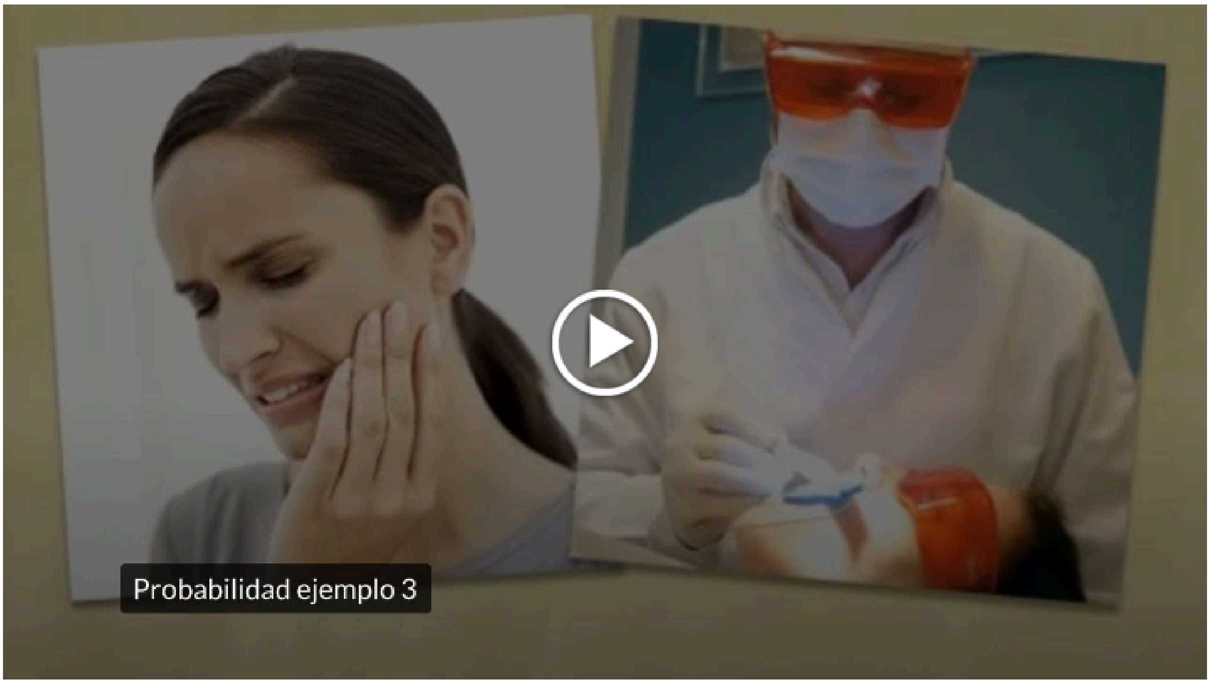
Significa "cuando el número de pruebas aumenta hasta que se aproxime a infinito"

Obviamente, en la práctica nunca podremos obtener la probabilidad de un hecho como dado por este límite: sólo podemos buscar una estimación próxima basada en un n grande.

Por comodidad, trataremos a la estimación de P(A) como si fuera realmente P(A), escribiendo la definición de probabilidad por frecuencia relativa:

$$P(A) = x/n \qquad \text{Ecuación E-3}$$

Definiendo P(A) de esta manera, se destaca que la probabilidad supone un concepto a largo plazo. En la práctica, esto significa que si echamos un dado perfecto 6 veces, es casi imposible esperar que cada una de las seis caras aparezca exactamente una vez. Pero si echamos el dado un número grande de veces, podemos esperar, a largo plazo, que cada una de las seis caras del dado aparezcan la sexta parte de las repeticiones.



Link del video: https://www.youtube.com/watch?v=RuL4R2mR9Is

Texto del video: Ejemplo 3

Antes de incluir la cobertura para ciertos tipos de problemas dentales en pólizas de seguros médicos para adultos con empleo, una compañía de seguros desea determinar la probabilidad de ocurrencia de esa clase de problemas, para que pueda fijarse la prima de seguros de acuerdo con esas cifras. Por ello, un especialista en estadística recopila datos sobre 10.000 adultos que se encuentran en las categorías de edad apropiadas y encuentra que 100 de ellos han experimentado el problema dental específico durante el año anterior. Por ello, la probabilidad de ocurrencia es:

P(A)=100/10.000=0,01.

La teoría clásica y la teoría de frecuencias relativas se llaman enfoques objetivos de probabilidad. La teoría clásica es objetiva porque se basa en un conjunto de supuestos. La de frecuencias relativas es objetiva porque la probabilidad de un hecho es determinada por repetidas observaciones empíricas.

Teoría Subjetiva

Esta teoría se refiere a la posibilidad de ocurrencia de un hecho asignada por una persona en particular. Observemos que, a diferencia del enfoque de frecuencias relativas, hay hechos que son imposibles de repetirse para su estudio y, por tanto, el estudio bajo ese enfoque es imposible. Suponga que se analiza cierta reacción química, nuclear, o un estallido social; situaciones que son únicas, y por tanto, no pueden repetirse bajo las mismas condiciones.

Por ejemplo, ¿quién podría haber predicho el estallido social que provocó la caída del gobierno nacional en 2001? ¿Se podría haber determinado la probabilidad de que la crisis que estalló en esas circunstancias sucediera?

Esto nos da la pauta de que existe una gran cantidad de problemas que están fuera del alcance de las teorías objetivas. Esta limitación, es la que dio lugar al punto de vista personalista sobre la probabilidad. Esta teoría no es la más popular, ya que hay corrientes que afirman que, precisamente por su carácter de subjetividad no se considera con validez científica, aunque en la vida diaria es de las más comunes que se utilizan al no apoyarse más que en el sentido común y los conocimientos previos, y no en resultados estadísticos.

Pero esta teoría no asigna probabilidades de manera "caprichosa" a gusto de quien hace el estudio, sino que se asignan probabilidades basadas en una combinación de experiencia, la opinión personal y el análisis de la situación en particular. Se supone que la persona que asigna dicha probabilidad se basa en su experiencia, conocimiento del tema y, por supuesto que tendrá una influencia de su opinión personal, lo que lo hace de alguna manera subjetivo.

La probabilidad subjetiva es especialmente útil para tomar decisiones en situaciones en las que no se puede determinar empíricamente la probabilidad de la ocurrencia de varios hechos.

Conceptos Básicos de Probabilidad

Para avanzar en el estudio de reglas y operaciones referidas a la probabilidad básica, incorporaremos una serie de conceptos y nos familiarizaremos con terminología que pasaremos a desarrollar a continuación.

Cabe aclarar que un tratamiento adecuado de la teoría de probabilidades requiere cierto nivel de conocimiento de la teoría de conjuntos, por tanto se recomienda al lector una revisión de conceptos y reglas básicas de operaciones con conjuntos.

- **Espacio muestral:** es un conjunto cuyos elementos representan todos los resultados posibles de un experimento aleatorio. Generalmente lo denotamos con la letra S.

Para el lanzamiento de una moneda:

S ={cara, ceca}

Puntos muestrales: son cada uno de los resultados posibles del espacio muestral.

que salga cara

Evento o hecho: es cualquier subconjunto de un espacio muestral.
Por ejemplo:

S ={el día está nublado, el día está soleado, el día está lluvioso}

Un evento posible sería "el día está lluvioso", o "el día está nublado" o "soleado"

- **Complemento de un evento:** incluye todos los elementos del universo que no forman parte del evento. Generalmente lo denotamos con una rayita arriba del nombre, o una letra c como supra índice. Es decir, si A es un evento, su complemento se denota $\bar{A}$ o A^c.

En el ejemplo anterior, si J= {el día está soleado}

J^c ={el día está nublado o el día está lluvioso}

- **Evento simple:** es aquel evento que puede describirse a través de una sola característica.

De la definición anterior, J es un evento simple.

- **Evento compuesto:** es aquel evento que se describe por dos o más características, o sea que puede descomponerse en dos o más eventos simples.

J^c es un evento compuesto.

Esto nos da la pauta de que existe una gran cantidad de problemas que están fuera del alcance de las teorías objetivas. Esta limitación, es la que dio lugar al punto de vista personalista sobre la probabilidad. Esta teoría no es la más popular, ya que hay corrientes que afirman que, precisamente por su carácter de subjetividad no se considera con validez científica, aunque en la vida diaria es de las más comunes que se utilizan al no apoyarse más que en el sentido común y los conocimientos previos, y no en resultados estadísticos.

Cuando se construye una teoría, en este caso de características matemáticas, uno se basa en ciertas ideas o afirmaciones sobre las cuales se construye todo lo demás, como las "reglas de un juego", donde uno conociéndolas, puede jugar y moverse dentro de ese juego. En la matemática y el pensamiento lógico en general también funciona todo como un "juego": se plantean las "reglas básicas" y, a partir de allí, se construye y desarrolla el resto de la teoría. A esas reglas se las llama "axiomas".

Axiomas básicos de probabilidad

- Axioma 1: si A es un evento de S, entonces: $0 \leq P(A) \leq 1$

  A esto se llama, a veces, la ley de no negatividad y afirma que la probabilidad de un hecho en un espacio muestral es no negativa, y también que no excede a 1. Además, la probabilidad es cero cuando el evento está representado por el conjunto sin elementos (el conjunto vacío).

- Axioma 2: sea S un espacio muestral, entonces $P(S)=1$.

  Un espacio muestral S puede considerarse como una "certeza" (un hecho que debe ocurrir en la realización del experimento).

Indique la opción correcta

1- El evento “usa medias rojas o blancas” es un evento simple.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- El complemento de “la comida está lista” es “la comida no está lista”.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- Un evento compuesto puede descomponerse en dos o más eventos simples.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Al realizar el experimento de lanzar un objeto desde cierta altura y medir el tiempo que tarda en llegar al piso, se puede concluir que la probabilidad de que el objeto tarde menos de 10 segundos en caer es 1,5.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- Al realizar el experimento de mezclar pinturas de color azul y amarillo, los posibles eventos que representarán el espacio muestral son 2.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

6- Si se analizan las probabilidades de aprobar o no un examen, según el enfoque clásico, si definimos los eventos A={aprobar el examen}, B={no aprobar el examen}, se tiene que P(A)=0,5.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[18]

[18] 1) Falso. 2) Verdadero. 3) Verdadero. 4) Falso. 5) Falso. 6) Verdadero.

Definimos los siguientes eventos:

A={asisten 2000 o más personas al evento}

B={asisten menos de 2000 personas al evento}

- Según el enfoque de probabilidad clásica a priori, los hechos A y B tienen la misma posibilidad de ocurrir, es decir que se consideran igualmente probables, y en consecuencia, la probabilidad de que asistan 2000 o más personas al evento de la Pinturería, utilizando la ecuación E-1:

  P(A)=1/2=0,5 (es decir un 50% de posibilidades)

- Según el enfoque de probabilidad clásica empírica o de frecuencias relativas, el equipo que organiza el evento podría recurrir a registros de personas que asistieron a este evento en años anteriores, donde se indica que en 8 de los 10 últimos años, asistieron 2000 o más personas. Así, utilizando la ecuación E-3:

  P(A)=8/10=0,8 (es decir un 80% de posibilidades)

- Por último, dado que usted integra este equipo que se encarga de las relaciones públicas de este evento, sobre la base de su conocimiento y experiencia en el tema, y basado también en otros factores que tienen que ver con el reconocimiento de esta empresa y la situación económica del país, podría suponer que existe una probabilidad porcentual de un 70% de que asistan a este evento 2000 o más personas, es decir una probabilidad de 0,7. En este caso, estaría aplicando el enfoque subjetivo.

SP8/ Ejercicio por resolver

Realice el experimento de lanzar una moneda al aire un número grande de veces (digamos 100 o más), registre los resultados obtenidos y compruebe qué relación existe entre la probabilidad de la salida de cara, según las diferentes teorías que serían aplicables a esta situación.

Indique la opción correcta

1- El enfoque clásico a priori se basa en la repetición del experimento observado.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- El enfoque frecuencialista se basa en el supuesto de igualdad de posibilidades.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- El enfoque subjetivo supone la asignación de probabilidades por parte de una persona en particular.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Los puntos muestrales son cada uno de los resultados posibles del espacio muestral.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- Un evento o hecho es un conjunto cuyos elementos representan todos los resultados posibles de un experimento aleatorio. Generalmente lo denotaremos con la letra S.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

6- Un evento compuesto es aquel evento que puede describirse a través de una sola característica.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

7- El Axioma 2 plantea que sea S un espacio muestral, entonces P(S)=1.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

8- La toma de decisiones, con o sin datos de muestra, se asocia con cierto grado de incertidumbre.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

9- La probabilidad se utiliza para expresar la posibilidad de ocurrencia de un determinado evento.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

10- La estadística, vista como un método de toma de decisiones, y ante la incertidumbre, se basa en la teoría de la probabilidad.

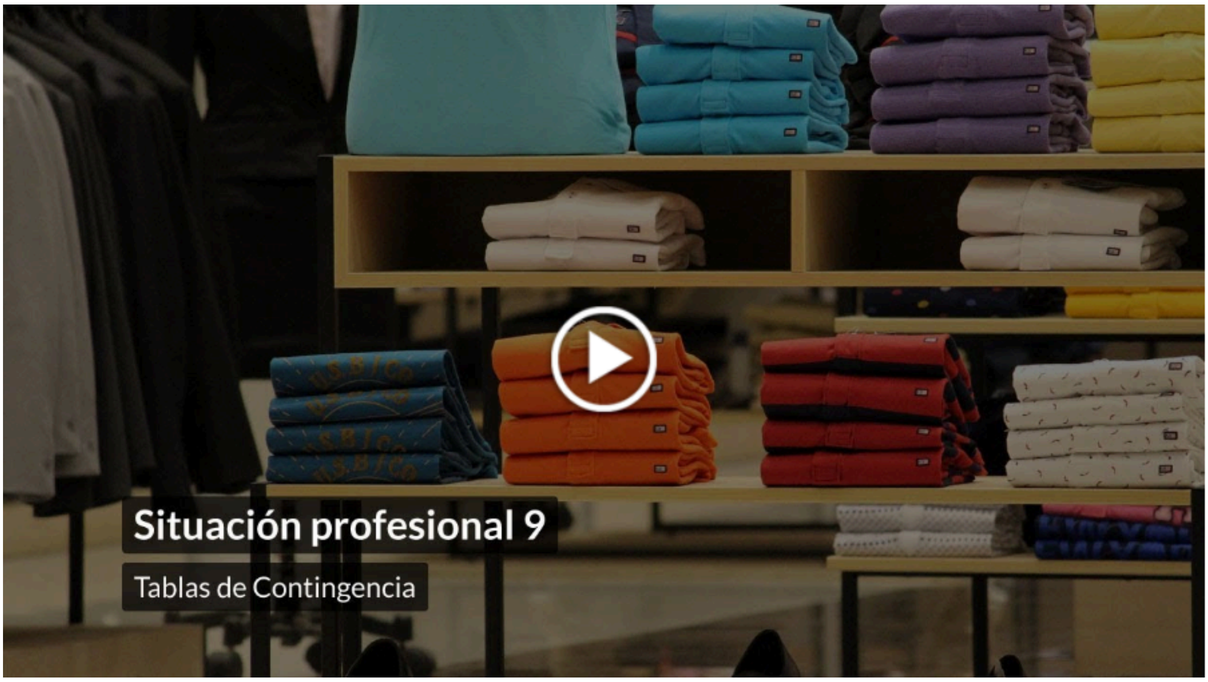
- ☐ Verdadero
- ☐ Falso

Respuestas correctas[19]

---

[19] 1) Falso. 2) Falso. 3) Verdadero. 4) Verdadero. 5) Falso. 6) Falso. 7) Verdadero. 8) Verdadero. 9) Verdadero. 10) Falso.

Situación profesional 9: Tablas de Contingencia



Link del video: https://www.youtube.com/watch?v=z6z9hnpPo8Y

Texto del video: Tienda "Los Hermanos" es una gran tienda de ropa informal, que se encuentra en Río Cuarto y tiene varias sucursales en zonas cercanas. Los propietarios de este negocio contratan a un grupo de profesionales que se encarga de la gestión contable, administración y estrategias para comercializar de manera más efectiva su mercadería, con el propósito de informarse más exactamente acerca de la mercadería que más se vende, a fin de comprar de forma más inteligente. A su vez, en un futuro cercano, desean tener el conocimiento necesario sobre qué productos se deben reforzar con campañas publicitarias convenientes para hacer crecer las ventas.

Para ello, consideran una muestra de 180 clientes: 110 mujeres y 70 hombres.

Se determina que, de las mujeres, 60 compran jeans, 30 compran remeras y camisas, y las restantes compran medias y ropa interior.

Además, se determinó que, en total, son 90 las personas que compraron jeans, 50 compraron remeras y camisas, y el resto ropa interior y/o medias.

Los propietarios de la tienda desean representar la información de alguna manera que facilite el análisis, a fin de poder sacar rápidas conclusiones.

SP9/H1: Tablas de contingencia y diagramas de Venn

Esta situación profesional nos preparará para reconocer diferentes formas de representar y visualizar un espacio muestral dado, con el propósito de obtener información clara y detallada, a los fines de determinar probabilidades, utilizando el enfoque clásico empírico.

Una tabla de contingencia es una tabla de clasificaciones cruzadas, que asigna a cada evento su probabilidad. Se lee luego, como una tabla de doble entrada.

Para la construcción de esta tabla, observe que se tiene una muestra con un total de 180 clientes habituales, de los cuales 110 son mujeres y 70 son hombres. Por lo tanto, en la primera columna de la tabla asignaremos la clasificación mujeres–hombres.

La primera fila se completará con los tipos de compra posible que se registraron en los clientes estudiados, es decir, compran jeans, remeras y camisas, o compran medias y/o ropa interior. La tabla se muestra a continuación:

Clientes	Compran Jeans	Compran remeras y camisas	Compran medias y/o ropa interior	Total
Mujeres	60	30	20	110
Hombres	30	20	20	70
Total	**90**	**50**	**40**	**180**

Dentro de las celdas colocamos los datos del problema; si se sabe que de las mujeres 60 de ellas compraron jeans y 30 compraron remeras y camisas, entonces deducimos que 20 compraron medias y/o ropa interior, ya que se tiene un total de 110 mujeres. Con ello completamos la primera fila.

Además, en total, 90 son las personas que compraron jeans, por tanto deducimos que si son 60 las mujeres que compraron jeans, serán 30 los hombres que compraron jeans; además si se tiene que en total 50 personas compraron camisas y/o remeras, y se tenía que eran 30 las mujeres que compraron camisas y remeras, deducimos que son 20 los hombres que compraron remeras y camisas. Por tanto, vamos completando la segunda fila, con un 30 en la primera columna, un 20 en la segunda y en consecuencia, si el total de hombres es 70, se completa la columna con un 20 de los hombres que compraron medias y/o ropa interior.

A su vez, si los totales que se tienen para la última fila son: 90 que compraron jeans y 50 que compraron remeras y camisas, para tener el total de la muestra de 180 personas, concluimos que son 40 los hombres y mujeres que compraron medias y/o ropa interior. Con ello completamos la última fila.

Usted podría haber deducido de manera distinta los valores que hemos calculado, sin embargo, su procedimiento será igualmente válido si los resultados coinciden con la tabla expuesta.

Otra forma de presentar el espacio muestral es usando un diagrama de Venn. Recordemos que los diagramas de Venn son diagramas que representan de manera gráfica los conjuntos; estos son una herramienta muy útil dentro de la teoría de conjuntos. Posiblemente sea necesario que revise algún texto referido a la representación gráfica de conjuntos para repasar conceptos básicos.

Para utilizar diagramas de Venn, debemos definir antes los posibles eventos simples:

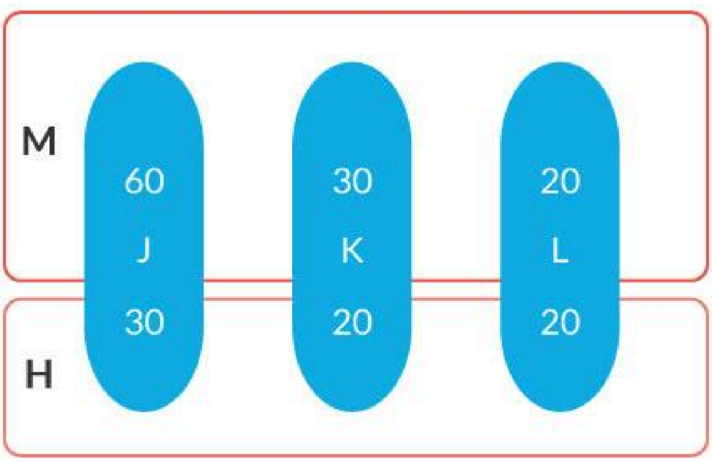
M = {el cliente es mujer}

H = {el cliente es hombre}

J = {el cliente compra jeans}

K = {el cliente compra remeras y camisas}

L = {el cliente compra medias y/o ropa interior}



Observe que:

M y J =60; M y K =30; M y L = 20

H y J =30; H y K =20; H y L = 20

Nótese que las intersecciones de los conjuntos J, K y L con M y H son los respectivos valores de la tabla de contingencia, en las intersecciones de columnas y filas correspondientes, y que la unión entre los conjuntos M y H contiene el total de los clientes muestreados.

Hasta ahora se centró la atención en el significado de probabilidad y en definir y representar eventos y espacios muestrales. Ahora, comenzaremos a desarrollar conceptos referidos a cálculos de probabilidad para diferentes situaciones particulares.

Probabilidad simple (o marginal)

Se refiere a la probabilidad de ocurrencia de un evento simple, como podría ser:

Ejemplo 1: la probabilidad de seleccionar un cliente que sea mujer, P (M) = 110/180 = 0,61 (lo que representa el 61%).

Ejemplo 2: la probabilidad de seleccionar un cliente que sea hombre, P (H) = 70/180 = 0,38 (lo que representa un 38%).

Ejemplo 3: la probabilidad de seleccionar un cliente que compre remeras y camisas, P (K) = 50/180 = 0,27 (lo que representa un 27%).

A la probabilidad simple, también se la denomina probabilidad marginal, puesto que el total de éxitos se puede obtener del margen apropiado de la tabla de contingencia.

Probabilidad conjunta

Lógicamente, si las probabilidades simples se refieren al cálculo de probabilidad de ocurrencia de un evento simple, la probabilidad conjunta se refiere a la probabilidad de eventos conjuntos o compuestos, como puede ser:

Ejemplo 4: la probabilidad de que sea mujer y compre jeans, P (M y J) = 60/180 = 0,33 (33%).

Ejemplo 5: la probabilidad de que sea hombre y compre medias y/o ropa interior, P (H y L) = 20/180 = 0,11 (11%).

Teniendo ahora claro el cálculo de probabilidades simples y conjuntas, una probabilidad marginal de un evento en particular se puede observar en forma alternativa.

Determinaremos por la siguiente forma alternativa, que será justificada a continuación, la probabilidad de que al seleccionar un cliente al azar, éste sea hombre:

P (H) = P (H y J) + P (H y K) + P (H y L)
= 30/180 + 20/180 + 20/180
= 70/180
= 0,38

Observe que esta probabilidad simple, fue calculada a partir de probabilidades conjuntas que tienen una característica especial que será analizada a continuación, y que el resultado obtenido coincide con el calculado anteriormente, realizando el cálculo como probabilidad marginal.

Ahora que demostramos que la probabilidad marginal de un evento consiste en un grupo de probabilidades conjuntas, podemos generalizar diciendo:

$$P(A) = P(A \text{ y } B1) + P(A \text{ y } B2) + \ldots + P(A \text{ y } Bk) \qquad \text{ecuación P-1}$$

donde B1, B2,..., Bk son **eventos mutuamente excluyentes**[20] y **colectivamente exhaustivos**[21] .

[20] Eventos mutuamente excluyentes: Son aquellos que al acontecer excluyen la posibilidad de acontecimiento de los otros; en términos de conjuntos, esto significa que las intersecciones entre ellos, tomados de dos en dos, son vacías.

[21] Eventos colectivamente exhaustivos: Son aquellos que, al menos uno de ellos tiene que acontecer en la realización de un experimento aleatorio, y su unión en términos de conjuntos da como resultado el espacio muestral.

Ejemplo 6: la probabilidad de seleccionar un cliente que sea mujer:

P (M) = P (mujer y compra jeans) + P (mujer y compra remeras y camisas) + P (mujer y compra medias y/o ropa interior)
= P (M y J) + P (M y K) + P (M y L)
= 60/180 + 30/180 + 20/180
= 110/180
= 0,61.

Regla de la adición

Hasta ahora desarrollamos los medios para determinar probabilidad de un evento A y la probabilidad del evento "A y B". Ahora, avanzando un poco más, deseamos examinar una regla para calcular la probabilidad del evento "A o B", regla que nos permitirá calcular la probabilidad de ocurrencia del evento A, o del evento B, o de ambos (A y B). Recordemos, que en términos de conjuntos, el conjunto "A o B", es la unión conjuntista de los mismos, y mirar un punto de esa unión es mirar un punto que pertenece al conjunto A, o al conjunto B, o que pertenece a ambos.

Ejemplo 7: el evento SEA MUJER O COMPRE JEANS, incluirá tanto a clientes mujeres como a hombres.

Analizando la tabla de contingencia, se puede determinar la probabilidad del evento mencionado de la siguiente manera:

P (M o J) = 60/180 + 30/180 +20/180 +30/180

= 140/180
=0,77 (77%).

El cálculo anterior demuestra que se han considerado los elementos comunes y no comunes a M y J. Esta operación es la que indicábamos anteriormente como la unión conjuntista, o unión de conjuntos. La probabilidad de ocurrencia del evento "A o B" se puede expresar con la siguiente regla, conocida como regla de la adición:

$$P(A \cup B) = P(A) + P(B) - P(A \text{ y } B) \qquad \text{ecuación P-2}$$

Aplicando esta regla general al ejemplo anterior, se obtiene el siguiente resultado:

P (MUJ) = P (M) + P (J) - P (M y J)

= 110/180 + 90/180 - 60/180

= 140/180

= 0,77

La intersección entre M y J debe restarse porque ya se ha incluido dos veces al calcular la probabilidad de M y J, lo que puede verse claramente al observar la tabla de contingencia o diagramas de Venn.

Eventos mutuamente excluyentes

Anteriormente definimos eventos mutuamente excluyentes. Observemos que dados eventos con estas características, y calculada la probabilidad de una unión de eventos, no será necesario restar la probabilidad conjunta, ya que la intersección es un evento que corresponde a un conjunto vacío lo que hace que dicha probabilidad conjunta sea cero; por tanto, basta sumar las probabilidades simples para tener el resultado de la probabilidad de la unión.

Ejemplo 8: volviendo a la situación profesional, supongamos que se quiere conocer la probabilidad de que un cliente compre jeans o compre remeras y camisas.

P (J o K) = P (J U K) = P (J) + P (K) – P (J y K)

=90/180 + 50/180 - 0/180

=0,77

Se entiende que los eventos "el cliente compre jeans" y "el cliente compre remeras y camisas" no pueden ocurrir simultáneamente, por ello como conjunto dan resultado el conjunto vacío, cuya probabilidad es nula.

Como se mencionaba antes, siempre que la probabilidad conjunta sea cero, se considera que los eventos involucrados son mutuamente excluyentes.

Luego, la regla de la adición para eventos mutuamente excluyentes se reduce a:

$$P(AUB) = P(A) + P(B) \qquad \text{ecuación P-3}$$

Eventos colectivamente exhaustivos

Como la expresión lo indica, los eventos colectivamente exhaustivos son aquellos que, colectivamente, recorren todas las posibilidades, o dicho de otra manera, tiene que ocurrir alguno de los eventos. Por tanto, se tendrá que la unión de todos ellos será el espacio muestral completo. Así, si A y B son colectivamente exhaustivos, P(A U B)=1.

Ejemplo 9: considere la probabilidad de seleccionar un cliente y que este sea hombre o mujer; dado que estos eventos son mutuamente excluyentes y usando la ecuación P-3, se tiene que

P(H o M) = P(H U M)= P(H)+P(M) =
70/180 + 110/180
=180/180
=1

Probabilidad Condicional

Uno podría preguntarse cómo encontrar ciertas probabilidades, si se tuviera conocimiento previo de alguna característica; es decir, si se tiene información previa sobre los eventos involucrados.

Cuando se está calculando la probabilidad de un evento A en particular, y se tiene información sobre la ocurrencia de otro evento B, esta probabilidad se conoce como probabilidad condicional P(A/B) (la probabilidad de ocurrencia de A, si se sabe que ocurre B), que se puede calcular de la siguiente forma:

$$P(A/B)= P(A \text{ y } B)/P(B) \qquad \text{ecuación P-4}$$

Ejemplo 10: si se supiera de antemano que el cliente es un hombre, ¿Cuál sería la probabilidad de que compre jeans?

Podríamos utilizar la ecuación P-4, con lo cual tendríamos que:

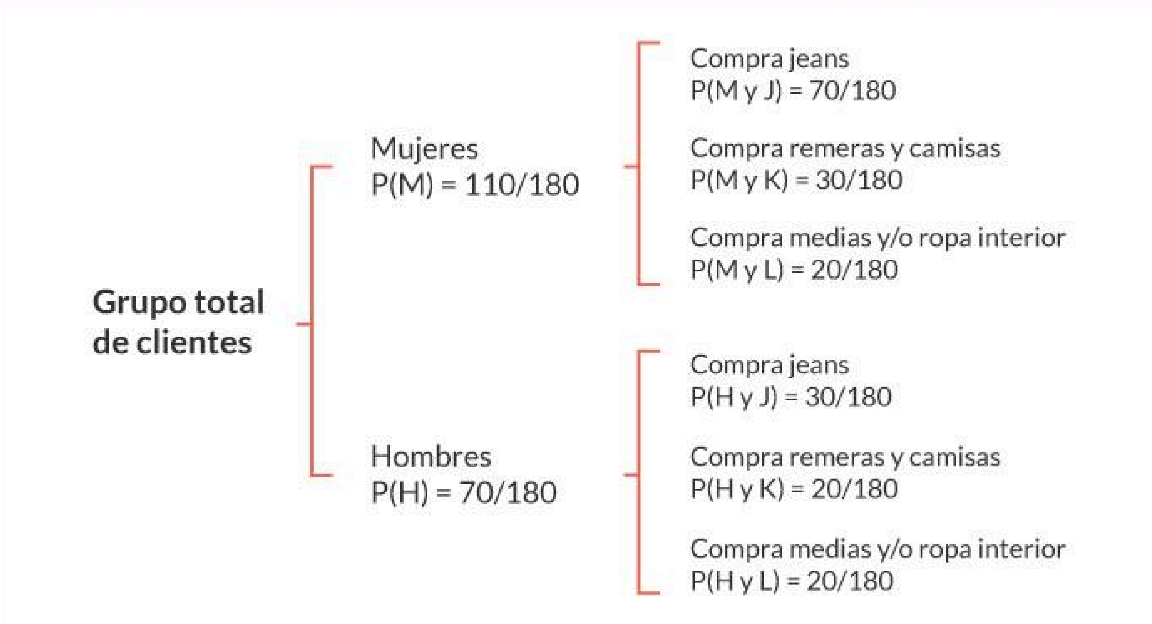
P (J/H)=P(J y H)/P(H)= (30/180)/(70/180)=30/70.

Sin embargo, observe que no es indispensable utilizar la fórmula, ya que podríamos, a través de la tabla de contingencia, determinar los valores. Al tener una "condición", se reduce el total de elementos a estudiar, por tanto, podríamos razonar diciendo que sobre un total de 70 hombres, 30 compran jeans; por lo tanto la probabilidad será P(J/H)=30/70, resultado que coincide con el obtenido usando la fórmula.

Árboles de decisión

Un árbol de decisión es otra manera de organizar y presentar los datos; es un diagrama con ramas y subramas, que resulta una forma alternativa de analizar las probabilidades.

La siguiente figura contiene un árbol de decisión para la información de la tabla de contingencia de la situación profesional:



De la figura se deduce que, al final de las ramas iniciales, se encuentran las probabilidades marginales de "mujer" y "hombre". Mientras que al final de las 3 subramas de cada una de las ramas, se encuentran las probabilidades conjuntas para la combinación entre cliente y tipo de compra.

La probabilidad condicional se puede obtener, simplemente, dividiendo la probabilidad conjunta de interés en la marginal que le corresponde.

Ejemplo 11: para determinar la probabilidad de que un cliente compre jeans, si se sabe de antemano que es hombre, utilizando el árbol de decisión se tiene:

P (J/H) = (30/180)/(70/180)

=30/70.

Independencia Estadística

Para explicar este concepto, comenzaremos presentando un ejemplo que nos ayudará a introducir ideas.

Supongamos que se lanza un dado dos veces. Definiremos los eventos A y B como sigue:

A={el primer dado muestra un número par}

B={el segundo dado muestra un 5 o un 6}

Por intuición, sabemos que los eventos A y B no están relacionados. Saber que B ocurre, por ejemplo, no proporciona información acerca de la ocurrencia de A. De hecho, el siguiente cálculo lo pone de manifiesto. Tomando como nuestro espacio muestral los 36 resultados igualmente posibles (que se dejan como ejercicio al lector), encontraremos que P(A)=18/36=1/2, P(B)=12/36=1/3, mientras que P(A y B)=6/36=1/6.

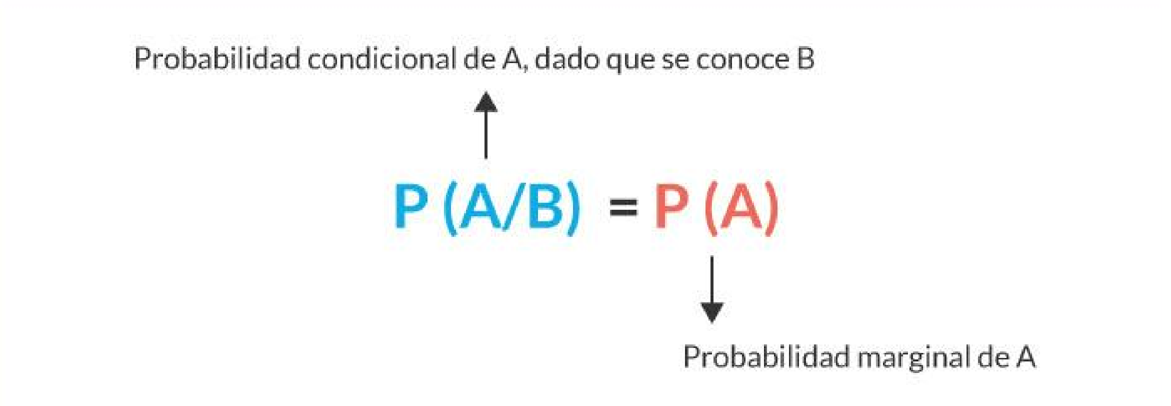
Por lo tanto,

P(A/B) = P(A y B)/P(B) = (1/6)/(1/3) = 1/2

Observe que si no se tuviera la condición, se tiene que P(A)=1/2, lo que coincide con el resultado anterior.

Así encontramos que, como era de suponer, la probabilidad no condicional es igual a la probabilidad condicional. Este resultado revela una información importante: el conocimiento previo de alguna condición no influye en la probabilidad buscada.

A esta característica se la denomina independencia estadística y se puede definir de la siguiente manera:



Luego, diremos que dos eventos A y B son estadísticamente independientes si y sólo si P(A/B)=P(A).

En este caso, "el primer dado muestra un número par" y "el segundo dado muestra un 5 o un 6" son eventos estadísticamente independientes. El conocimiento de uno no afecta ni modifica de manera alguna la probabilidad del segundo.

Ejemplo 12: en referencia a la situación que se planteaba al comienzo de la situación profesional, sobre la tienda "Los Hermanos", quisiéramos determinar si el evento "compra jeans" es independiente del evento "el cliente es hombre". Para ello realizamos los siguientes cálculos:

P(J/H)=30/70, donde este valor representa la probabilidad de que compre jeans, si se sabe que el cliente es hombre.

P(J)=90/180=1/2, donde este valor representa la probabilidad de que compre jeans.

Observe que estos valores son distintos, por tanto podemos afirmar que desde un punto de vista estadístico, estos eventos NO son independientes o, lo que es lo mismo, son estadísticamente dependientes.

La regla de la multiplicación

La fórmula que utilizamos para calcular una probabilidad condicional se puede manipular de manera algebraica, de forma tal que la probabilidad conjunta de A y B se pueda determinar por la probabilidad condicional de uno de los eventos.

De la fórmula P-4, tenemos que:

P(A/B)= P(A y B)/P(B)

Si nuestro objetivo es obtener la probabilidad del evento "A y B", trabajando algebraicamente vemos que:

$$P(A \text{ y } B)= P(A/B).P(B) \qquad \text{ecuación P-5}$$

Ejemplo 13: recordemos el ejemplo que presentábamos en la sección de independencia estadística. Allí considerábamos que se tiraba un dado 2 veces, y definíamos los siguientes eventos:

A={el primer dado muestra un número par}

B={el segundo dado muestra un 5 o un 6}

Donde, además, calculamos que P(A)=1/2, P(B)=1/3 y P(A/B)=1/2.

Recordemos que determinamos que A y B son estadísticamente independientes.

Si quisiéramos determinar la probabilidad conjunta del evento A y B, y usando la ecuación P-5, se obtiene que P(A y B)=P(A/B).P(B)=(1/2).(1/3)=1/6

Por otro lado, observe que P(A).P(B)=1/6. Así, se cumple que P(A y B)=P(A).P(B). Este resultado no es casual, ya que siempre se cumplirá cuando los eventos analizados sean independientes.

Por lo tanto, la regla de la multiplicación para eventos independientes se expresa de la siguiente forma:

P(A y B)=P(A).P(B)ecuación P-6

Por lo tanto, hay dos formas de determinar la independencia estadística:

1. Los eventos A y B son estadísticamente independientes sí y sólo si P(A/B)=P(A)
2. Los eventos A y B son estadísticamente independientes sí y sólo si P(A y B)=P(A).P(B)

Ahora, conociendo la regla de la multiplicación se puede reescribir la ecuación P-1 de la siguiente forma. Si P(A)=P(A y B1)+P(A y B2)+ …+P(A y Bk)

Entonces, aplicando la regla de la multiplicación se tiene:

P(A)=P(A/B1)P(B1)+P(A/B2)P(B2)+ …+P(A/Bk)P(Bk) Ecuación P-7

Donde B, B, …, B son eventos mutuamente excluyentes y colectivamente exhaustivos.

Ejemplo 14: consultando la tabla de contingencia del problema de "La tienda Los Hermanos", se puede calcular la probabilidad de que un cliente compre jeans de la siguiente manera:

P(compre Jeans)= P(compre jeans/mujer).P(mujer)+P(compre jeans/hombre).P(hombre)

=(60/110).(110/180)+(30/70).(70/180)=60/180+30/180=90/180=1/2.

La probabilidad condicional toma en cuenta información sobre la ocurrencia de un evento para predecir la probabilidad de otro. Este concepto puede ampliarse para revisar las probabilidades en base a nueva información, y determinar la probabilidad de que un efecto en particular se deba a una causa específica.

El procedimiento para revisar estas probabilidades se conoce como teorema de Bayes (debido a que originalmente fue desarrollado por el reverendo Tomas Bayes (1702-1761)).

Este teorema también es llamado "fórmula para la probabilidad de las causas".

El teorema de Bayes se puede desarrollar a partir de las definiciones de probabilidad condicional y marginal de la siguiente forma:

$$P(A \text{ y } B) = P(A/B).P(B) \qquad \text{ecuación P-5}$$

Pero también

$$P(A \text{ y } B) = P(B/A) \,.\, P(A) \qquad \text{ecuación P-8}$$

Y a partir de las ecuaciones P-5 y P-8 se tiene:

$$P(B/A) \,.\, P(A) = P(A/B) \,.\, P(B) \qquad \text{ecuación P-9}$$

Por lo que al dividir en P(A) se obtiene:

$$P(B/A) = [\ P(A/B)\ .\ P(B)\ ]\ /\ P(A) \qquad \text{ecuación P-10}$$

Sin embargo, según la ecuación P-7

$$P(A) = P(A/B1)\ P(B1) + P(A/B2)\ P(B2) + \ldots + P(A/Bk)\ P(Bk)$$

Por lo que el teorema de Bayes puede enunciarse de la siguiente manera:

$$P(Bi/A) = \frac{P(A/B_i)\ .\ P(Bi)}{P(A/B_1)\ P(B_1) + P(A/B_2)\ P(B_2) + \ldots + P(A/B_k)\ P(B_k)}$$

Donde Bi es el evento número i de k eventos mutuamente excluyentes.

Indique la opción correcta

1- Una tabla de contingencia es una tabla de clasificaciones cruzadas, que asigna a cada evento su probabilidad.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Los diagramas de Venn son diagramas que representan de manera gráfica los conjuntos.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- A la probabilidad simple, también se la denomina probabilidad conjunta.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Siempre que la probabilidad conjunta sea cero, se considera que los eventos involucrados son mutuamente excluyentes.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- La unión de todos los elementos colectivamente exhaustivos será el espacio muestral incompleto.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

6- Un árbol de decisión es otra manera de organizar y presentar los datos.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

7- La regla de la multiplicación para eventos independientes se expresa de la siguiente forma: P(A y B)=P(A).P(B) ecuación P-6.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

8- El teorema de Bayes es llamado también "fórmula para la probabilidad de las consecuencias".

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[22]

[22] 1) Verdadero. 2) Verdadero. 3) Falso. 4) Verdadero. 5) Falso. 6) Verdadero. 7) Verdadero. 8) Falso.

SP9/ Ejercicio resuelto

A partir de la organización realizada anteriormente a través de Diagramas de Venn, Tablas de Contingencia y Árbol de Decisión y calculando las probabilidades referidas a esta situación, se desprenden las siguientes conclusiones:

a. Existe mayor probabilidad de que el cliente sea mujer a que sea hombre.

b. Las mujeres tienen mayor probabilidad de comprar jeans, remeras y camisas, y los hombres tiene la misma probabilidad de comprar medias y ropa interior que las mujeres.

c. Desde un punto de vista estadístico, el hecho de que el comprador sea hombre o mujer, no es estadísticamente independiente al tipo de prenda que compra.

d. Además, al analizar el stock, el orden de prioridad es:

1. jeans

2. remeras y camisas

3. medias y/o ropa interior

SP9/ Ejercicio por resolver

Suponga que analiza cierta enfermedad en la piel, y la posible influencia de un factor de riesgo: la cama solar. Para ello se considera una muestra de 300 personas, de las cuales 100 están enfermas y 200 no. De las personas enfermas, 30 estuvieron expuestas a este factor de riesgo, y del total son 100 las personas que estuvieron expuestas a este posible factor de riesgo. En base a estos datos construya una tabla de contingencia y responda:

a. ¿Cuál es la probabilidad de que una persona tenga la enfermedad?

b. Si se sabe que la persona se expuso a cama solar, ¿cuál es la probabilidad de que tenga la enfermedad?

c. ¿Cuál es la probabilidad de que no tenga la enfermedad y se haya expuesto a este factor de riesgo?

Indique la opción correcta

1- Dados los eventos:

☐ Verdadero

☐ Falso

Indique la opción correcta

2- Dados los eventos:

W={el número es múltiplo de 2}

☐ Verdadero

☐ Falso

Indique la opción correcta

3- Dados los eventos:

☐ Verdadero

☐ Falso

Indique la opción correcta

4- Las probabilidades simples se refieren a eventos simples.

☐ Verdadero

☐ Falso

Indique la opción correcta

5- La probabilidad condicional toma en cuenta información sobre la ocurrencia de un evento para predecir la probabilidad de otro.

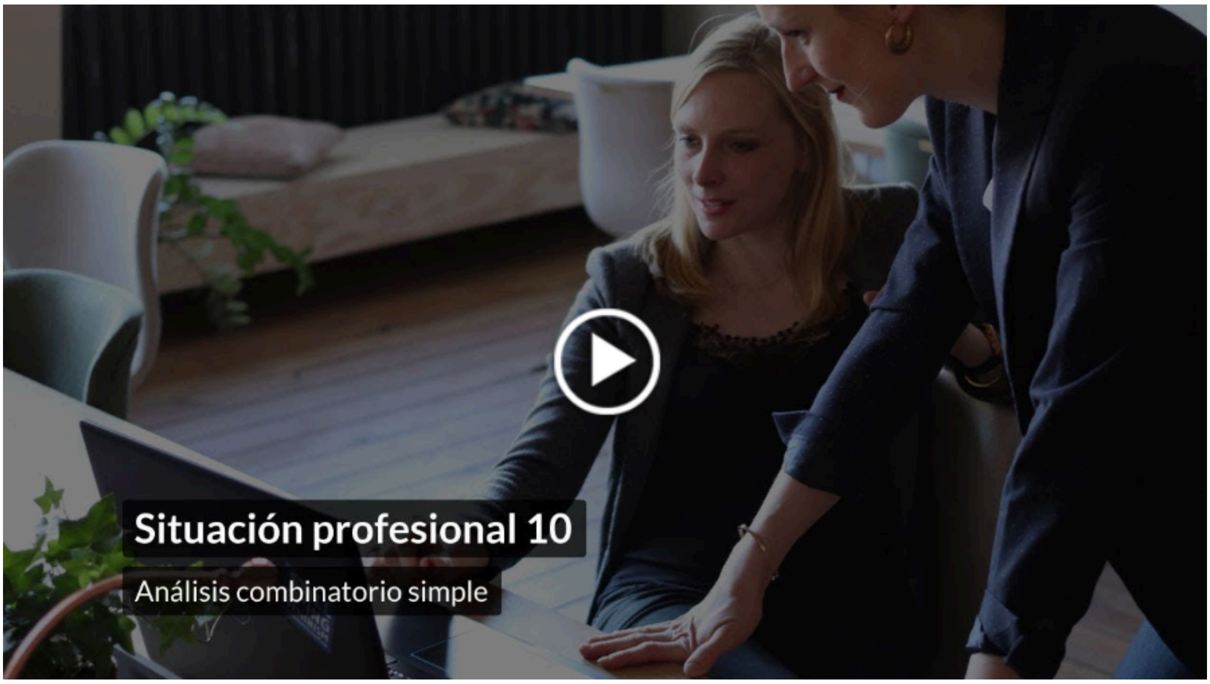
☐ Verdadero

☐ Falso

Respuestas correctas[23]

[23] 1) Falso. 2) Falso. 3) Falso. 4) Verdadero. 5) Verdadero.

Situación profesional 10: Análisis combinatorio simple



Link del video: https://www.youtube.com/watch?v=X6NwKUR9gy8

Texto del video: El Departamento de Recursos Humanos de la Empresa "LA NUEVA ERA" recibe una invitación para el Congreso "El comercio en la era de las Tecnologías de la Información y las Comunicaciones" a desarrollarse en la ciudad de Córdoba, organizada por el área de Relaciones Institucionales del Colegio Universitario IES Siglo 21. Dicha invitación es direccionada al área de comercio internacional, ya que se dirige a un equipo de cuatro profesionales expertos en el tema que deben ser seleccionados del grupo de treinta personas con que cuenta el área.

Además los cuatro profesionales expertos deben ser considerados por una escala jerárquica de acuerdo a sus Curriculum Vitae.

El jefe del área le solicita a usted que le averigüe entre cuántos equipos probables se deben elegir, y de cuántas maneras deben ser los ordenamientos de los treinta profesionales expertos.

Es importante, para su futuro desempeño profesional, tener conocimiento sobre los conceptos de permutaciones, variaciones y combinaciones, de manera tal de contar todas las opciones de subdivisión y ordenamiento de una colección de elementos no repetidos.

Del mismo modo, veremos el caso de permutaciones, donde consideraremos elementos repetidos.

FACTORIAL DE UN NÚMERO

Definición:
Como podemos observar, el factorial de un número natural es el producto del número natural por todos los naturales que lo preceden.

$$0! = 1$$

$$n! = n(n-1)(n-2)\ldots3.2.1$$

Ejemplos:

$$1! = 1$$

$$2! = 2 \,.\, 1$$

$$3! = 3 \,.\, 2 \,.\, 1$$

Principio multiplicativo

Ejemplo:

Link del video: https://www.youtube.com/watch?v=yeELNn7kkF8

Texto del video: Una familia decide comprar una heladera y un lavarropas. Si en el lugar donde harán la compra hay 4 tipos de heladeras y 2 clases de lavarropas ¿de cuántas maneras distintas pueden realizar la compra de ambos objetos a la vez?

Respuesta: $N = 4 \times 2 = 8$

Observación: este principio puede extenderse a más de dos operaciones.

Definición:

Si una operación puede efectuarse de n maneras diferentes y realizada una cualquiera de ellas, una segunda operación puede efectuarse de p maneras distintas, entonces el número total (N) de maneras diferentes en que pueden realizarse, a la vez, ambas operaciones son:

$$N = n \times p$$

Permutaciones simples

Ejemplo:

Link del video: https://www.youtube.com/watch?v=Vh7TVnHo0FA

Texto del video: Una abuela tiene que cuidar 3 nietos ¿de cuántas maneras distintas, nombrándolos uno por uno, puede llamarlos a almorzar?

Respuesta: P3 = 3! = 6

Definición:

Son permutaciones simples, de n elementos distintos, todas las agrupaciones de esos n elementos, dispuestos linealmente, sin que ninguno falte o se repita. Estas agrupaciones se diferencian entre sí, sólo por el orden de sus elementos.

El número de permutaciones simples que pueden realizarse con n elementos distintos son:

$$P_n = n!$$

Permutaciones con repetición

Ejemplo:

Link del video: https://www.youtube.com/watch?v=zveauchJ320

Texto del video: ¿Cuántos números de 6 cifras se pueden formar con los dígitos

1 , 1 , 1 , 2 , 2 y 3?

Respuesta: P6, 3, 2 = 6! = 60

3! X 2!

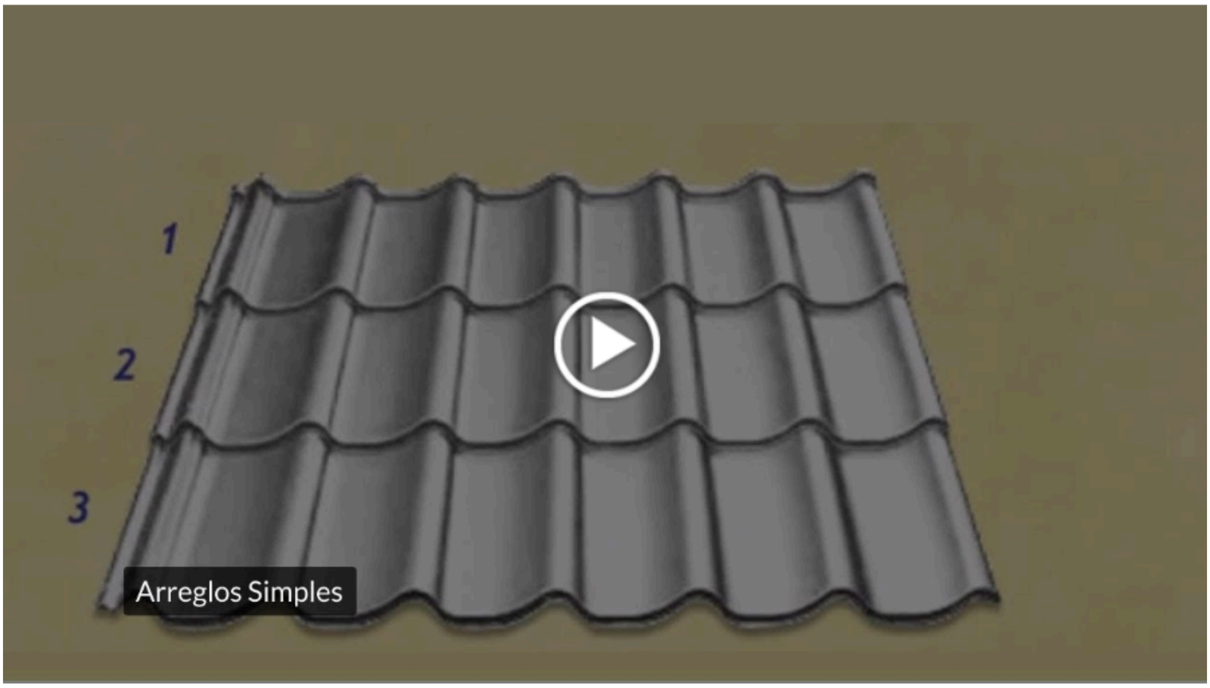
Definición:

Son permutaciones con repetición de n elementos, no todos distintos, todas las agrupaciones de n elementos, formadas por aquellos, dispuestos linealmente y sin que ninguno falte.

El número de permutaciones con repetición que pueden realizarse con n elementos, donde existen r elementos iguales entre sí (de una misma clase) y el resto distintos entre sí y distintos también a los anteriores $(P_{n,r})$, es:

Arreglo (variaciones) simples

Ejemplo:



Link del video: https://www.youtube.com/watch?v=mMnnWoO9nPw

Texto del video: ¿Cuántos techos diferentes, de tres franjas horizontales de igual ancho y de colores distintos, pueden confeccionarse a partir de siete colores diferentes?

Definición:

Son arreglos (o variaciones) simples, todas las agrupaciones de k elementos, dispuestos linealmente, que se pueden formar a partir de n elementos distintos ($k < n$), sin que ninguno se repita. Estas agrupaciones se diferencian entre sí, por los elementos que las componen o por su orden.

El número de variaciones de k elementos que pueden formarse a partir de n elementos distintos (V_k^n) , es:

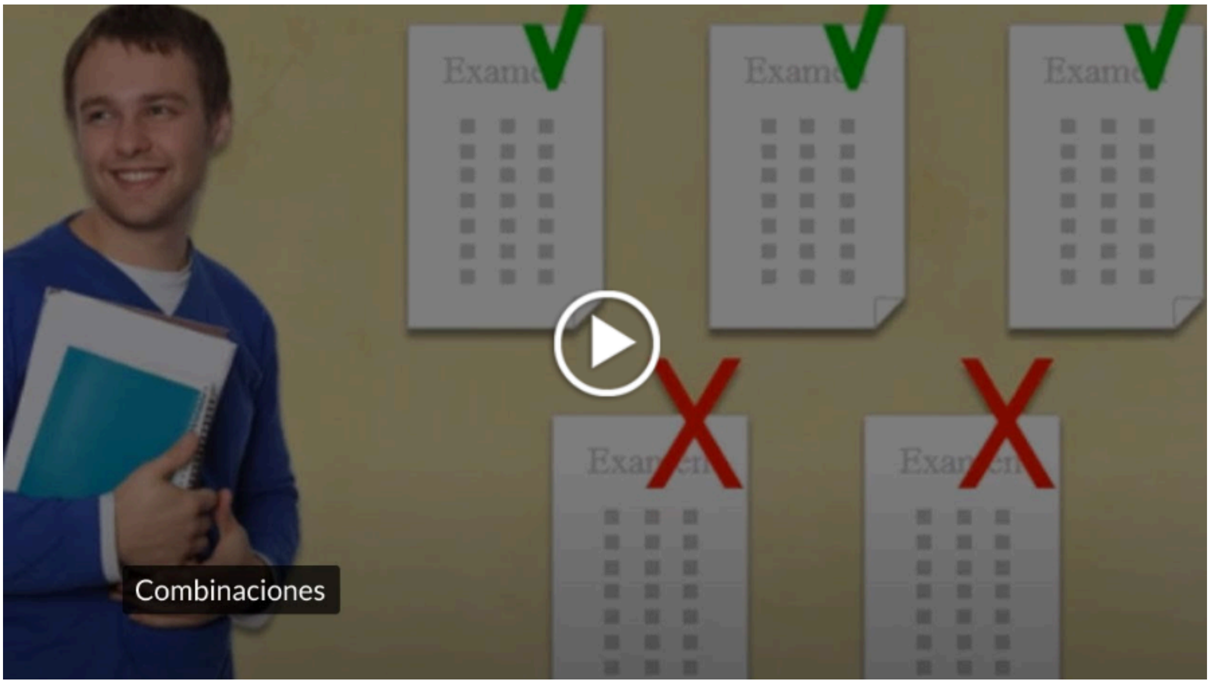
$$\left(V_k^n\right) = \frac{n!}{(n-k)!}$$

En el ejemplo anterior

$$V_3^7 = \frac{7!}{4!} = 210$$

Combinaciones

Ejemplo:



Link del video: https://www.youtube.com/watch?v=7NCwNdWza3M

Texto del video: Un alumno decide rendir tres de cinco exámenes que le corresponden ¿De cuántas maneras distintas puede elegir rendir los tres exámenes?

Definición:

Son combinaciones simples, todas las agrupaciones de k elementos, dispuestos linealmente, que se pueden formar a partir de n elementos distintos (k £ n), sin que ninguno se repita y sin importar el orden de ellos. Estas agrupaciones se diferencian entre sí, sólo por los elementos que las conforman.

El número de combinaciones simples de k elementos, que pueden formarse a partir de n elementos distintos (C_k^n), es:

$$C_k^n = \frac{n!}{k!\,.\,(n-k)\,!}\ ;\ (k \leq n)$$

Indique la opción correcta

1- 0! =

- ☐ 0
- ☐ 1
- ☐ 10

Indique la opción correcta

2- 2! =

- ☐ 2
- ☐ 1
- ☐ 0

Indique la opción correcta

3- 6! =

- ☐ 720
- ☐ 120
- ☐ 520

Indique la opción correcta

4- Las posibles permutaciones de 4 elementos, P4, es igual a:

- ☐ 48
- ☐ 56
- ☐ 24

Indique la opción correcta

5- Las posibles permutaciones de 6 elementos, P6, es igual a:

- ☐ 120
- ☐ 720
- ☐ 520

Indique la opción correcta

6- Las posibles permutaciones de 2 elementos, P2, es igual a:

- ☐ 2
- ☐ 4
- ☐ 1

Respuestas correctas[24]

[24] 1) 1. 2) 2. 3) 720. 4) 24. 5) 720. 6) 2.

Según la situación profesional planteada, observamos lo siguiente:

Las cantidades de equipos probables que se pueden armar son:

1) Si interesa el orden: son 30 variaciones en 4 V:

n! / (n-k)! = 30! / (30 – 4)! = 30! / 26! = 26! x27x28x29x30 / 26! =

27x28x29x30= 657720

2) Si no interesa el orden son 30 combinaciones de 4 bC:

n! / k! (n-k)! = 30! / 4! (30-4)! = 26! x27x28x29x30 / 4! 26! =

27x28x29x30/ 4! = 27405

SP10/ Ejercicio por resolver

La consultora en sistemas "TECNO.NET" cuenta con un grupo de especialistas en software conformada por 7 hombres y 12 mujeres; el Director decide que deben ser especializados en el exterior sobre un nuevo producto, 5 hombres y 10 mujeres. Calcule cuántos grupos pueden formarse respetando el criterio dado por el Jefe.

Indique la opción correcta

1- Se define como factorial de un número natural al producto de dicho número natural por todos los números naturales que lo preceden, cuya expresión es: n ! = n (n-1) (n-2) 3. 2. 1

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- Permutaciones son todas las agrupaciones de k elementos, dispuestos linealmente, que se pueden formar a partir de n elementos distintos (k £ n), sin que ninguno se repita y sin importar el orden de ellos. Estas agrupaciones se diferencian entre sí, sólo por los elementos que las conforman.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- Son combinaciones simples de n elementos distintos todas las agrupaciones de esos n elementos, dispuestos linealmente, sin que ninguno falte o se repita. Estas agrupaciones se diferencian entre sí, sólo por el orden de sus elementos.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Son arreglos o variaciones simples todas las agrupaciones de k elementos, dispuestos linealmente, que se pueden formar a partir de n elementos distintos (k > n), sin que ninguno se repita. Estas agrupaciones se diferencian entre sí por los elementos que las componen o por su orden.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

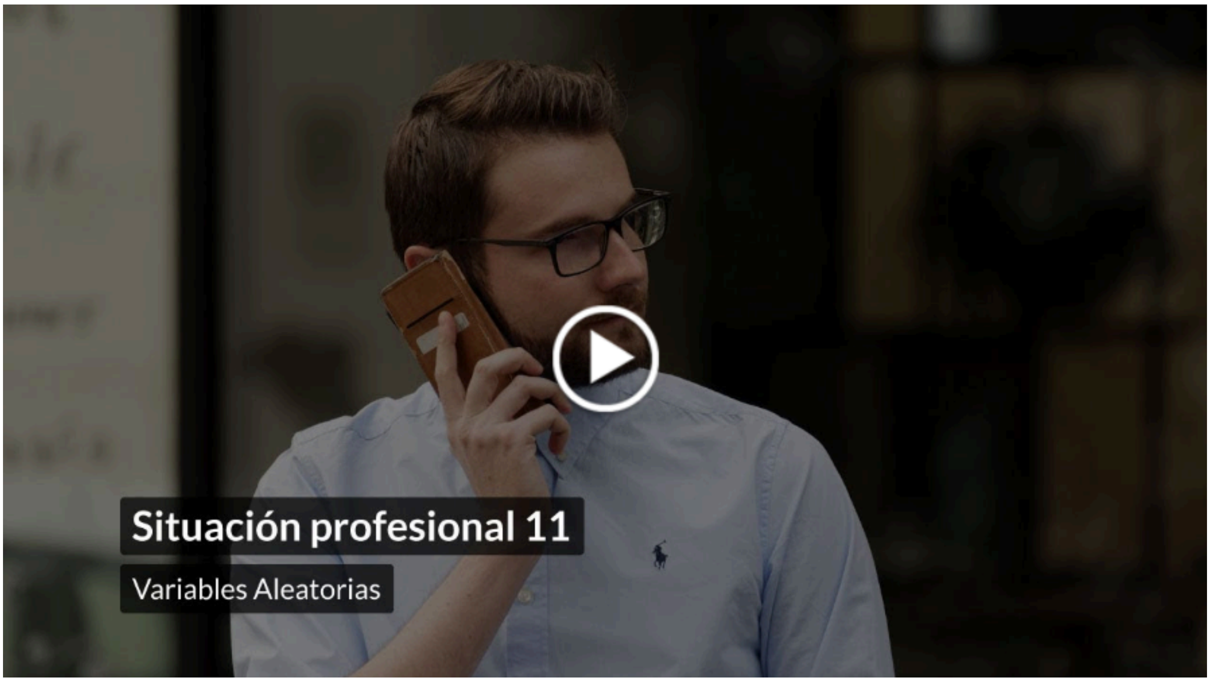
5- Son permutaciones con repetición de n elementos, no todos distintos, todas las agrupaciones de n elementos, formadas por aquellos, dispuestos linealmente y sin que ninguno falte.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[25]

[25] 1) Verdadero. 2) Falso. 3) Falso. 4) Verdadero. 5) Verdadero.

Situación profesional 11: Variables Aleatorias



Link del video: https://www.youtube.com/watch?v=5j6Ie2m9-8M

Texto del video: La cadena de electrodomésticos con sucursales en todo el país "Electro-casa S.A.", lanzó un concurso en el que los participantes deben, para entrar en el mismo, hacer una llamada a un número dado, con un costo de $1,50 la llamada y, a partir de la misma, se genera un cupón, con una capacidad máxima de 10.000 cupones. El día del sorteo, se seleccionará de la urna y al azar, 3 cupones sin reposición: el primer premio es una orden de compra de $3000, el segundo premio es una orden de compra de $2000, y el tercer premio es una orden de compra de $500, para "comprar lo que quieras" en dicho negocio.

Usted está interesado en saber, monetariamente, cuál será el valor esperado de ganar, en caso de participar en este concurso, a fin de decidir si llama a la línea para participar.

SP11/H1: Variables aleatorias: caso discreto y continuo

Hasta ahora, se estudiaron algunas reglas de probabilidad y técnicas de conteo a través del desarrollo del análisis combinatorio.

Uno de nuestros objetivos en esta situación profesional es desarrollar el concepto de esperanza matemática, que veremos más adelante.

Para comenzar, necesitamos saber lo que es una variable aleatoria discreta; definimos variable cuantitativa discreta como el fenómeno de interés, o en estudio, cuyas respuestas o resultados se pueden expresar numéricamente. La clasificación de discreta obedece a que proviene de un proceso de contar. Es decir, que se refiere a un fenómeno en estudio, cuyo resultado es numérico, y específicamente su resultado es un número natural.

Definimos distribución de probabilidad para variable aleatoria discreta como una relación mutuamente excluyente entre todos los resultados numéricos posibles para esa variable aleatoria y sus respectivas probabilidades de ocurrencia.

Link del video: https://www.youtube.com/watch?v=IzQsWNK-Bxw

Texto del video: En la siguiente tabla se representa la distribución teórica de la suma de puntos (que representaremos con X) que se obtiene al arrojar dos dados simultáneamente.

X	2	3	4	5	6	7	8	9	10	11	12	Total
P(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	36/36=1

Por ejemplo, para X=4, las posibilidades son: el primer dado tiene un 1 y el segundo un 3, o los dos dados tienen un 2, o el primer dado tiene un 3 y el segundo un 1. Así, de manera análoga, se construye la siguiente tabla:

X	2	3	4	5	6	7	8	9	10	11	12	Total
P(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	36/36=1

Observe que esta tabla es análoga a una distribución de frecuencias relativas, donde se expresa la probabilidad de ocurrencia en lugar de la frecuencia relativa.

Recuerde que podemos pensar en las distribuciones de probabilidad, como formas teóricas o ideales en el límite de distribuciones de frecuencias relativas cuando el número de observaciones crece y se convierte en un número muy grande, hasta aproximarse al infinito, tal como lo planteamos en el enfoque probabilístico de frecuencias relativas. Es, en parte, por esta razón que podemos pensar en las distribuciones de probabilidad como si fueran distribuciones de poblaciones, y a las distribuciones de frecuencia relativa como distribuciones de muestras de esa población.

Las distribuciones de probabilidad se pueden representar gráficamente como las distribuciones de frecuencia relativa, graficando en ejes coordenados, con los cuales ya nos hemos familiarizado, donde dibujaremos X vs. P(X).

De la misma manera, acumulando probabilidades se obtienen distribuciones de probabilidad acumulada. La función asociada con esta distribución se llama función de distribución de probabilidad. Algunas de estas funciones de distribución de probabilidad que representan fenómenos de interés serán analizadas más adelante.

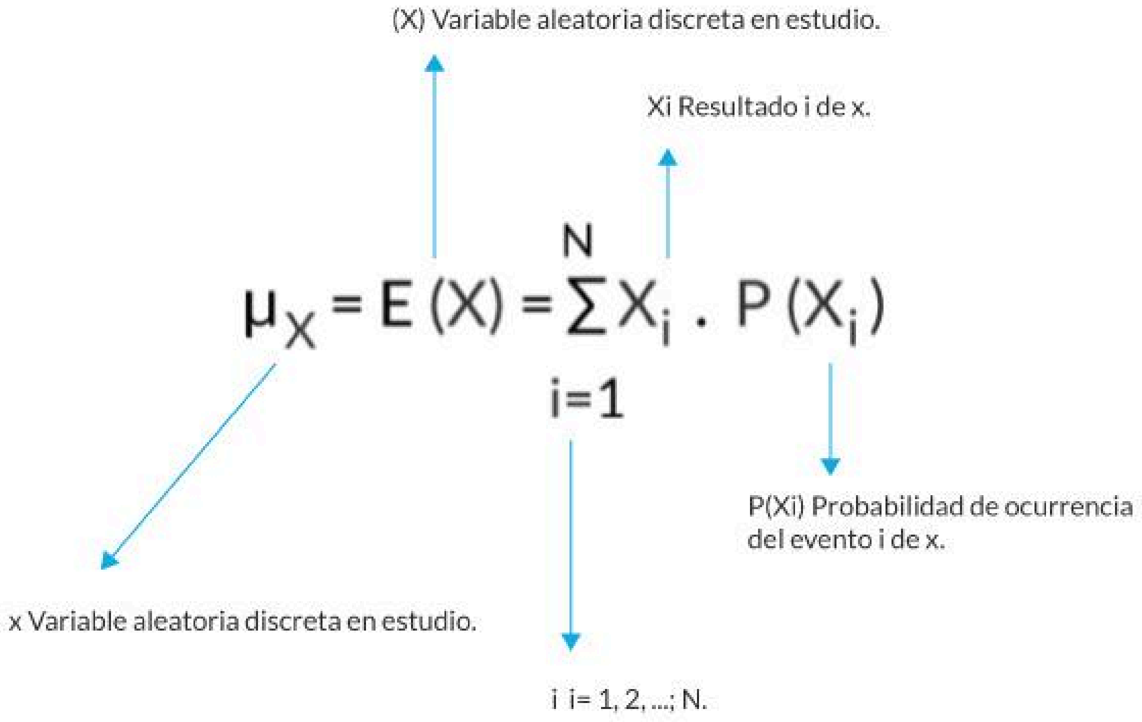
Recordemos que desde el principio de la materia, al analizar las situaciones profesionales planteadas, era de nuestro interés indagar sobre la muestra o población en estudio, y lo hacíamos a partir de valores de tendencia central, y la distribución de los datos o dispersión en los mismos, a través de valores de dispersión.

A continuación, estudiaremos las características principales de una distribución de probabilidad discreta: la media o esperanza matemática, y la desviación estándar.

Valor esperado o esperanza matemática

La media de una distribución de probabilidad (que denotamos μX) es lo que definiremos como esperanza matemática de la variable X, o valor esperado de la variable aleatoria. Este valor esperado de una variable aleatoria discreta se puede considerar como su promedio ponderado sobre todos los resultados posibles, siendo estas ponderaciones la probabilidad relacionada con cada uno de los resultados.

Así, el valor esperado de la variable aleatoria discreta X, representada como E(X), se calcula de la siguiente manera:



Observe que para la distribución de probabilidad del ejemplo de los dados, el valor esperado será:

$$\mu_X = E(X) = 2.(1/36) + 3.(2/36) + 4.(3/36) + 5.(4/36) + 6.(5/36) + 7.(6/36) + 8.(5/36) + 9.(4/36) + 10.(3/36) + 11.(2/36) + 12.(1/36) = 252/36 = 7$$

Note que el valor esperado es un promedio, es decir que este valor conseguido puede interpretarse como que, luego de realizar muchas veces la experiencia de lanzar dos dados y sumar sus resultados, en promedio se conseguirá que la suma es 7.

Suponga la situación del juego de dado ¿cuánto dinero estaría dispuesto a apostar para tener la posibilidad de tirar los dos dados que pagarán, en pesos, la cantidad que apareciera como suma de las caras?

Esto quiere decir que, en una tirada particular se podrían obtener $2, $3, $4, …, $12; y aquí es donde radica un poco la importancia de "el valor esperado". Después de muchos lanzamientos se obtiene entonces que se podría conseguir una utilidad promedio de $7 por lanzamiento.

Si se desea que sea un juego justo, ni jugadores,ni oponentes (en este caso sería "el casino"), deben tener ventaja alguna. Por lo tanto, para jugar se debería estar dispuesto a pagar $7 por lanzamiento. Si por ejemplo el casino decide cobrar $10 por lanzamiento, la esperanza es perder en este juego, a largo plazo, $3 por lanzamiento (7-10=-3). Observe que por lo general, en el casino en cualquier juego de azar, la utilidad esperada a largo plazo para el participante es negativa, de lo contrario, no sería "negocio" para "la casa". Esto puede resultar una desilusión para todos aquellos que sueñan con ganar un juego de azar y hacerse ricos, y aunque es cierto que existe la probabilidad de ganar y podría suceder, debe conocer que la esperanza de ganar en realidad no es positiva.

Varianza y desvío estándar para una variable aleatoria

La varianza de una variable aleatoria discreta, que representaremos como σ_X^2 se define como el promedio ponderado de las discrepancias elevadas al cuadrado, entre cada resultado posible y su media, siendo estas "ponderaciones" la probabilidad relacionada con cada uno de los resultados.

La varianza puede, entonces, expresarse de la siguiente manera:

$$\sigma_X^2 = \sum_{i=1}^{N} (X_i - \mu_x)^2 \, P(X_i)$$

La desviación estándar de una variable aleatoria, que representaremos

con σ_X, se puede expresar de la siguiente manera:

$$\sigma_X = \sqrt{\sum_{i=1}^{N} (X_i - \mu_x)^2 \, P(X_i)}$$

Para la distribución de probabilidad del ejemplo de los dados, la varianza y la desviación estándar resultan:

$$\sigma^2_X = (2-7)^2 1/36 + (3-7)^2 2/36 + (4-7)^2 3/36 + (5-7)^2 4/36 + (6-7)^2 5/36 + (7-7)^2 6/36 + (8-7)^2 5/36 + (9-7)^2 4/36 + (10-7)^2 3/36 + (11-7)^2 2/36 + (12-7)^2 1/36$$

$$\sigma^2_X = 5{,}83 \qquad \text{y} \qquad \sigma_X = 2{,}415$$

En términos del juego, la retribución media es de $7, con una desviación estándar de $2,41.

De acuerdo a la regla empírica estudiada anteriormente, se podría esperar que el 68% (la mayoría) de los pagos recibidos se encuentren en ($\mu x - \sigma x$, $\mu x + \sigma x$). Esto quiere decir que, de cada lanzamiento se debería esperar un pago promedio entre $4,6 y $9,4; y no ganar con mucha frecuencia si el juego costara $10.

Como posibles participantes del juego, la pregunta más importante a esta altura es si este juego realmente resulta o no rentable, más que conocer la probabilidad de que salga cierta combinación de dados.

Para responder a este interrogante, introduciremos un nuevo concepto: el valor monetario esperado, que representaremos con E(V). El valor monetario esperado se define como el promedio que se obtiene al seleccionar una estrategia particular (por ejemplo jugar o no jugar).

En el caso del juego de dados, los valores monetarios esperados oscilan entre -$8,00 y $2, ya que debemos recordar que suponemos que cada lanzamiento cuesta $10. Así, si pagamos $10 para jugar y por ejemplo sale la suma "3", nos pagarán $3, por tanto el valor monetario será de -$7. Con este razonamiento construimos la siguiente tabla:

X	2	3	4	5	6	7	8	9	10	11	12	Total
P(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1
Valor ($)	- 8	- 7	- 6	- 5	- 4	- 3	- 2	- 1	0	1	2	

El valor esperado en caso de jugar será:

E(V)=0,00, lo cual es bastante razonable.

En base a la tabla, tendremos entonces que el valor monetario esperado en caso de jugar será:

$$E\,(V)= (-8).\ 1/36 + (-7).\ 2/36 + (-6).\ 3/36 + (-5).\ 4/36 + (-4).\ 5/36 +$$

$$(-3).\ 6/36 + (-2).\ 5/36 + (-1).\ 4/36 + 0.3/36 + 1.2/36 + 2.1/36 = -3$$

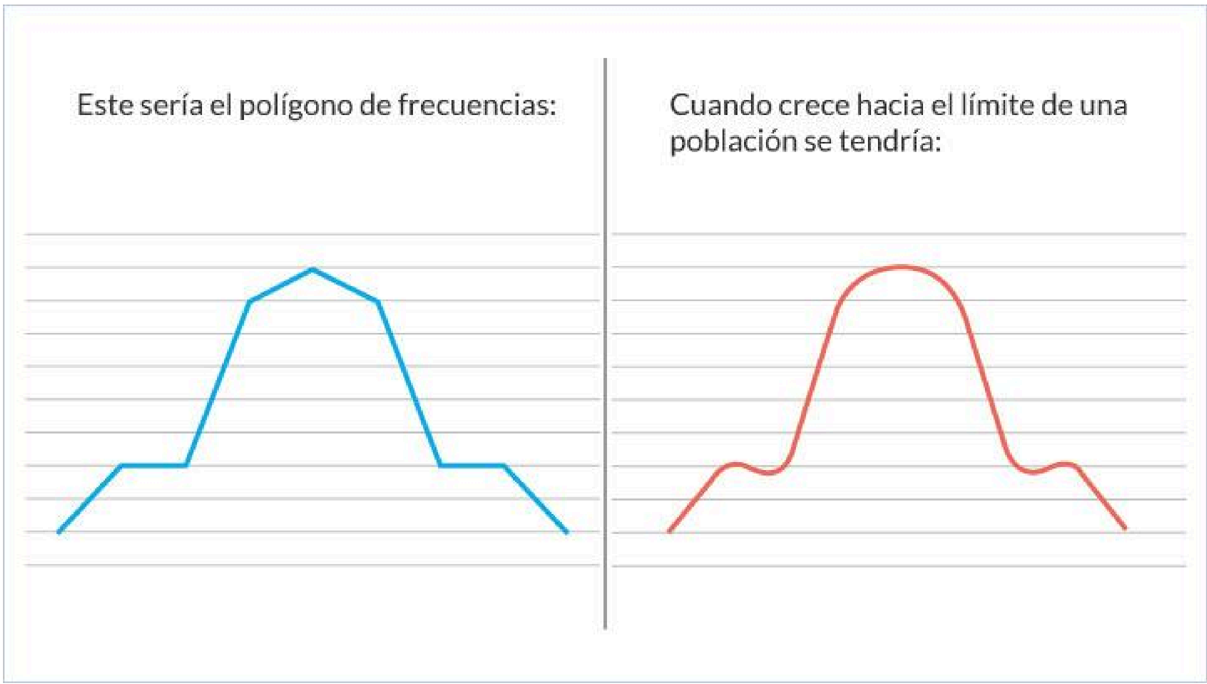
Observe que la remuneración a largo plazo por participar del juego es negativa.

Distribuciones de probabilidad para variables aleatorias continuas

Sabemos que una variable aleatoria continua es aquella variable en estudio que toma valores dentro de un intervalo continuo de números. Las ideas y conceptos introducidos anteriormente para variables discretas, pueden extenderse de manera natural para las variables continuas.

Observe que, si por ejemplo se considera un polígono de frecuencias P(Xi) vs. Xi, éste se convierte en el caso teórico o límite de una población, en una curva continua que puede muchas veces representarse mediante una función llamada función de densidad de probabilidad.

Estos conceptos serán estudiados en profundidad en las próximas situaciones profesionales.



Indique la opción correcta

1- La distribución de probabilidad para variable aleatoria discreta es una relación mutuamente excluyente entre todos los resultados numéricos posibles para esa variable aleatoria y sus respectivas probabilidades de ocurrencia.

☐ Verdadero

☐ Falso

Indique la opción correcta

2- Podemos pensar en las distribuciones de probabilidad como si fueran distribuciones de muestras de población, y a las distribuciones de frecuencia relativa como distribuciones de población.

☐ Verdadero

☐ Falso

Indique la opción correcta

3- El valor esperado de una variable aleatoria discreta se puede considerar como su promedio ponderado sobre todos los resultados posibles.

☐ Verdadero

☐ Falso

Indique la opción correcta

4- La varianza de una variable aleatoria puede expresarse de la siguiente manera:

$$\sigma_X^2 = \sum_{i=1}^{N} i=1N \ (Xi - \mu_X)^2 \ P(Xi)$$

☐ Verdadero

☐ Falso

Indique la opción correcta

5- La variable aleatoria discreta se refiere a un fenómeno en estudios, cuyo resultado no es numérico.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[26]

[26] 1) Verdadero. 2) Falso. 3) Verdadero. 4) Falso. 5) Falso.

SP11/ Ejercicio resuelto

Recordemos que usted está interesado en conocer si resulta o no rentable participar del concurso: "comprá lo que quieras", más que conocer que posibilidades de que salga su número de cupón en el sorteo. Para ello, usaremos las herramientas recién desarrolladas.

En primer lugar, consideramos la distribución de probabilidad de la situación planteada:

Premio ($)	Valor (premio - $1,50)	Probabilidad
3.000	2.998,50	1 / 10.000
2.000	1.998,50	1 / 10.000
500	498,50	1 / 10.000
0 (no ganar)	- 1,50	9.997 / 10.000

Luego, el valor monetario esperado en caso de participar resulta:

$$E\,(V)= 2.998{,}50 \times 1 / 10.000 + 1.998{,}50 \times 1 / 10.000 +$$
$$498{,}50 \times 10.000 - 1{,}50 \times 9.997 / 10.000 = -1{,}149$$

Y el E(V) en caso de NO participar sería E(V)=0.

Por tanto, usted debe saber, que en caso de participar, la remuneración a largo plazo será negativa, es decir que perderá $1.149 por cada cupón con el que participe en el concurso.

¿Sorprendido? Aquí radica el "negocio" de los juegos… el valor esperado para el jugador, a largo plazo siempre resulta negativo, es decir que a la larga, el jugador tiende a perder dinero. Le proponemos un desafío: con las herramientas ya desarrolladas, analice alguno de los juegos que conoce (como la quiniela, la ruleta, etc.) y en base a ello determine la conveniencia o no de participar en los mismos; así podrá usted mismo sacar sus propias conclusiones.

Para ello, debe determinar primero la distribución de probabilidad del juego de interés, asignando las probabilidades a la lista de eventos mutuamente excluyentes y colectivamente exhaustivos.

En la realidad, muchas veces sucede que la asignación de probabilidades se dificulta ya que no se conoce de antemano las probabilidades de ocurrencia de los diferentes eventos; en estos casos se acude a utilizar el enfoque clásico, es decir que se toma el mismo valor para todos los eventos posibles.

En otros casos suelen estimarse las probabilidades de los eventos de diversas maneras, aquí entran en consideración los otros dos enfoques estudiados anteriormente: puede contarse con información de experiencias anteriores que puede utilizarse (enfoque frecuencialista), o bien una persona idónea puede proporcionar probabilidades según el enfoque subjetivo; además en algunos casos podría suceder que la variable discreta estudiada tenga un comportamiento particular y pueda ajustarse bajo alguna distribución en particular, como las que serán analizadas más adelante.

SP11/ Ejercicio por resolver

Una empresa que se dedica a publicidad y marketing para empresas analiza el impacto de una campaña publicitaria lanzada hace 90 días, sobre una nueva marca de bebida gaseosa. Para ello, analizan una muestra de 100 personas que normalmente consumen gaseosas en sus hogares, y se les pregunta cuantas gaseosas de esta marca compraban antes y después de dicha campaña. Los valores registrados fueron:

Cantidad de gaseosas por día	Frecuencia (antes)	Frecuencia (después)
1	2	3
2	3	4
3	12	7
4	15	20
5	15	17
6	13	15
7	8	10
8	10	7
9	14	9
10	8	8
Total	100	100

a. Determine las dos distribuciones de probabilidad: la de antes y la de después de la promoción.
b. Construya histogramas de frecuencias para ambas distribuciones y superponga dichos gráficos para comparar.
c. Calcule las medias en cada caso.
d. Calcule las desviaciones estándar.
e. Responda:

1. ¿Cuál es la probabilidad de que se venda, antes de la promoción, exactamente 6 gaseosas? ¿Y después de la misma?

2. ¿Cuál es la probabilidad de que se vendan, antes de la promoción, al menos 7 gaseosas? ¿Y después de la misma?

3. ¿Cuál es la probabilidad de que se vendan, antes de la promoción, a lo sumo 5 gaseosas? ¿Y después de la misma

Indique la opción correcta

1- La variable aleatoria discreta se refiere a un fenómeno en estudio, cuyo resultado es numérico, y específicamente su resultado es un número natural.

☐ Verdadero
☐ Falso

Indique la opción correcta

2- Las distribuciones de probabilidad son como si fueran distribuciones de muestras de poblaciones, y las distribuciones de frecuencia relativa son como distribuciones de poblaciones.

☐ Verdadero
☐ Falso

Indique la opción correcta

3- La función asociada con las distribuciones de probabilidad acumulada se llama función de distribución de probabilidad equitativa.

☐ Verdadero
☐ Falso

Indique la opción correcta

4- La esperanza matemática de la variable X es lo mismo que el valor esperado de la variable aleatoria.

☐ Verdadero
☐ Falso

Indique la opción correcta

5- La varianza de una variable aleatoria discreta se representa como µx.

☐ Verdadero
☐ Falso

Indique la opción correcta

6- Se define como variable aleatoria discreta a aquella variable en estudio cuyas respuestas o resultados se pueden expresar ____.

☐ Con letras
☐ Numéricamente
☐ Con un cero

Indique la opción correcta

7- Definimos como esperanza matemática o valor esperado a ______ de la distribución de probabilidad.

- ☐ La Clase
- ☐ La Media
- ☐ Un Número

Indique la opción correcta

8- La desviación estándar es _____ de la varianza.

- ☐ La raíz cuadrada
- ☐ La raíz cúbica
- ☐ La potencia

Indique la opción correcta

9- Una variable aleatoria continua es aquella variable en estudio que toma valores dentro de un ______ de números.

- ☐ Conjunto finito
- ☐ Intervalo continuo
- ☐ Intervalo cerrado

Indique la opción correcta

10- El valor monetario esperado es _____ que se obtiene al seleccionar una estrategia particular en una situación concreta.

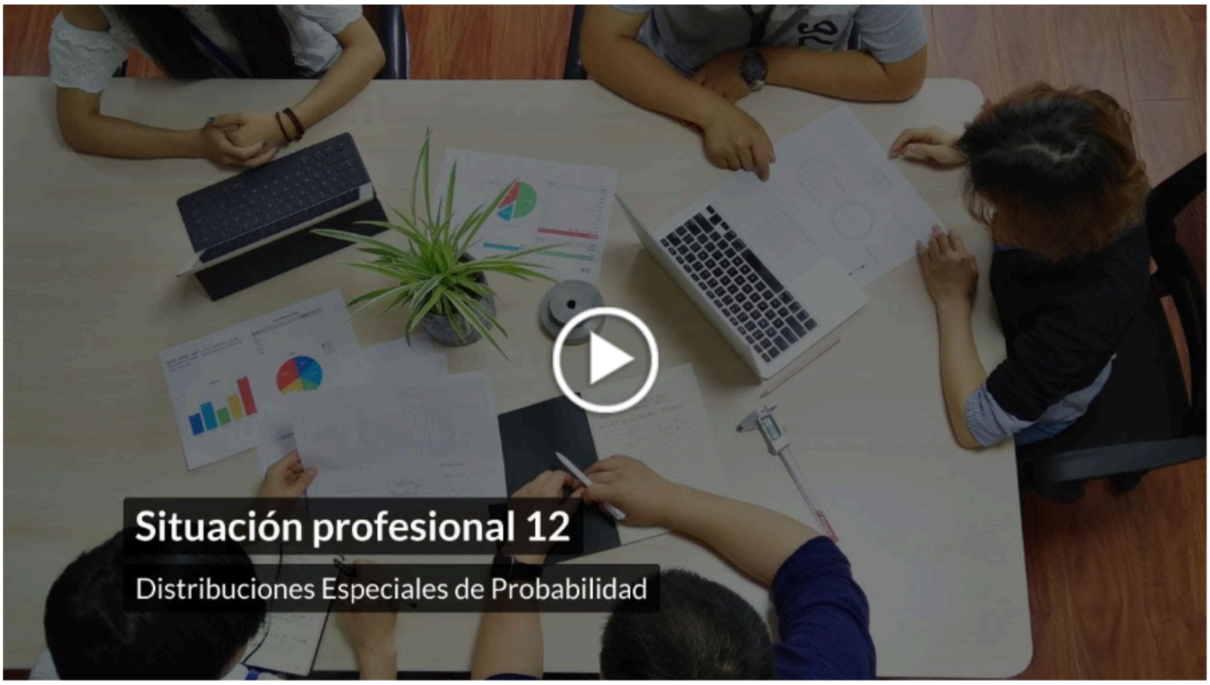
- ☐ El mayor valor posible
- ☐ El menor valor posible
- ☐ El promedio

Respuestas correctas[27]

---

[27] 1) Verdadero. 2) Falso. 3) Falso. 4) Verdadero. 5) Falso. 6) Numéricamente. 7) La Media. 8) La raíz cuadrada. 9) Intervalo continuo. 10) El promedio.

Situación profesional 12: Distribuciones Especiales de Probabilidad



Link del video: https://www.youtube.com/watch?v=CyQNF4RK7YQ

Texto del video: La Subsecretaría de Protección Ambiental de la Municipalidad de Córdoba lanzó un proyecto de "créditos verdes" que consiste en que aquellas personas que pertenecen a su base de datos y que cumplan con ciertos requisitos de crédito, pueden obtener una tarjeta "verde" que es aceptada por los comerciantes del área, de acuerdo a las negociaciones previamente realizadas.

Los informes de meses anteriores indican que el 20 % de todos los solicitantes de dichas tarjetas no son aceptados. Dado que la aceptación o el rechazo es un proceso de Bernoulli, el Subsecretario le solicita a usted como estudiante avanzado, que le averigüe sobre la base de datos, tomando diez solicitudes, lo siguiente:

- ¿Cuál es la probabilidad de que puntualmente dos solicitudes sean rechazadas?
- ¿Cuál es la probabilidad de que puntualmente cuatro sean rechazadas?
- ¿Cuál es la probabilidad de que menos de tres sean rechazadas?
- ¿Cuál es la probabilidad de que más de cinco sean rechazadas?
- ¿Cuál es la probabilidad de que realice 650 ensayos con una probabilidad de éxito de 0,02 y le indique qué distribución de probabilidad binomial, poisson o normal debería ser?

SP12/H1: Distribuciones de probabilidad de variable discreta

Distribuciones de probabilidad

Para el caso de ciertos experimentos aleatorios, los mismos muestran características propias que permiten establecer las probabilidades correspondientes a los diferentes valores que pueden adoptar la variable aleatoria por intermedio de una fórmula o modelo matemático que permite su cálculo.

Tenemos diferentes tipos de modelos, según si la variable aleatoria es discreta o continua y, dentro de cada clase de variables, hallaremos también diferentes modelos conforme con las características del experimento analizado.

Distribuciones de variables discretas

En cuanto a los modelos correspondientes a este tipo de variable aleatoria, estudiaremos las siguientes distribuciones especiales de probabilidad:

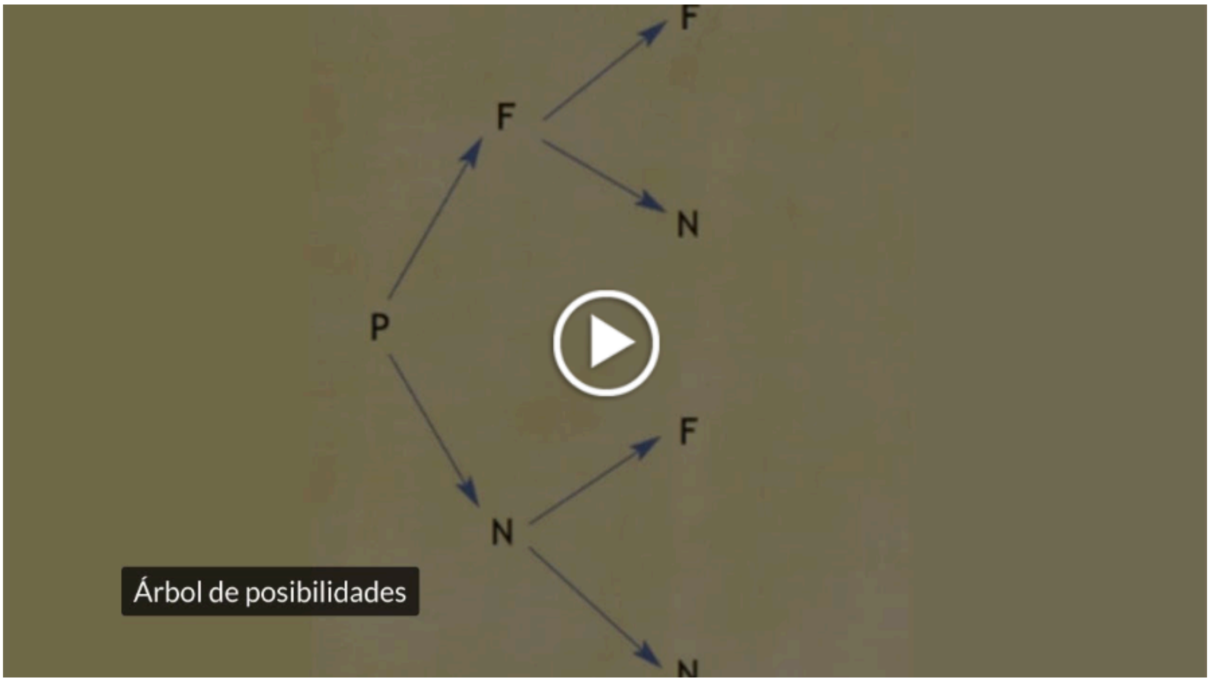
1. **Distribución Binomial**
2. **Distribución Hipergeométrica**
3. **Distribución de Poisson**

1. Distribución Binomial

Al formalizar una encuesta de personas que fuman, se conoce que la probabilidad de que una persona, seleccionada al azar, deje de fumar es de 0,70. Se selecciona al azar una muestra de tres personas y se les pregunta si van a dejar de fumar. Se busca determinar la distribución de probabilidad de la variable aleatoria X = "cantidad de fumadores que van a continuar fumando".

Con los conceptos aprendidos hasta ahora, podemos definir el espacio muestral y establecer la distribución de probabilidad correspondiente a la variable aleatoria en estudio. Luego, analizando esa distribución, la caracterizamos y podremos llegar a establecer un modelo adecuado, para posteriormente generalizarlo.

En el ejemplo presentado, el experimento se refiere a preguntarle a cada persona, si va a continuar fumando. Podemos establecer un árbol de posibilidades y con él, definir el espacio muestral.



Link del video: https://www.youtube.com/watch?v=5HrO7_LR8Iw

Texto del video: Este árbol se establece a partir de la primera persona (P) para el cual existen dos posibilidades: que fume (F) o que no fume (N), a partir de cada una de ellas podemos imaginar a la segunda persona,

que en cada caso dispondrá nuevamente de las dos posibilidades y así sucesivamente, hasta finalizar con la cantidad de personas involucradas.

Una vez construido el árbol, recorriendo cada rama, podemos definir todos los eventos posibles y, de este modo, confeccionar el espacio muestral correspondiente:

$$1^{\circ})\quad S = \{(FFF), (FFN), (FNF), (FNN), (NFF), (NFN), (NNF), (NNN)\}$$

Características del experimento binomial

Considerando el ejemplo, se observa que el experimento se repite en idénticas condiciones.

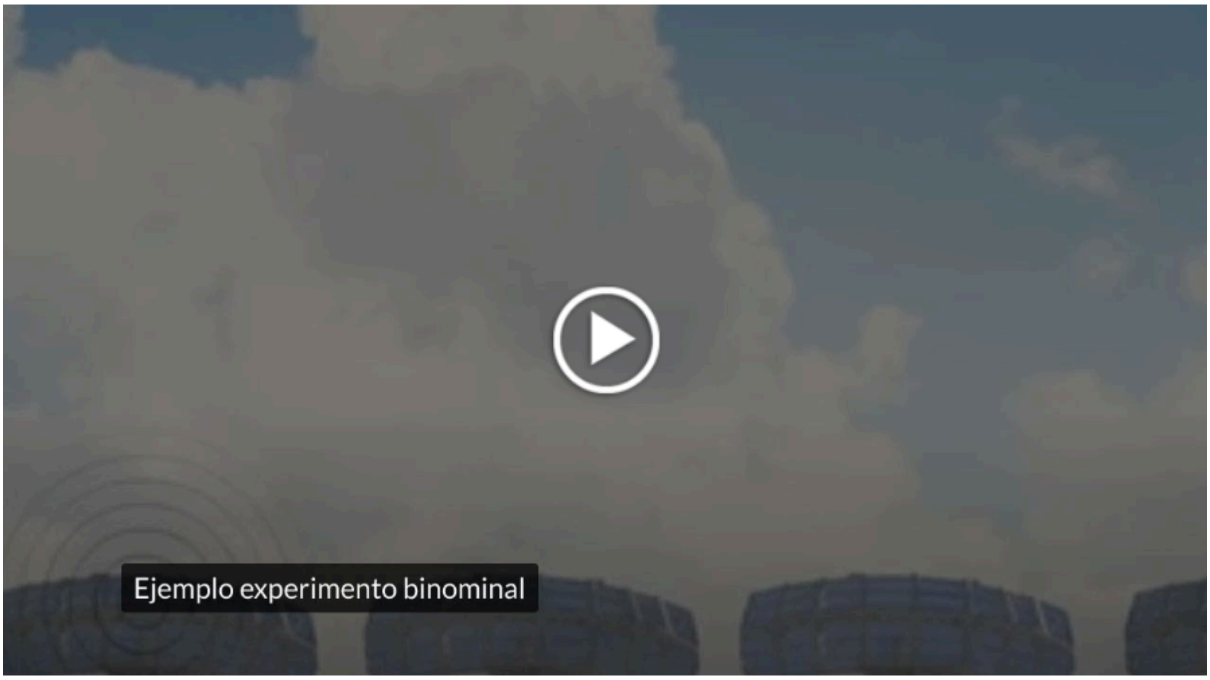
Por cada repetición podemos tener solamente dos resultados, a los cuales denominaremos éxito y fracaso, respectivamente.

La probabilidad de éxito se conserva constante de repetición en repetición. Los resultados obtenidos de cada experimento no se ven alterados por los de los anteriores (son independientes); y para el caso de la variable en estudio, es la cantidad de éxitos observados en las repeticiones.

Si sintetizamos estas observaciones, tendremos las características que debe cumplir un experimento para ser un experimento binomial.

1. El experimento se repite n veces en idénticas condiciones.
2. Cada prueba puede tener únicamente uno de dos resultados que llamaremos éxito o fracaso.
3. La probabilidad de éxito en una prueba es "p" y se mantiene constante durante todo el experimento. La probabilidad de fracaso será $q = 1 - p$.
4. Las pruebas son independientes.
5. La variable aleatoria en estudio x, es el número de éxitos observados en las n pruebas.

Un ensayo con repeticiones, en el cual cada repetición tiene dos posibles resultados, se conoce como Proceso de Bernoulli y es el que se describe a través de la distribución binomial.



Link del video: https://www.youtube.com/watch?v=5meGeVbKQCs

Texto del video: Un sistema de vigilancia y control aéreo, consta de 4 radares idénticos que trabajan independientemente y transmiten su información a una unidad de control central. Supongamos que cada uno de ellos tiene una probabilidad de 0,95 de detectar un avión que se interna en el espacio aéreo controlado. La variable aleatoria de interés es X = el número de unidades que no detectan un avión que se interna en el espacio aéreo controlado. Determine si este experimento cumple con los requisitos para ser considerado binomial.

Analizaremos ahora los requisitos:

1. Consta de 4 ensayos que se repiten en idénticas condiciones (verificar si cada radar no detectó el avión).
2. Cada ensayo tiene únicamente dos resultados posibles: no detecta el avión (éxito), detecta el avión (fracaso).
3. La probabilidad de éxito es p = 0,05.
4. Los eventos son independientes.
5. Se desea saber el número de éxitos (cantidad de radares que no detectan un avión) en las cuatro pruebas.

Luego de este análisis, podemos afirmar que el experimento es binomial.

CÁLCULO DE LA PROBABILIDAD EN UN EXPERIMENTO BINOMIAL

Tomando nuevamente el ejemplo de los fumadores, podemos establecer la probabilidad perteneciente a cada valor de la variable como un término del desarrollo del cubo de un binomio, cuyos términos son p y q (p= 0,70 y q = 0,30):

$$P(x=0) + P(x=1) + P(x=2) + P(x=3) = \sum_{i=0}^{3} P(x_i) = 1$$

$$P(s) = q^3 p^0 + 3\,p\,q^2 + 3\,p^2\,q + q^0\,p^3 = (p+q)^3 = 1$$

En el caso de no haber consultado a tres personas, se hubieran consultado a cuatro, entonces el desarrollo correspondería a la cuarta potencia del binomio p + q. Esto sucede en el ejemplo del sistema de vigilancia y control aéreo, donde los valores que puede asumir la variable aleatoria son 0, 1, 2, 3 y 4, de acuerdo con la cantidad de radares que no detectan el avión.

Podemos observar que, por ejemplo, al valor tres de esta variable le conciernen los eventos en los cuales tres radares no detectan el avión y una unidad sí lo hace, por lo tanto, considerando que el éxito es "que el radar no detecte el avión" tendremos que la probabilidad de x = 3 será :

$$p^3\,q$$

Multiplicado por la cantidad de distintas formas en que puede darse esta situación. En otras palabras, nos podemos preguntar como obtener los coeficientes del desarrollo de la cuarta potencia de un binomio.

Dichos coeficientes son el resultado de las combinaciones de cuatro elementos tomados de tres en tres, es decir:

$$P\,(x = 3, n = 4, p = 0{,}05) \;=\; C^{4}_{3}\, p^{3}\, q^{1}$$

En lo que respecta al cuatro, el mismo corresponde a la cantidad de veces que se repite la prueba y el tres, el valor de la variable aleatoria cuya probabilidad queremos indagar.

Resumiendo, si queremos conocer la probabilidad de un valor x de una variable aleatoria con distribución binomial en un experimento que se repite n veces, y la probabilidad de éxito es p, deberemos aplicar el modelo:

$$P\,(X = x, n, p) \;=\; C^{n}_{x}\, p^{x}\, q^{n-x}$$

Para el caso del ejemplo del sistema de vigilancia y control aéreo (que el radar no detecte el avión), calcularemos las probabilidades:

x	p (X = x)
0	$C^{4}_{0}\ 0{,}05^{0} \,.\, 0{,}95^{4} = \dfrac{4!}{0!\,(4-0)!}\ 0{,}8145 = 0{,}8145$
1	$C^{4}_{1}\ 0{,}05^{1} \,.\, 0{,}95^{3} = \dfrac{4!}{1!\,(4-1)!}\ 0{,}0429 = 0{,}1715$
2	$C^{4}_{2}\ 0{,}05^{2} \,.\, 0{,}95^{2} = \dfrac{4!}{2!\,(4-2)!}\ 2{,}26 \,.\, 10^{-3} = 0{,}0135$
3	$C^{4}_{3}\ 0{,}05^{3} \,.\, 0{,}95^{1} = \dfrac{4!}{3!\,(4-3)!}\ 1{,}1865 \,.\, 10^{-4} = 4{,}75 \,.\, 10^{-4}$
4	$C^{4}_{4}\ 0{,}05^{4} \,.\, 0{,}95^{0} = \dfrac{4!}{4!\,(4-4)!}\ 6{,}25 \,.\, 10^{-5} = 6{,}25^{-6}$

Esperanza y varianza en una distribución binomial

Tomando como punto de partida la definición de esperanza matemática o valor esperado de una variable aleatoria y tomando que las probabilidades para el experimento binomial se obtienen mediante el modelo matemático visto anteriormente, se llega a la siguiente expresión para su cálculo:

$$E(x) = n \,.\, p$$

Deduciendo del mismo modo para la varianza, se obtiene:

$$Var(x) = n \,.\, p \,.\, q$$

Calculemos ahora estos valores para los ejemplos mencionados anteriormente.
Tomando el ejemplo de los fumadores, el valor esperado será:

$$E(x) = 3 \,.\, 0{,}70 = 2{,}1 \;;\; Var(x) = 2{,}1 \,.\, 0{,}30 = 0{,}63$$

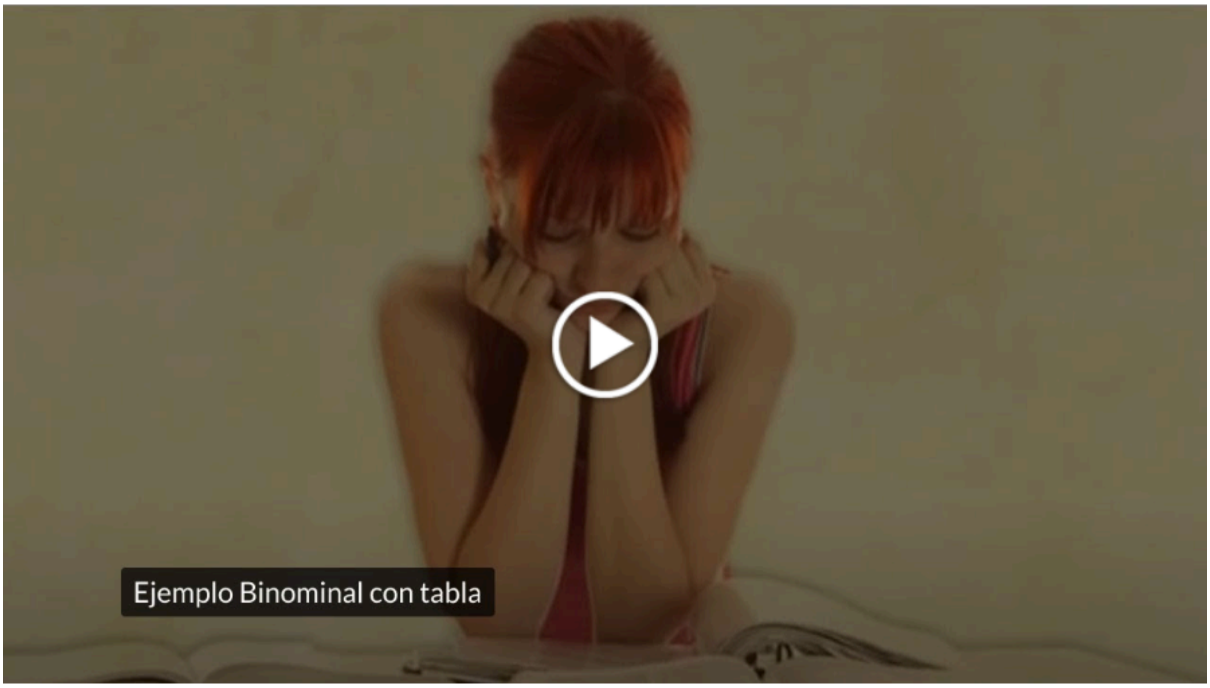
Podemos mencionar que si repetimos el experimento una gran cantidad de veces, en promedio habrá algo más de dos fumadores que continúan fumando y obtendremos una dispersión en los valores de 0,63.

CÁLCULO DE LA PROBABILIDAD EN UN MODELO BINOMIAL MEDIANTE TABLA

Pensemos que necesitamos calcular la probabilidad de que un estudiante responda a lo sumo 6 preguntas correctamente. Responder esto significaría calcular las probabilidades acumuladas desde x = 0 hasta x = 6, es decir P(x = 0) + P(x = 1) + … + P(x = 6), que implica realizar todos los cálculos pertinentes.

Para ayudar a realizar estos cálculos, existen las tablas de probabilidad para los distintos modelos. Nos referiremos ahora del uso de la tabla de función de probabilidad binomial (**anexo 1**), que nos indicará la probabilidad acumulada desde x = 0 hasta el valor de la variable deseado.

Instruiremos el manejo de esta tabla resolviendo el siguiente ejercicio:



Link del video: https://www.youtube.com/watch?v=gqq1RNwMXaQ

Texto del video: Se desea averiguar cuál es la probabilidad que, dado un examen de historia para alumnos del último año de nivel medio con 10 preguntas de verdadero o falso, el alumno conteste (optando al azar) menos de 7 preguntas correctamente (ya que 6 es el mínimo de respuestas correctas para aprobar), expresado en símbolos sería: $P(X \leq 6; n=10; p=0,5)$

En primera instancia buscamos el valor de n correspondiente, en este caso n=10, que se encuentra en la primera columna de la tabla, en la segunda columna, y en el grupo de valores de x posibles para n=10, seleccionamos el correspondiente de x (x=6 en nuestro caso), el valor de p (p=0,5), lo encontramos en la primera fila, en la intersección de la fila (determinada por los valores de n y x) y la columna (definida por el valor de p), encontraremos la probabilidad acumulada hasta x=6. En este caso $P(x\leq$ a $6, n=10; p=0,5) = 0,8281$.

Si necesitamos determinar la probabilidad de que responda al menos 7 preguntas correctamente, para ello, debemos restar a la probabilidad acumulada correspondiente a x=10, el valor acumulado para x=6, es decir:

$$P(x \geq 7, n=10, p=0{,}5) = P(x \leq 10, n=10, p=0{,}5) - P(x \leq 6, n=10, p=0{,}5)$$

$$= 1 - 0{,}8281$$

$$= 0{,}1719$$

Como podemos observar, es el complemento a 1 de la probabilidad obtenida anteriormente.

Si necesitáramos conocer la probabilidad de que contestara exactamente 7 preguntas en forma correcta, lo podríamos hacer de la siguiente manera:

$$P(x=7, n=10, p=0{,}5) = P(x \leq 7, n=10, p=0{,}5) - P(x \leq 6, n=10, p=0{,}5)$$

$$= 0{,}9453 - 0{,}8281$$

$$= 0{,}1172$$

Podemos indicar que la probabilidad de que el alumno referido responda exactamente 7 preguntas correctamente (siendo que su elección en cada una es el azar), es de 0,1172 o 11,72 %

Parámetros que definen la distribución:

n: tamaño de la muestra

P: probabilidad de éxito

X: cantidad de éxitos en la muestra

$$P(X = x, n, p) = Cn^{x}\, p^{x}\, q^{n-x}$$

Esperanza y varianza:

$$E(x) = n \,.\, p \qquad\qquad Var(x) = n \,.\, p \,.\, q$$

2. Distribución Hipergeométrica

Son experimentos similares a los binomiales, con la diferencia que analizaremos a partir del siguiente ejemplo:

1. Un producto medicinal se envía en lotes de 20 artículos. La prueba para determinar si un artículo es defectuoso es muy costoso, así que el fabricante hace una muestra de la producción en vez de inspeccionar la totalidad.
2. Un plan de muestreo diseñado para minimizar el riesgo, necesita que se muestreen 5 artículos de cada lote y el rechazo del mismo si resulta más de uno defectuoso.

Supongamos que sabemos que un lote contiene 4 artículos defectuosos, ¿cuál es la probabilidad de que sea rechazado?

Estudiemos el problema continuando el procedimiento adoptado en el caso de experimentos binomiales. El experimento reside en observar uno a uno los 20 artículos de un lote y decidir si son defectuosos o no.

La variable aleatoria de estudio es "cantidad de artículos defectuosos", por lo tanto nuestro éxito será que el artículo sea defectuoso, en caso contrario obtendremos un fracaso.

Nos preguntamos si la probabilidad de éxito, se mantiene constante de ensayo en ensayo.

Conocemos que de los 20 artículos, 4 son defectuosos, por lo cual, al extraer el primer artículo y ejecutar la observación, existe una probabilidad de 4 en 20 de obtener éxito.

A continuación saquemos un segundo artículo para su estudio (sin reponer el anterior al lote) ¿se mantendrá la probabilidad anterior? Claro que no, este nuevo valor dependerá de lo obtenido anteriormente. Aquí es donde surge la diferencia que existe con el experimento binomial, en el hipergeométrico, la probabilidad de éxito NO SE MANTIENE CONSTANTE.

Cálculo de la probabilidad en el modelo hipergeométrico

Tomando el ejemplo anterior podremos construir el modelo matemático. Los datos son los siguientes:

1. En el caso de los lotes, los mismos constan de 20 artículos. Esta será la población por analizar y lo designaremos con la letra N (Para este caso N = 20)

Surge de la población que tenemos 4 éxitos (4 artículos defectuosos), lo indicaremos con la letra M, o sea M = 4. Además, N-M = 16 fracasos.

La muestra seleccionada es de 5 artículos, n=5

Si nuestra intención es saber la probabilidad de que la variable aleatoria tome el valor 2, o sea x=2. Por lo tanto, en la muestra de 5 habrá n-x = 3 fracasos.

Sintetizando, tenemos:

$$N=20 \qquad M=4 \qquad N-M=16 \qquad n=5 \qquad x=2 \qquad n-x=3$$

Para comprender la forma en que se calcula esta probabilidad, recurriremos a la Teoría Clásica del cálculo de probabilidades, es decir, casos favorables sobre casos posibles.

Empezando por los casos posibles, se trata de saber cuántos grupos diferentes de 5 elementos se pueden conformar con los 20 disponibles, o sea, cuántas combinaciones es posible conformar con 20 elementos tomándolos de 5 en 5:

$$C_5^{20} = \frac{20!}{5!\ (20-5)!}$$

Si tomamos los casos favorables, tendremos que ver cuántas formas diferentes pueden seleccionar 2 elementos (éxitos en la muestra) de un grupo de 4 (éxitos en la población), esto es:

$$C_2^4 = \frac{4!}{2!\ (4-2)!}$$

Y de cuántas formas diferentes pueden seleccionarse 3 elementos (fracasos en la muestra) de un grupo de 16 (fracasos en la población):

$$C_3^{16} = \frac{16!}{3!\ (16-3)!}$$

En este momento, por cada una de las maneras de selección de los éxitos, pueden darse de todas las formas diferentes, la selección de los fracasos; de este modo, se multiplicarán los casos posibles obteniendo:

Casos posibles = $C_2^4 \cdot C_3^{16}$ por lo tanto:

$$P(x = 2, N=20, M=4, n=5) = \frac{\text{Casos Favorables}}{\text{Casos Posibles}} = \frac{C_2^4 \cdot C_3^{16}}{C_5^{20}}$$

Y resolviendo:

$$P(x = 2, N=20, M=4, n=5) = \frac{\frac{4!}{2!\,(4-2)!} \; \frac{16!}{3!\,(16-3)!}}{5!\,(20-5)!} = 0{,}217$$

El resultado obtenido significa, que en las condiciones dadas, en el 21,7 % de los casos, un lote será rechazado.

Resumiendo, para N elementos en la población con M éxitos, seleccionando una muestra de n elementos y queriendo establecer la probabilidad de que la variable aleatoria x adopte un valor x tendremos:

$$P(X= x, N, M, n) = \frac{C_x^M \; C_{n-x}^{N-M}}{C_n^N}$$

Tomando un ejemplo:

1. En el Aeropuerto "Córdoba", el responsable de la Aduana, decide revisar 3 de 16 embarques provenientes del extranjero por vía aérea.

Si la selección es aleatoria y 5 de los 16 embarques contienen contrabando, encuentre la probabilidad de:

a. No localizar ningún embarque con contrabando.

b. Localizar al menos 2 embarques con contrabando.

c. Localizar a lo sumo 1 embarque con contrabando.

En primer término obtengamos del enunciado los siguientes datos:

N=16

M=5

n=3

X=0 para el caso 1

X ≥ 2 para el caso 2

X ≤ 1 para el caso 3

1

$$P(x=0, N=16, M=5, n=3) = \frac{C_0^5 \cdot C_3^{11}}{C_3^{16}} = 0{,}295$$

Por lo tanto, en casi el 30% de los embarques revisados, no encontrará contrabando.

2

$$P(x \geq 2, N=16, M=5, n=3)$$
$$= P(x=2, N=16, M=5, n=3) + P(x=3, N=16, M=5, n=3)$$
$$= 0{,}2143$$

Probabilidad de que encuentre dos o más embarques con contrabando.

3

$$P(x \leq 1, N=16, M=5, n=3)$$
$$= P(x=1, N=16, M=5, n=3) + P(x=0, N=16, M=5, n=3)$$
$$= 0{,}7857$$

Probabilidad de que encuentre un contrabando.

Esperanza y varianza en una distribución hipergeométrica

Las siguientes expresiones que pueden demostrarse tomando las definiciones de esperanza y varianza de una variable aleatoria, para el caso hipergeométrico resultan:

$$E(x) = n\frac{M}{N}$$

$$Var(x) = n\,\frac{M}{N}\,\frac{(N - M)}{(N)}\,\frac{(N - n)}{(N - 1)}$$

El último factor que surge en la expresión de la varianza es:

$$\frac{N - n}{N - 1}$$

Se designa factor de corrección para poblaciones finitas y aproximándose a 1, cuando el tamaño de la muestra sea más chico frente al tamaño de la población.

Parámetros que definen la distribución:

N: tamaño de la población

n: tamaño de la muestra

M: número de éxitos en la población

x: número de éxitos en la muestra

$$P(X = x, N, M, n) = \frac{C_x^M \; C_{n-x}^{N-M}}{C_n^N}$$

Esperanza y varianza:

$$E(X) = n\frac{M}{N} \qquad Var(x) = n\frac{M}{N}\,\frac{N-M}{N}\,\frac{N-n}{N-1}$$

3. Distribución de Poisson

El modelo se puede utilizar en dos situaciones:

1. **Como aproximación al modelo binomial**
2. **Como proceso aleatorio**

1. Como aproximación al modelo binomial

En entornos donde el tamaño de la muestra es muy grande (matemáticamente n tiende a infinito) y la probabilidad de éxito es muy pequeña, la distribución de Poisson tiene en cuenta los casos "poco frecuentes", y se logra a partir del límite para n que tiende a infinito, a la expresión del modelo binomial. En este desarrollo surge un nuevo parámetro, que es landa y que es igual a n.p .

La expresión matemática del modelo de Poisson es:

$$P(X=x, \lambda) = \frac{e^{-\lambda} \quad \lambda^{x}}{X\,!}$$

El punto que adoptaremos para aproximar un experimento binomial mediante el modelo de Poisson, será que la probabilidad de éxito sea baja y que el número de repeticiones sea alto de manera tal que el producto n.p resulte menor a 5.

Esperanza y varianza de la variable con distribución de Poisson

La esperanza como la varianza en este modelo son iguales a su parámetro landa, es decir:

$$E(x) = \lambda \,;\, Var(x) = \lambda$$

2. Como proceso aleatorio

El modelo de Poisson también permite una muy buena aproximación para el número de eventos, que ocurren en algunas ocasiones en el espacio, en el tiempo, en un volumen o cualquier otra magnitud, por ejemplo:

Una empresa dedicada a la fabricación de lavarropas inspecciona 1000 aparatos, a medida que salen de la línea de producción. Registra la cantidad de defectos por aparato y confecciona la siguiente tabla:

x	f	h
0	955	0,955
1	42	0,042
2	2	0,002
3	1	0,001
4	0	0,000

1. Estimar el número promedio de defectos por aparato.

$$E(x) = \sum p(xi) \, . \, xi = 0{,}0049 = \lambda = 1!$$

2. Calcular la probabilidad de que los futuros compradores reciban un aparato con un solo defecto.

$$p\,(x=1, \lambda\, 0 = 0{,}049) = \frac{e^{-0{,}049} \, . \, \lambda}{1!}$$

$$= 0{,}0467$$

Por lo tanto, la cantidad esperada de defectos por aparato, es notablemente menor a 1 (0,049) y la probabilidad de que los compradores reciban un aparato con un solo defecto es de 0,0467, o lo que es lo mismo, en un 4,67 % de los casos, los aparatos tendrán un único defecto.

Cálculo de la probabilidad de Poisson mediante tabla

Igualmente que para las situaciones referentes al modelo binomial, también es posible recurrir a una tabla de probabilidades para las variables aleatorias con distribución de Poisson. Por ejemplo, cuál es la probabilidad de que en cualquier período de 10 minutos, lleguen a la casilla, por lo menos, 20 vehículos. Por lo tanto, en símbolos sería:

$$1° \;\; P(x \leq a\ 20, \lambda=25) = P(x=0, \lambda=25) + (P(x=1, \lambda=25) + ... + P(x=20, \lambda=25)$$

Ubicando la tabla del **Anexo 2**, se busca primero el valor de λ que corresponde a la situación, en el caso de λ=25, en la primera columna de la tabla; para obtener el valor de x lo hallaremos en la primera fila del sector correspondiente.

En la intersección de la fila y columna seleccionadas, hallamos el valor de la probabilidad acumulada:

$$1° \ P(x \leq 20, \lambda=25) = 0{,}185$$

Por lo tanto, existe un 18,5 % de probabilidades de que logren llegar a la casilla de peaje 20 vehículos o menos, siendo que en promedio arriban 25 vehículos en un período de 10 minutos.

Parámetros que definen la distribución:

$\lambda = n \, . \, p$ como aproximación de la distribución binominal
Cantidad de éxitos por unidad de tiempo, volumen, superficie, etc. por unidad de tiempo.

X = cantidad de éxitos en la muestra

$$P(X=x, \lambda) = \frac{e^{-\lambda} \lambda^{x}}{X!}$$

Esperanza y varianza:

$E(x) = \lambda$ $\qquad$ $Var\,(x) = \lambda$

Debe verificarse que:

$$\sum_{i=0}^{\infty} p(x_i) = 1$$

Indique la opción correcta

1- Un experimento binomial debe cumplir las siguientes condiciones: el experimento se repite "n" veces en idénticas condiciones, cada prueba puede tener únicamente uno de dos resultados que se denominan éxito o fracaso, la probabilidad de éxito en una prueba es "p" y se mantiene constante durante todo el experimento, la probabilidad de fracaso será q=1-p, _________, y la variable aleatoria en estudio "x", es el número de éxitos observados en las "n" pruebas.

- ☐ Las pruebas son dependientes
- ☐ Las pruebas son independientes
- ☐ Las pruebas son determinísticas

Indique la opción correcta

2- Las distribuciones de variables discretas son _______.

- ☐ distribución binomial
- ☐ distribución hipergeométrica
- ☐ distribución binomial, distribución hipergeométrica y distribución poisson

Indique la opción correcta

3- En el experimento binomial, se observa que el experimento ________ .

- ☐ se repite en diferentes condiciones
- ☐ se repite en idénticas condiciones
- ☐ no se repite

Indique la opción correcta

4- En el experimento binomial, por cada repetición podemos tener solamente _______ resultados respectivamente.

- ☐ 4
- ☐ 0
- ☐ 2

Indique la opción correcta

5- En el experimento binomial la probabilidad de éxito se conserva _______ de repetición en repetición.

- ☐ constante
- ☐ variable
- ☐ íntegra

Indique la opción correcta

6- En el experimento binomial, los resultados obtenidos de cada experimento son ____.

- ☐ dependientes
- ☐ independientes
- ☐ continuos

Indique la opción correcta

7- En el modelo hipergeométrico la probabilidad de éxito _______.

- ☐ no existe
- ☐ se mantiene constante
- ☐ no se mantiene constante

Indique la opción correcta

8- El criterio adoptado para aproximar un experimento binomial mediante el modelo de Poisson, es que la probabilidad de éxito sea _______ y que el número de repeticiones sea alto, de modo que el producto n.p resulte menor a 5.

- ☐ baja
- ☐ alta
- ☐ media

Indique la opción correcta

9- La distribución de Poisson se puede utilizar en dos situaciones:

- ☐ como aproximación al modelo binomial y como proceso aleatorio
- ☐ como esperanza y varianza en una distribución binomial
- ☐ como aproximación del modelo hipergeométrico mediante la binomial

Indique la opción correcta

10- La distribución de Poisson como aproximación al modelo binomial se realiza en entornos donde el tamaño de la muestra es muy __________ y la probabilidad de éxito es muy pequeña.

- ☐ pequeña
- ☐ grande
- ☐ indistinta

Indique la opción correcta

11- Para aproximar un experimento binomial mediante el modelo de Poisson, será que la probabilidad de éxito sea baja y que el número de repeticiones sea alto de manera tal que el producto n . p resulte menor a:

- ☐ 5
- ☐ 10
- ☐ 15

Indique la opción correcta

12- El modelo de Poisson, ____________ permite una muy buena aproximación para el número de eventos, que ocurren en algunas ocasiones en el espacio, en el tiempo, en un volumen o cualquier otra magnitud.

- ☐ como aproximación al modelo binomial
- ☐ como proceso aleatorio
- ☐ como aproximación al modelo hipergeométrico

Respuestas correctas[28]

[28] 1) Las pruebas son independientes. 2) distribución binomial, distribución hipergeométrica y distribución poisson. 3) se repite en idénticas condiciones. 4) 2. 5) constante. 6) independientes. 7) no se mantiene constante. 8) baja. 9) como aproximación al modelo binomial y como proceso aleatorio. 10) grande. 11) 5. 12) como proceso aleatorio.

SP12/H2: Distribuciones de probabilidad de variable continua

Si observamos los problemas estadísticos que suelen presentarse regularmente en la realidad, veríamos que no todas las variables aleatorias se ajustan a la definición de variables aleatorias discretas. Por ejemplo, la investigación sobre el número de artículos defectuosos en un lote de n artículos fabricados implica una variable aleatoria discreta, porque el número de defectuosos es finito y toma uno de los valores 0, 1, 2, ... o n.

Tengamos en cuenta, por otra parte, la precipitación pluvial diaria en una determinada zona, en ese caso, teóricamente con un equipo de medición de gran precisión, podríamos relacionar a cada posible dato de cantidad de lluvia, un punto único en un intervalo.

Por lo tanto, cada uno de los números infinitos de puntos numerables del intervalo numérico, sería un valor posible distinto del dato de la cantidad de lluvia. El tipo de variable aleatoria que representa cualquier valor en un intervalo, se llama variable aleatoria continua, y son las distribuciones de probabilidad de este tipo de variables las que analizaremos a continuación.

Es primordial tener claro que la variable aleatoria siempre tomará un número determinado de valores, pero es continua porque puede tomar un número infinito de valores reales.

La distribución de probabilidad de una variable aleatoria discreta la alcanzamos estableciendo una probabilidad de cada uno de los posibles valores de la variable.

Para el caso de las variables continuas, esto no puede hacerse de la misma forma, no es posible matemáticamente establecer probabilidades distintas de cero al número infinito de valores que puede adoptar este tipo de variable, por consiguiente, habrá que desarrollar otro método para asignar probabilidades en este caso.

En este método, es necesario recapitular lo estudiado en el tema distribuciones de frecuencias de variables aleatorias continuas. En dicha ocasión, a partir de la tabla de distribución de frecuencias, elaboramos el gráfico correspondiente, llamado histograma, que podía ser de frecuencias absolutas o relativas.

Es significativo recordar que la característica principal de este gráfico es que se refiere a un gráfico de superficie y que la frecuencia absoluta o la frecuencia relativa (según el histograma) estaba dada por la superficie encerrada por el gráfico entre los valores de la variable considerados.

Link del video: https://www.youtube.com/watch?v=P340c116ABw

Texto del video: Podemos observar el gráfico de la figura 1 que representa un histograma de frecuencias relativas, donde la superficie sombreada muestra la proporción de observaciones (frecuencia relativa) incluidas entre los valores de la variable a y b.

En el momento que estudiamos probabilidad, pudimos analizar que, si el conjunto de observaciones era muy grande (tiende a infinito) el valor de la frecuencia relativa tiende a la probabilidad (teoría frecuencial de la probabilidad). Por lo tanto, si tenemos una gran cantidad de datos, podemos aseverar que la superficie encerrada por el histograma de frecuencias relativas entre dos valores de la variable, presenta la probabilidad de encontrar un valor de la misma comprendido entre los dos dados.

También tuvimos oportunidad de observar que el polígono de frecuencias es otro gráfico de superficie que se utiliza para este tipo de variable, su superficie es igual a la del histograma correspondiente. Por lo mencionado, se puede afirmar lo mismo respecto de la superficie encerrada por el polígono de frecuencia relativa entre dos valores de la variable aleatoria.

Figura 2

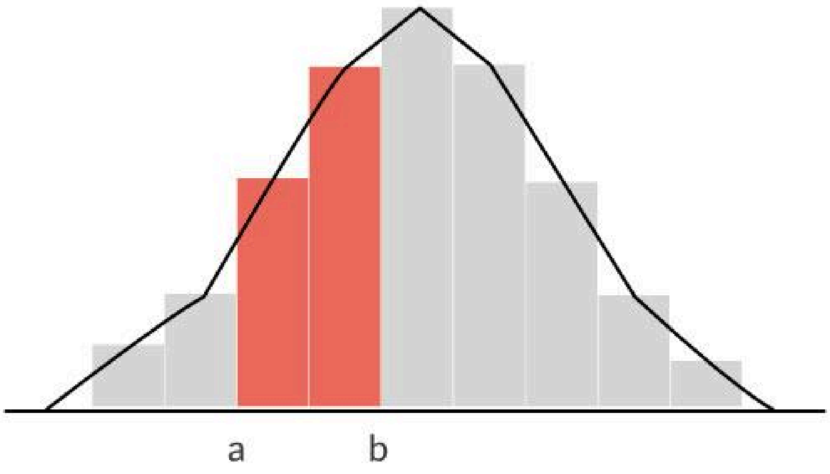
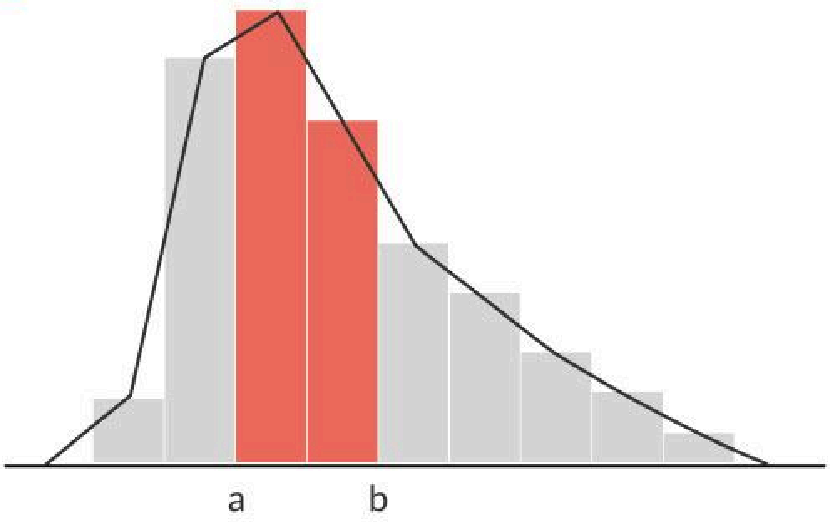


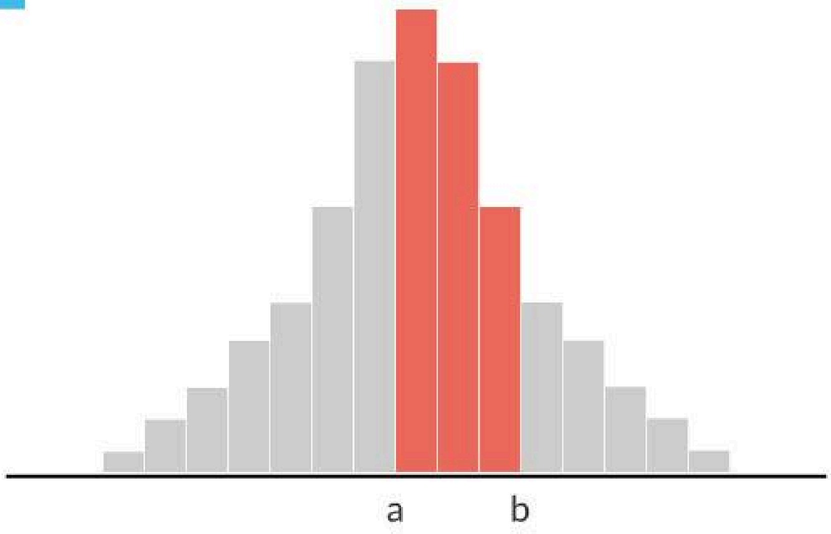
Figura 3



Podemos observar en los histogramas y polígonos de las figuras 2 y 3, que la idea anterior es válida no sólo cuando las gráficas son simétricas.

Deberíamos tener en cuenta, que tenemos más observaciones, de modo que es posible disminuir la amplitud de cada clase o intervalo y, a la vez, aumentar la cantidad de intervalos.

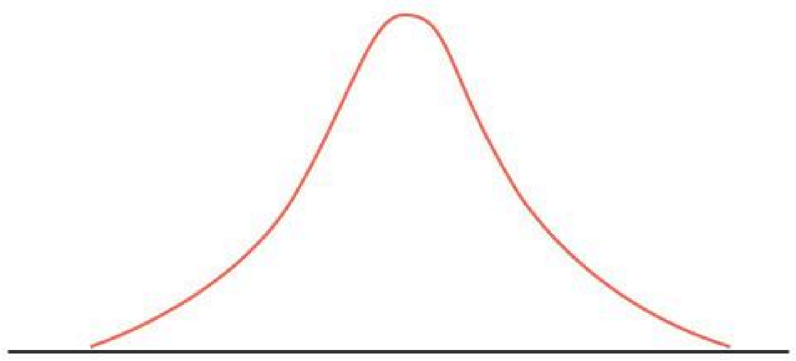
Figura 4



La zona sombreada muestra la probabilidad de hallar un valor de la variable comprendido entre a y b. Podemos observar en la figura 4 que si trazáramos el polígono de frecuencias relativas, este se iría aproximando a una curva, y lo haría cada vez más a medida que tuviéramos mayor cantidad de observaciones, ya que nos permitiría formar clases de menor amplitud y, por lo tanto, mayor cantidad de intervalos.

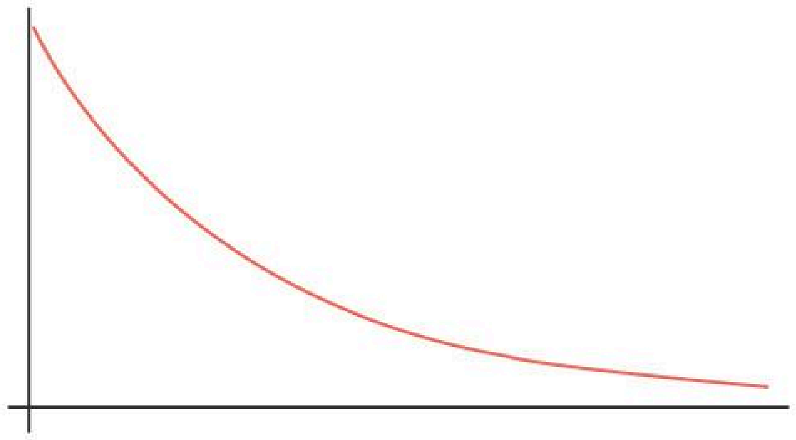
Continuando con el análisis teórico, si la cantidad de observaciones tiende a infinito, la amplitud de cada uno de los intervalos tenderá a cero y obtendremos un número infinito de ellos. En este caso el polígono de frecuencias se convertiría en una curva continua como la figura 5.

Figura 5



Del mismo modo, podría resultar una curva asimétrica, dependiendo lo mencionado, de los datos originales. En la figura 6 podemos ver una distribución de forma exponencial.

Figura 6



Las distribuciones que se han presentado, son distribuciones teóricas, es decir, indican funciones matemáticas y no situaciones reales. Sin embargo, son una excelente aproximación en mucho de los casos que se presentan en la realidad, con la ventaja de que el cálculo de superficie (o sea la probabilidad) se puede efectuar de manera mucho más precisa.

El cálculo se realiza con un recurso matemático llamado integración, pero también, existen tablas que nos darán la información de la probabilidad que necesitamos de manera más sencilla. El caso específico, en que la distribución es simétrica y tiene forma acampanada, es de singular importancia por la cantidad de situaciones prácticas en las que ocurre.

Seguidamente, estudiaremos particularmente dicha distribución.

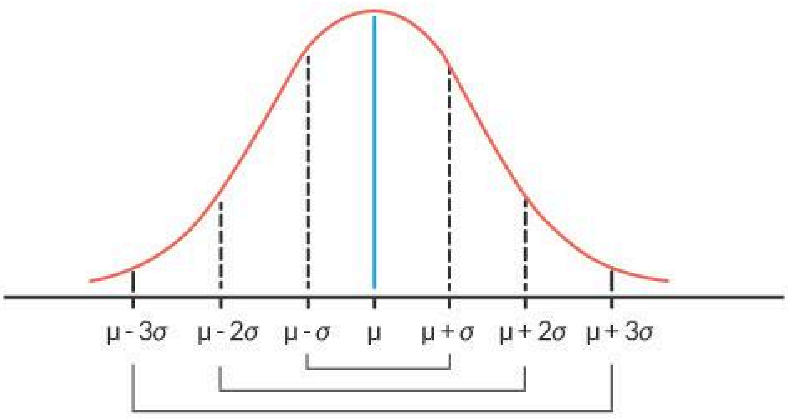
Una cuestión para destacar es que, cuando trabajamos con distribución de probabilidad de variables continuas, a diferencia de lo que sucedía con las discretas, la probabilidad de obtener un valor determinado de la variable es nula. Esto sucede, ya que dentro de un intervalo determinado por dos valores de la variable tenemos infinitos valores posibles, de manera tal que la probabilidad de que se dé uno específico de ellos es cero.

Link del video: https://www.youtube.com/watch?v=lBVT6-brPIE
Texto del video: Por ejemplo, si tenemos una alta probabilidad de que el promedio de lluvia mensual en un determinado mes esté ubicado entre 240,00 mm y 280,25 mm, seguramente, la probabilidad de que ese mes llueva exactamente 252,16 mm es difícil de cumplir, en realidad tiende a cero.

Distribución de probabilidad normal

Como fue analizado anteriormente, una distribución es normal, cuando la misma es simétrica, acampanada y tiene que cumplir con la regla empírica.

Figura 7



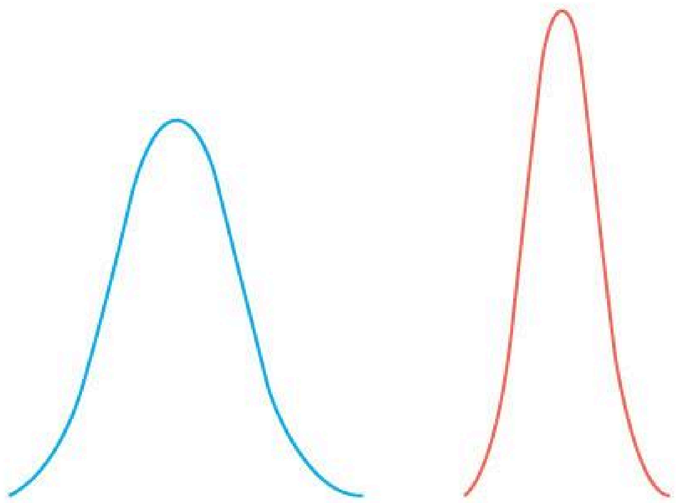
Entre μ – σ y μ + σ está comprendido, aproximadamente, el 68% de la superficie encerrada por la curva, entre μ – 2 σ y μ + 2 σ está comprendido, aproximadamente, el 95 % de la superficie y entre μ – 3 σ y μ + 3 σ se encontrará casi el 100 % de la superficie total encerrada por la curva (figura 7).

Dicha curva recibe el nombre de curva de distribución normal o curva de Gauss. Debido a su simetría, coinciden media, mediana y moda.

Estandarización de la variable con distribución normal

Si tuviésemos dos distribuciones normales con diferente media y del mismo modo, diferente desviación estándar, las formas de las curvas podrían ser como las presentadas en la figura 8.

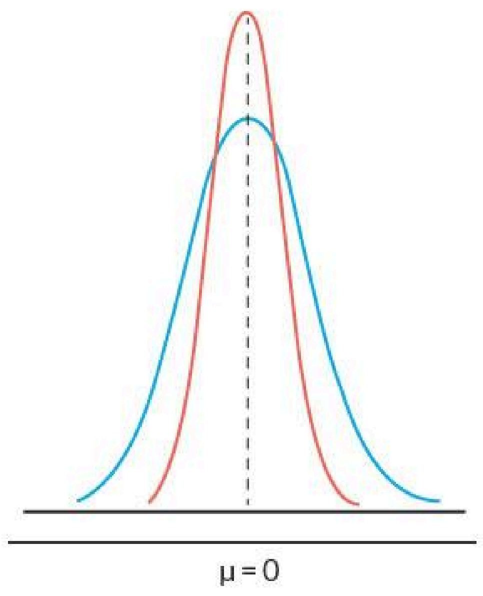
Figura 8



Si en cada una de las curvas, a cada valor de la variable x le restamos su correspondiente media, esto va a repercutir en un corrimiento de toda la distribución, de tal forma que se centre en el valor cero.

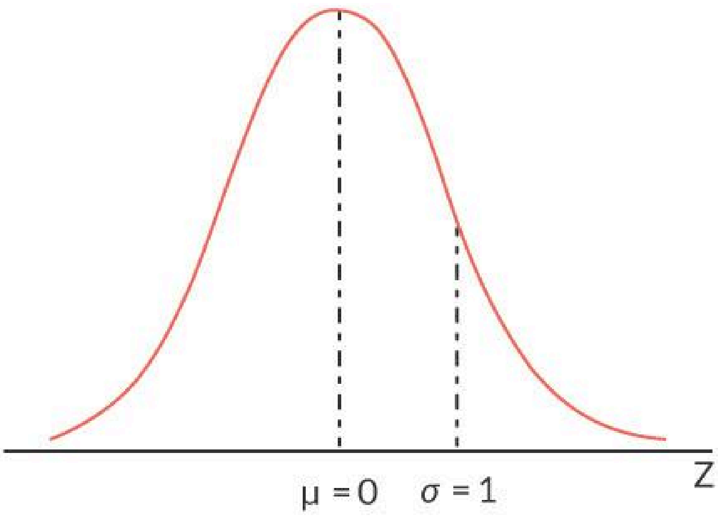
Por lo cual, tendríamos dos distribuciones normales superpuestas con el centro de ambas en cero, diferirían únicamente en su desviación estándar (Figura 9).

Figura 9



Si a continuación dividimos las diferencias (x – μ) por la correspondiente desviación estándar de cada distribución, ambas se adaptarán a una tercera, que toma el nombre de distribución normal estandarizada.

Figura 10



Por lo tanto, hemos logrado como resultado una nueva variable (z) que toma el nombre de variable normal estándar y sus valores se obtienen de la siguiente forma:

$$Z = \frac{x - \mu}{\sigma}$$

De este modo, logramos que a toda distribución normal se le haga corresponder una única distribución normal llamada distribución normal estandarizada, esta relación se establece de forma tal que la superficie encerrada entre dos valores (a y b) de una distribución normal es igual a la superficie encerrada entre los correspondientes valores Za y Zb de la distribución normal estandarizada (figura 10).

De esta forma podemos recurrir a una única tabla que nos dará la superficie (probabilidad) buscada. La variable Z, entonces, tiene distribución normal con media μ = 0 y desviación estándar σ = 1. Simbólicamente: Z símbolo N(0;1)

Uso de la tabla para la determinación de las probabilidades en una distribución normal

En primera instancia, trabajaremos directamente con la variable z, después plantearemos una situación con la variable x, la estandarizaremos y solucionaremos la situación. Utilizaremos la tabla del **Anexo 3** .

En dicha tabla, encontraremos la superficie bajo la curva normal desde menos infinito hasta el valor de la variable z que necesitamos, el gráfico que encabeza la tabla muestra lo mencionado. En la primera columna encontraremos el valor de la parte entera y el primer decimal de la variable z y, en la primera fila, obtendremos el segundo decimal de z.

La distribución normal como aproximación de la binomial

Una gran cantidad de distribuciones de probabilidad muestran la particularidad de que al cumplir con ciertos criterios, adquieren (aproximadamente) la forma de la distribución normal. La distribución binomial es una de las mencionadas anteriormente. Cuando el número n de pruebas en un experimento binomial es grande y p no es demasiado chico ni tampoco muy próximo a 1, es tomado como criterio de aproximación que n.p y n.q ambos sean mayores que 5.

Por ejemplo:

Una variable x tiene distribución binomial con parámetros característicos n = 40 y p =0,40. Establezca la probabilidad de que la variable tome valores menores a 15.

Entonces queremos saber: $P\ (x < 15;\ n = 40,\ p = 0{,}40)$, si buscamos en la tabla de la distribución binomial, no hallaremos esta probabilidad.

Por lo tanto, nos queda calcularla por intermedio de su modelo matemático o bien comprobar si cumple con los criterios de aproximación mediante la distribución normal y, si los cumple, requerir de la misma.

Los criterios son n.p y n.q ambos mayores que 5, en este caso 40. 0,40 = 16 y 40 . 0,60 = 24. Se cumplen los criterios, por lo tanto, debemos determinar la media y la desviación estándar:

$$\mu = n \cdot p \quad ; \quad \mu = 16$$

$$\sigma = \sqrt{n \cdot p \cdot q} \quad ; \quad \sigma = \sqrt{9{,}6} \quad ; \quad \sigma = 3{,}1$$

Entonces:

$$P(x < 15 \ ; \ n = 40 \ ; \ p = 0{,}40) = P(x < 15 \ ; \ \mu = 16 \ ; \ \sigma = 3{,}1)$$

$$= P\left(z < \frac{15 - 16}{3{,}1}\right)$$

$$= 0{,}3745$$

La probabilidad de que la variable considere valores inferiores a 15 es de 0,3745.

Si en cambio, quisiéramos saber la probabilidad de que la variable considere valores iguales o menores a 15, deberíamos llevar a cabo lo que se llama corrección por continuidad, que consiste en tomar valores de x menor a 15,5 ya que al ser la distribución normal una distribución de variable continua, la probabilidad de x = 15, como fue estudiado, es cero; por lo tanto, para incluir el valor 15 tenemos que ampliar el intervalo, de manera tal que este valor quede dentro del mismo.

Por lo tanto:

$$P(x \leq 15 \ ; \ n = 40 \ ; \ p = 0{,}40) = P(x < a\ 15{,}5 \ ; \ \mu = 16 \ ; \ \sigma = 3{,}1)$$

$$= P\left(z < \frac{15{,}5 - 16}{3{,}1}\right)$$

$$= P(z < -0{,}16)$$

$= 0{,}4364$ *como vemos cambia el valor de la probabilidad*

Esperanza y varianza:

$$E(x) = \mu \qquad\qquad Var(x) = \sigma^2$$

Debe estandarizarse la variable: $z = \dfrac{x - \mu}{\sigma}$ $\quad z \sim N(0;1)$

$$P(X \leq x\,;\ \mu\ ;\ \sigma) = P\left(z \leq \frac{x - \mu}{\sigma}\right)$$

Indique la opción correcta

1- En una distribución de probabilidad de variables continuas, a diferencia de lo que ocurre con las discretas, la probabilidad de obtener un determinado valor de la variable es _______ .

- ☐ uno
- ☐ nulo
- ☐ infinito

Indique la opción correcta

2- Una distribución normal debe ser _______ .

- ☐ asimétrica, acampanada y debe cumplir con la regla empírica
- ☐ simétrica, acampanada y debe cumplir con la regla empírica
- ☐ simétrica, acampanada y no debe cumplir con la regla empírica

Indique la opción correcta

3- La curva de distribución normal o curva de Gauss es _______ .

- ☐ simétrica, coinciden media, mediana y moda
- ☐ simétrica, no coinciden media, mediana y moda
- ☐ asimétrica, coinciden media, mediana y moda

Indique la opción correcta

4- No es posible tener infinitas tablas de probabilidades, por lo cual se hace una transformación a la variable con distribución normal de modo de hacerle corresponder otra distribución, también normal, que tiene el nombre de _______.

- ☐ distribución hipergeométrica estandarizada
- ☐ distribución binomial estandarizada
- ☐ distribución normal estandarizada

Indique la opción correcta

5- En la estandarización de la variable con distribución normal, tenemos generalmente dos distribuciones normales con diferente media y también diferente desviación estándar. Si en cada una de ellas a cada valor de la variable "x" le restamos su correspondiente media, resulta en un corrimiento de toda la distribución, que se centra en el valor cero. Por lo tanto tendríamos dos distribuciones normales superpuestas, con el centro de ambos en cero, que diferirían sólo en su ______ .

- ☐ esperanza
- ☐ mediana
- ☐ desviación estándar

Indique la opción correcta

6- La distribución binomial presenta la particularidad de que al cumplir con ciertos requisitos, adquieren la forma de la distribución normal cuando el número "n" de pruebas en un experimento binomial es _________ y "p" _________, suele tomarse como criterio de aproximación que n.p y n.q ambos sean mayores que ______________ ___________.

- ☐ grande, no es demasiado chico ni tampoco muy próximo a 1, 5
- ☐ pequeño,no es demasiado chico ni tampoco muy próximo a 1, 5
- ☐ grande, es muy chico y muy próximo a 1, 5

Respuestas correctas[29]

[29] 1) nulo. 2) simétrica, acampanada y debe cumplir con la regla empírica. 3) simétrica, coinciden media, mediana y moda. 4) distribución normal estandarizada. 5) desviación estándar. 6) grande, no es demasiado chico ni tampoco muy próximo a 1, 5.

Veamos la resolución de la situación profesional planteada:

- ¿Cuál es la probabilidad de que puntualmente dos solicitudes sean rechazadas?
  Distribución binomial de parámetros 10; 0.20.

  Rta: 0,30

- ¿Cuál es la probabilidad de que puntualmente cuatro sean rechazadas?
  Distribución binomial de parámetros 10; 0.20.

  Rta: 0,09

- ¿Cuál es la probabilidad de que menos de tres sean rechazadas?
  Distribución binomial de parámetros 10; 0.20.

  Rta: 0,68

- ¿Cuál es la probabilidad de que más de cinco sean rechazadas?
  Distribución binomial de parámetros 10; 0.20.

  Rta: 0,006

- ¿Cuál es la probabilidad de que realice 650 ensayos con una probabilidad de éxito de 0,02 y le indique qué distribución de probabilidad binomial, Poisson o normal debería ser?
  Es un experimento binomial con probabilidad de éxito 0,02 y n= 650.

Pero al ser n grande y p pequeño, se puede aproximar la distribución de Poisson con parámetro $\lambda = n.\,p = 13$.

SP12/ Ejercicio por resolver

En los siguientes casos, decida cuál es la distribución de probabilidad más apropiada para la variable aleatoria X (binomial, Poisson, normal):

a. X= "cantidad de llamadas que recibe una central en un intervalo de tiempo"
b. X= "cantidad de automóviles que pasan diariamente por un peaje"
c. X = "puntuaciones obtenidas en una prueba de inteligencia"
d. X = "sueldos de los empleados de una empresa"
e. X= "cantidad de personas que se manifestaron a favor de un producto en una encuesta"
f. X = " tiempo de espera en una cola del banco"
g. X = "cantidad de aciertos al blanco en diez tiros de bala"

Indique la opción correcta

1- Para averiguar la probabilidad de un valor X de una variable aleatoria con distribución binomial en un experimento que se repite n veces y la probabilidad de éxito es p debemos realizar la siguiente operación:

$$P(X=x, n, p) = Cn^{x}\ p^{x}\ q^{n-x}$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- El valor de la varianza en una distribución binomial es: Var(x) = n . p

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- La varianza en una distribución hipergeométrica y se expresa:

$$Var(x) = \frac{n\,M}{N}\ \frac{(N-M)}{(N)}\ \frac{(N-n)}{(N-1)}$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- La expresión matemática del modelo de Poisson es:

$$P(X=x, \lambda) = \frac{e^{-\lambda}\ \lambda^{x}}{X!}$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- La esperanza de una distribución normal es:

$$E(x) = \sigma^2$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

6- La varianza de una distribución normal es:Var(x) = μ

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

7- En una distribución normal la estandarización de la variable es:

$$z = \frac{x - \mu}{\sigma} \ , \ z \sim N(0;1)$$

$$P(X = x; \mu; \sigma) = P\left(z = \frac{x - \mu}{\sigma} ; \mu = 0 ; \sigma = 1\right)$$

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[30]

[30] 1) Verdadero. 2) Falso. 3) Verdadero. 4) Verdadero. 5) Falso. 6) Falso. 7) Verdadero.

Situación profesional 13: Distribución de Muestreo

Link del video: https://www.youtube.com/watch?v=25Uifpb1cXg

Texto del video: La empresa "CHAT" que fabrica grandes sistemas de computación en la ciudad de Córdoba, siempre se ha destacado por la confiabilidad de las unidades de procesamiento central (CPU) que entrega al mercado, pero su presidente está preocupado por los resultados de los últimos informes recibidos. En promedio 112 CPU al día son devueltos por problemas en algunas de sus partes, con una desviación estándar de 62.

El señor Marcial, presidente de la empresa, ha decidido que necesita estar seguro en un 80 % de que, en promedio, no se devolverán más de 122 CPU al día durante los siguientes 50 días, caso contrario pedirá una auditoría en los procedimientos.

Por ello, le encomienda a usted como pasante, que estudie la situación y lo asesore en función de los criterios dados, si es conveniente o no una auditoría en los procedimientos.

SP13/H1: Distribución de muestreo

Siguiendo la referencia del texto de estudio anterior "Probabilidad y estadística", la meta al efectuar un muestreo es poder inferir sobre los parámetros de una población a través de las características que nos presenta una muestra de los valores estadísticos tales como: la media, el desvío estándar o la proporción de los elementos de la población que cumplen con una característica determinada.

Por lo tanto, la media de la muestra nos da la posibilidad de inferir sobre la media poblacional; del mismo modo, el desvío estándar o la proporción de los elementos con las características en estudio de la muestra, también nos permite inferir sobre el comportamiento del desvío estándar o proporción poblacional.

Es indudable que los valores estadísticos de una muestra no van a coincidir con los parámetros poblacionales, siempre habrá una diferencia, que nace justamente de la variabilidad del muestreo, por lo tanto, los valores estadísticos variarán de muestra a muestra y, por consiguiente, podremos acercar con bastante aproximación las diferencias entre un valor estadístico y el parámetro correspondiente.

Cuando se procura inferir sobre un parámetro a través del estadístico muestral es importante determinar qué tan cercano está dicho valor del poblacional, y esto es consecuencia de tres factores:

1. El parámetro en estudio
2. El tamaño de la muestra
3. La dispersión poblacional

Variabilidad de estadísticos

Al comienzo del presente texto, definimos el concepto de Estadística y en la definición de la misma determinábamos el concepto de muestra. Seguidamente, vimos estadística descriptiva, luego, probabilidades y las distribuciones de variables aleatorias. Luego estudiamos conceptos sobre muestreo y los distintos tipos de muestra que pueden tomarse de acuerdo con las características de la población y, por consiguiente de acuerdo con la variable en estudio, tratando siempre de establecer los parámetros de la población mediante la estimación a través de los estadísticos muestrales.

De acuerdo a lo estudiado, éstos podrían ser: la media de la muestra para estimar la media poblacional o, de la misma manera, el desvío estándar o la proporción de la población en cuanto a determinada característica.

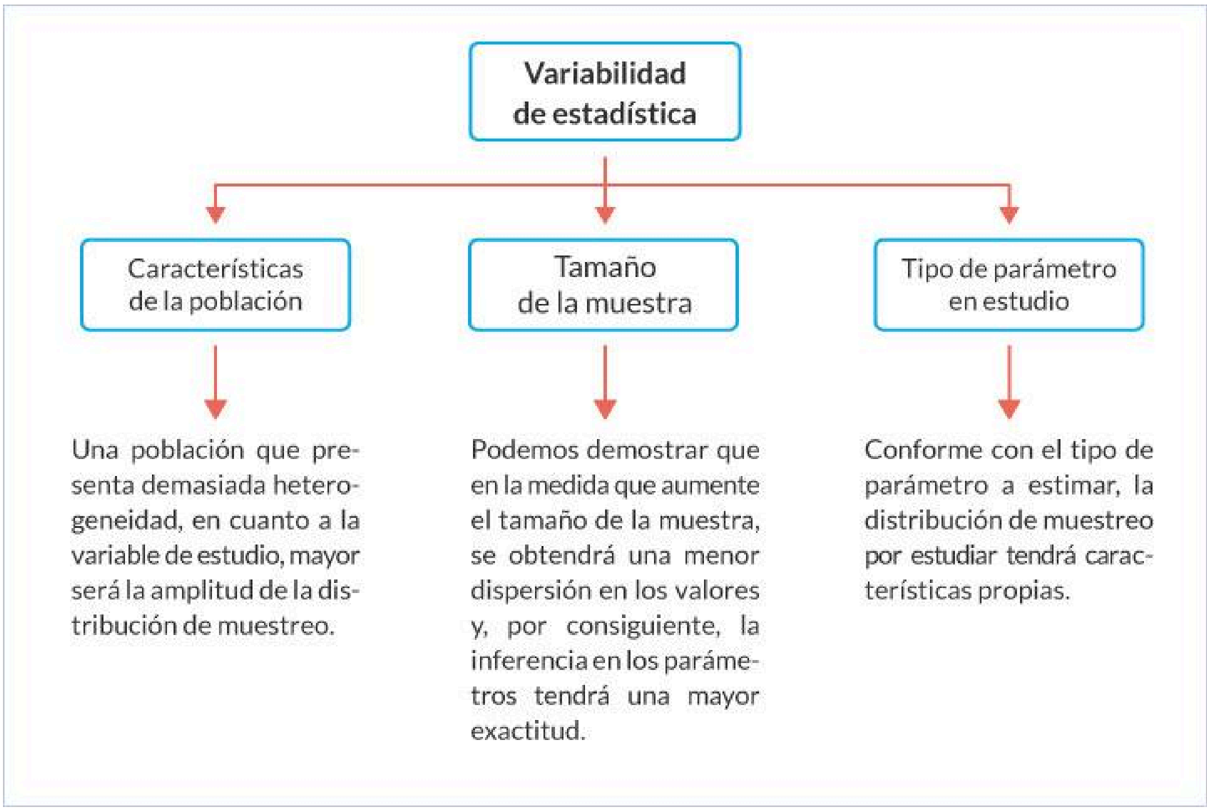
Si en una población se toma un número especificado N de muestras y, suponiendo que sea la media el parámetro de la población por establecer, es indudable que las medias de las distintas muestras serán diferentes entre sí y, por lo tanto, distintas de las media de la población, diferencias éstas que surgen como lógicas si se tiene en cuenta lo aleatorio del muestreo.

Si en una población se toma un número especificado N de muestras y, suponiendo que sea la media el parámetro de la población por establecer, es indudable que las medias de las distintas muestras serán diferentes entre sí y, por lo tanto, distintas de las media de la población, diferencias éstas que surgen como lógicas si se tiene en cuenta lo aleatorio del muestreo.

Por lo mencionado anteriormente, para poder hacer alguna inferencia sobre la media de la población en función de los valores estadísticos, será necesario tener en cuenta la forma en que varían los mismos en el muestreo.

Probaremos más adelante que, estudiadas las variaciones que experimentan los estadísticos en las distintas muestras, se crea una distribución a la que bien podríamos denominar distribución de muestreo de la media, la que muestra la característica de ser del tipo normal, tratándose siempre de muestras obtenidas en forma totalmente aleatoria.

Resumiendo podemos decir que la variabilidad producida en los estadísticos de las diferentes muestras depende de:



El concepto de distribución de muestreo es preliminar al estudio del teorema central del límite, necesario para el estudio de la inferencia estadística.

Veamos un ejemplo:

Distribución en el muestreo de la media

La fábrica de equipos electrónicos de precisión "El Águila" hnos., cuenta con un plantel de seis operarios a los que designaremos como (A, B, C, D, E y F), y cuyos salarios mensuales son respectivamente (1,2,3,3,4 y 5) multiplicados por $200, de tal forma que todos los resultados que se obtengan tendrán que ser multiplicados por la cantidad mencionada. Los salarios constituirán la población en estudio.

Con la finalidad de determinar los parámetros de la población crearemos una tabla del mismo tipo a las desarrolladas en la primera situación profesional:

Operarios	Salarios x $200	Media μ	(xi - μ)	$(xi - \mu)^2$
A	1	3	-2	4
B	2	3	-1	1
C	3	3	0	0
D	3	3	0	0
E	4	3	1	1
F	5	3	2	4
Sumas	18		0	10

Deslice el cursor sobre las fórmulas para ampliar la información.

Parámetros poblacionales

$$\mu = \frac{18}{6} = 3 \longrightarrow$$ Con lo que entregaría un sueldo promedio de $ 600

$$\sigma^2 = \frac{10}{6} = 1,67 \longrightarrow$$ Varianza

$$\sigma = \sqrt{1,67} = 1,29 \longrightarrow$$ Desvío estándar: 1,29 . $200 = $258

Tomemos ahora de esta población, todas las muestras posibles de dos elementos, a fin de establecer:

1. Las medias de cada una de las muestras (dichas medias formarán una nueva distribución a la que llamaremos distribución muestral).
2. La media de la distribución muestral.
3. La desviación estándar de las medias muestrales (el que llevará el nombre de error estándar de las medias).
4. Probabilidad de ocurrencia de las medias muestrales.

Si resolvemos lo solicitado:

1. Dada la población de 6 asalariados, podemos extraer tantas muestras de tamaño 2 como las combinaciones de 6 elementos tomados de a 2.

$$C_2^6 = 15$$

Por lo tanto, con los seis salarios que forman nuestra población en estudio se pueden tomar 15 combinaciones de dos elementos cada una de ellas y que son:

Muestras		$\bar{X}$ ($)
AB	(1,2)	1,5
AC	(1,3)	2
AD	(1,3)	2
AE	(1,4)	2,5
AF	(1,5)	3
BC	(2,3)	2,5
BD	(2,3)	2,5
BE	(2,4)	3
BF	(2,5)	3,5
CD	(3,3)	3
CE	(3,4)	3,5
CF	(3,5)	4
DE	(3,4)	3,5
DF	(3,5)	4
EF	(4,5)	4,5

Del mismo modo en que determinamos los parámetros de la población, estableceremos ahora los estadísticos de la distribución de las medias de las muestras, distribución ésta a la que ya designamos como distribución muestral.

$\bar{X}_i$	f_i	$\bar{X}i \cdot f_i$	fri = fi/n	$(\bar{X}_i - \mu)$	$(\bar{X}_i - \mu)^2$	$fi \cdot (\bar{X}_i - \mu)^2$
1,5	1	1,5	1/15	- 1,5	2,25	2,25
2	2	4	2/15	- 1	1	2
2,5	3	7,5	3/15	- 0,5	0,25	0,75
3	3	9	3/15	0	0	0
3,5	3	10,5	3/15	0,5	0,25	0,75
4	2	8	2/15	1	1	2
4,5	1	4,5	1/15	1,5	1,5	2,25
Total	**15**	**45**				**10**

2. La media de las medias muestrales es:

$$\mu = \frac{\sum (Xi\ fi)}{\sum fi} = \frac{45}{15} = 3$$

Valor al que empleada la relación nos entrega el sueldo medio de:

3.$200 = $600, valor este que coincide con la media poblacional.

3. La desviación estándar de las medias muestrales es:

$$\sigma \bar{x} = \sqrt{\frac{\sum (\bar{x}i - \mu)^2}{\sum fi}} = \sqrt{\frac{10}{15}} = 0{,}816$$

También llamado error estándar es:

$$\sigma \bar{x} = 0{,}816 \cdot \$200 = \$163{,}20$$

4. La probabilidad de ocurrencia de las medias muestrales viene dada por la columna de frecuencias relativas (fi/n).

De este modo, la probabilidad de que una muestra de tamaño 2, extraída de esa población, proporcione un promedio salarial de 4.$200 es de: 2/15 = 0,13

Por lo tanto, la probabilidad de que seleccionado uno de los trabajadores de manera totalmente azarosa, su sueldo sea de $800 es de:

$$P(\bar{x} = \$800) = 0{,}13$$

Como conclusión preliminar podemos decir que:

a. La media de la distribución de las medias de todas las muestras posibles de la población de tamaño (n=2) es igual a la media de la población.

b. El desvío estándar de dicha distribución, también denominado como error estándar, es menor al desvío estándar poblacional.

Volviendo a la población, podemos tomar todas las muestras posibles de cuatro elementos de cada una de ellas (n=4). El número de muestras posibles estará dado por las combinaciones de 6 tomadas de cuatro en cuatro:

$$C_4^6 = 15 \text{ muestras}$$

Muestras		$\bar{X}$. $200
ABCD	(1,2,3,3)	2,25
ABCE	(1,2,3,4)	2,50
ABCF	(1,2,3,5)	2,75
ABDE	(1,2,3,4)	2,50
ABDF	(1,2,3,5)	2,75
ABEF	(1,2,4,5)	3,00
ACDE	(1,3,3,4)	2,25
ACDF	(1,3,3,5)	3,00
ACEF	(1,3,4,5)	3,25
ADEF	(1,3,4,5)	3,25
BCDE	(2,3,3,4)	3,00
BCDF	(2,3,3,5)	3,25
BCEF	(2,3,4,5)	3,50
BDEF	(2,3,4,5)	3,50
CDEF	(3,3,4,5)	3,75

$\bar{X}i$	fi	f/n	fi . $\bar{X}i$	$(\bar{X}i - \mu)$	$(\bar{X}i - \mu)^2$	$f_i . (\bar{X}i - \mu)^2$
2,25	1	1/15	2,25	- 0,75	0,5625	0,5625
2,50	2	2/15	5,00	- 0,50	0,25	0,50
2,75	3	3/15	8,25	- 0,25	0,0625	0,1875
3,00	3	3/15	9,00	0	0	0
3,25	3	3/15	9,75	0,25	0,0625	0,1875
3,50	2	2/15	7,00	0,50	0,25	0,50
3,75	1	1/15	3,75	0,75	0,5625	0,5625
Total	**15**		**45**			**2,5**

La media de las medias muestrales es:

$$\mu = \frac{\sum (Xi . fi)}{\sum fi} = \frac{45}{15} = 3$$

Sueldo promedio = 3.$200 = $600

La desviación estándar de las medias muestrales es:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{X}_i - \mu)^2 \, fi}{\sum fi}} \quad \sqrt{\frac{2,5}{15}} = 0,403$$

$$\sigma_{\bar{x}} = 0,408 \,.\, \$200 = 0,408 \,.\, \$200 = 0,408 \,.\, \$200$$

$$\sigma_{\bar{x}} = \$\,81,60$$

Conclusiones

a. La media de la distribución de las medias de todas las muestras posibles que de n=4 elementos de cada una de ellas se puede tomar, de una población de 6 valores, es igual a la media de la población, lo que también ocurría en el caso de n=2.
b. La distribución tiende a adquirir las características propias de una distribución normal.
c. El desvío estándar de la distribución es menor que el desvío estándar poblacional.
d. El desvío estándar de la distribución muestral de n=4 es menor que el desvío estándar de la distribución muestral de n=2, es decir que, a medida que el número de elementos que componen la muestra aumenta, disminuye la dispersión de la distribución.

Tamaño de la muestra: n	Media	Desviación estándar	Rango o recorrido
2	3	0,816	4,5 - 1,5 = 3
4	3	0,408	3,75 - 2,25 = 1,5

Graficando ambas distribuciones, podemos observar que se ajustan a una curva normal.

Como es poco práctico conseguir todas las medias muestrales en cada estudio, la curva normal se utiliza usualmente para aproximar las probabilidades de las medias muestrales de una distribución en el muestreo.

La normalidad de una distribución en el muestreo de la media es denominado el "Teorema del límite central", cuyo enunciado es:

Si de una población de media μ y desvío estándar σ , se extraen todas las muestras posibles del mismo número de elementos n cada una de ellas, y de cada muestra obtenemos su media, la distribución de todas esas medias tendrá una distribución del tipo normal, independientemente del tipo de distribución que posea la población, con una media igual a:

la media poblacional μ y un desvío estándar √x menor al desvío estándar poblacional, denominado error estándar cuyo valor es igual a:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

En función del enunciado de dicho teorema, totalmente demostrado en el análisis de las distribuciones realizadas en la población de los salarios de la fábrica de equipos electrónicos de precisión, tanto para n=2 como para n=4, podemos realizar el siguiente análisis:

1. Cuando la población es suficientemente grande y está normalmente distribuida, la distribución de las medias muestrales será normal.
2. En el caso que la población no esté distribuida normalmente, la distribución de las medias muestrales se aproximará a una distribución normal, si el tamaño de la muestra es bastante grande (30 o más elementos).
3. Para la distribución normal de las medias muestrales, la misma tiene una media igual a:

$$\bar{x} = \mu$$

4. Una desviación estándar:

Desviación estándar de la población

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Tamaño de la muestra

5. Dada la situación con una población finita de tamaño N y si la relación entre

   n > 0,04

   N

   Entonces se hace necesario aplicar un factor de corrección:

$$F_c = \sqrt{\frac{(N-n)}{(N-1)}}$$

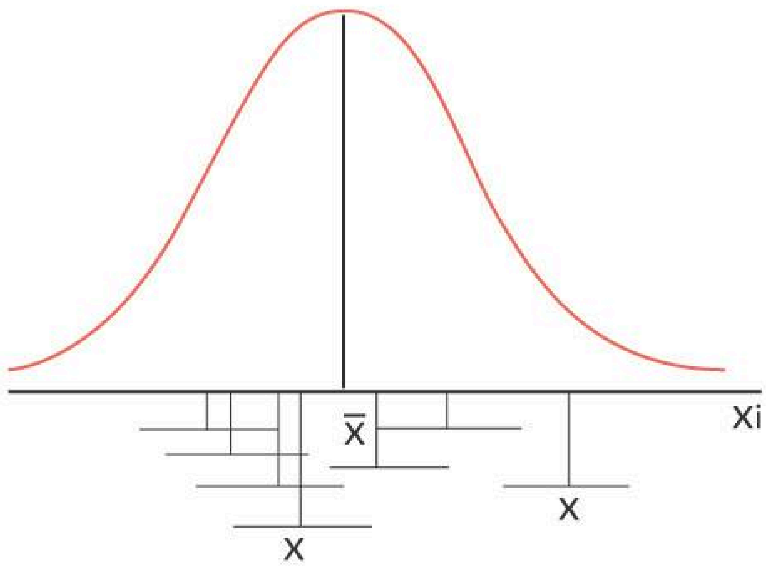
De manera tal que:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Observación

En el caso de la distribución de las medias muestrales, las mismas poseen una media que es igual a la media poblacional, según lo comprobado, pero que de ninguna forma tendrá que coincidir con las medias de las muestras, salvo en caso especiales.

Observemos el siguiente esquema:



En donde cada una de las xi pertenece a la media de una de las muestras. Después, para encontrar la probabilidad de una media muestral mediante la curva normal, debemos conocer:

1. **μ** : media de la población
2. **Xi**: media de la muestra
3. **σx**: desvío estándar de la media (error estándar)

Distribución en el muestreo de la proporción

La importancia de encontrar la proporción (o porcentaje) de los elementos de una población que cumplen con una determinada característica dada, es sumamente necesario.

Por ejemplo:

La proporción de votantes o no, la proporción de niños vacunados.

La relación entre la proporción de la población y la distribución en el muestreo de la proporción se presenta en el siguiente ejemplo:

Link del video: https://www.youtube.com/watch?v=ZMrNC71IJeE
Texto del video: En una empresa tenemos 6 personas A, B y C casados; X,Y y Z solteros.
Tomamos a las 6 personas como una población y le fijaremos un 1 a los casados y 0 a los solteros.

Queremos encontrar:

a. Los parámetros de la población.
b. Las proporciones de casados en las muestras de tamaño 4.
c. La media de las proporciones muestrales.
d. La desviación estándar de las proporciones muestrales.

Solucionemos por parte lo solicitado:
a. Los parámetros de la población:

Personas	Xi	Xi - p	$(xi - p)^2$
A	1	0,5	0,25
B	1	0,5	0,25
C	1	0,5	0,25
X	0	- 0,5	0,25
Y	0	- 0,5	0,25
Z	0	- 0,5	0,25
	3		**1,5**

$$p = \mu = \frac{3}{6} = 0{,}5 \quad \Rightarrow \quad q = 1 - p = 0{,}5$$

$$\sigma = \sqrt{\frac{1{,}5}{6}} = 0{,}5$$

También se puede usar la fórmula:

$$\sigma = \sqrt{p \,.\, q}$$

$$\sigma = \sqrt{0{,}5 \,.\, 0{,}5} = 0{,}5$$

b. Proporción de casados en las muestras de tamaño 4.

De la misma forma en que el desvío estándar de la distribución de las medias muestrales:

$$\sigma \bar{x} = \frac{\sigma}{\sqrt{n}}$$

En la distribución de las medias de proporciones se sabe que:

$$\sigma_{\bar{p}} = \sqrt{\frac{P * Q}{n}}$$

Y utilizando el factor de corrección para cuando la relación n/N ≥ 0,05

$$\sigma_{\bar{p}} = \sqrt{\frac{P \cdot Q}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$$

$$\sigma p = \sqrt{\frac{0,5 \cdot 0,5}{4}} \cdot \sqrt{\frac{6 - 4}{6 - 1}} = 0,158$$

Se multiplica por el corrector:

$$\sqrt{\frac{N - n}{N - 1}}$$

Porque $\frac{n}{N} >$ a $0.05 \quad \frac{4}{6}$

En lo referido a las proporciones, continuamos con las mismas conclusiones que para la media.

Para hallar la probabilidad de una proporción muestral de una distribución en el muestreo mediante la curva normal, es necesario conocer:

P = proporción de la población

p = proporción de la muestra

$$\sigma \bar{p} = \sqrt{\frac{P \, . \, Q}{n}}$$

Desviación estándar de la proporción

Recuerde que cuando la relación entre el tamaño de la muestra (n) y el de la población (N) es mayor o igual a 0.05 (esto implica que n es mayor o igual al 5% de N), entonces:

$$\sigma \bar{x} \ \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}} \qquad \text{y} \qquad \sigma p = \sqrt{\frac{P \, . \, Q}{n}} \sqrt{\frac{N - n}{N - 1}}$$

En el caso que el valor de N es lo suficientemente grande, la relación n/N es pequeña, y por lo tanto, el factor de corrección $\sqrt{\frac{N - n}{N - 1}}$ tiende a ser igual a 1 y su incidencia es nula.

Indique la opción correcta

1- Cuando se pretende inferir sobre un parámetro a través del estadístico muestral, es de vital importancia el determinar qué tan próximo está dicho valor del poblacional y esto depende de _____.

- ☐ el parámetro en estudio, tamaño de la muestra
- ☐ el parámetro en estudio, tamaño de la muestra, dispersión poblacional
- ☐ el parámetro en estudio, tamaño de la muestra, coeficiente de variación

Indique la opción correcta

2- En general podemos decir que la variabilidad producida en los estadísticos de las distintas muestras están relacionadas con ___.

- ☐ características de la población, tamaño de la muestra, tipo de parámetro de estudio
- ☐ características de la población
- ☐ características de la población, tamaño de la muestra

Indique la opción correcta

3- El Teorema del límite central dice: “si de una población de media μ y desvío estándar σ, se extraen todas las muestras posibles del ______ número de elementos “n”, cada una de ellas, y de cada muestra obtenemos su media, la distribución de todas esas medias tendrá una distribución del tipo normal, independientemente del tipo de distribución que posea la población, con una media x igual a la media poblacional μ y un desvío estándar σ x (sigma xmedia) es menor al desvío estándar poblacional, denominado error estándar”.

- ☐ infinito
- ☐ mismo
- ☐ característico

Indique la opción correcta

4- Para el caso de las proporciones, se sigue las mismas conclusiones que para la media. Para encontrar la probabilidad de una proporción muestral de una distribución en el muestreo mediante la curva normal, es necesario conocer _______.

- ☐ proporción de la población, proporción de la muestra, desviación estandar de la proporción
- ☐ proporción de la población
- ☐ ninguna de las anteriores

Indique la opción correcta

5- Si en una población se toma un número especificado N de muestras y suponiendo que sea la media el parámetro de la población por establecer, es indudable que las medias de las distintas muestras serán _____.

- ☐ iguales
- ☐ semejantes
- ☐ diferentes

Indique la opción correcta

6- Estudiadas las variaciones que experimentan los estadísticos en las distintas muestras, se crea una distribución a la que bien podríamos denominar_____, la que muestra la característica de ser del tipo normal, tratándose siempre de muestras obtenidas en forma totalmente aleatoria.

- ☐ distribución de muestreo de la media
- ☐ distribución de muestreo de la varianza
- ☐ distribución de muestreo de la desviación estándar

Indique la opción correcta

7- La distribución en el muestreo de la media tiende a adquirir las características propias de una distribución ______.

- ☐ hipergeométrica
- ☐ Poisson
- ☐ normal

Indique la opción correcta

8- La distribución en el muestreo de la media, el desvío estándar de la distribución es _____ que el desvío estándar poblacional.

- ☐ menor
- ☐ mayor
- ☐ igual

Respuestas correctas[31]

[31] 1) el parámetro en estudio, tamaño de la muestra, dispersión poblacional. 2) características de la población, tamaño de la muestra, tipo de parámetro de estudio. 3) mismo. 4) proporción de la población, proporción de la muestra, desviación estandar de la proporción. 5) diferentes. 6) distribución de muestreo de la media. 7) normal. 8) menor.

La inferencia estadística determina juicios acerca de algún aspecto de una población, estudiando una parte de ella llamada muestra.

Cuando se prueba una copa de vino para conocer su gusto, cuando se hojea un libro para ver si nos gusta o no, cuando se realiza el "boca de urna", en realidad se está muestreando.

El muestreo estadístico es similar, aunque sus métodos son más precisos.

Muestras y población

De acuerdo a lo mencionado precedentemente, podemos decir:

a. El estudio de toda la población se denomina censo.
b. El estudio de una parte de la población se llama muestra. Es necesario tener presente que toda muestra debe cumplir con la condición de ser debidamente representativa.

Ventajas de la muestra respecto del censo

1. Es más económico.
2. Ocupa menos tiempo. Un censo puede consumir tanto tiempo que el fenómeno que se estudia puede cambiar.
3. Una muestra podría ser más precisa. En un censo son necesarios más entrevistas, supervisores, etc. Por consiguiente, en la medida en que sea necesario ocupar más gente, disminuirá la calidad de las mismas y se hará más dificultoso el control.

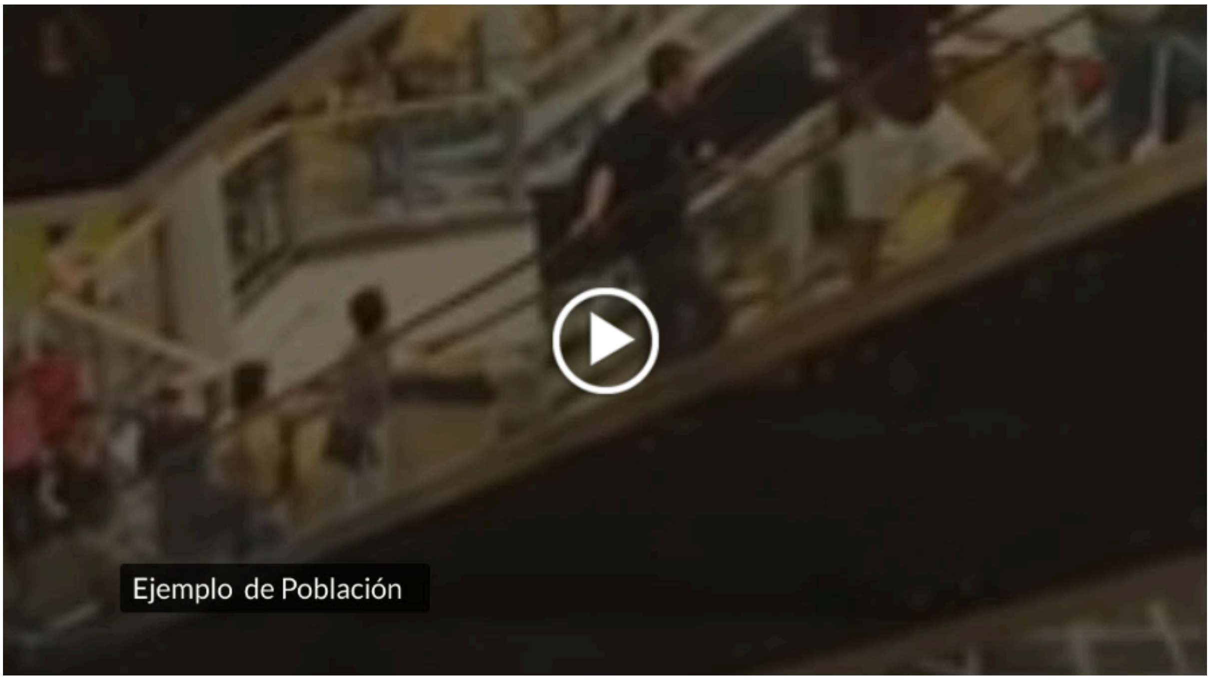
Conceptos fundamentales para el muestreo

a. **Elementos**
   Es la unidad acerca de la cual se solicita información.
b. **Población**

Es el conjunto de todos los elementos. Se debe definir en términos de:

- Elementos
- Unidad de muestreo
- Alcance
- Tiempo

Ejemplo 1:



Link del video: https://www.youtube.com/watch?v=6P0HJL6s5s4

Texto del video: Determinar la intención de compra en un Centro Comercial

Elemento: Mujeres de 15 a 60 años

Unidades de muestreo: Mujeres de 15 a 60 años (en este caso concuerda con el elemento)

Alcance: Córdoba, ciudad

Tiempo: del 1 de septiembre al 20 de octubre

Ejemplo 2:

Medición de la demanda de productos químicos industriales

Elemento: Ingenieros Químicos

Unidad de muestreo: Empresas que compran más de $600000 en químicos por año

Alcance. Zona noroeste del país

Tiempo: en el año 2009

c. **Unidad de Muestreo**

Son los elementos o el elemento disponible para su selección en alguna etapa del proceso de muestreo. En el muestreo de una sola etapa, el elemento y las unidades del muestreo son lo mismo (primer ejemplo).

En el segundo ejemplo, los elementos de interés son los ingenieros químicos, no obstante, arribamos a ellos por intermedio de un proceso de dos etapas.

En la primera, se selecciona una muestra de empresas que compran más de $ 600.000 en productos químicos industriales por año. Luego, en la segunda etapa, en dichas empresas se selecciona una muestra de ingenieros Químicos.

Un proceso de muestreo puede tener el número de etapas que el investigador determine. Por ejemplo: una muestra en cuatro etapas podría ser:

Etapa 1: Ciudades con más de 600.000 habitantes
Etapa 2: Barrios de ciudades (de la muestra)
Etapa 3: Núcleos familiares (de la muestra de manzanas)
Etapa 4: Hombres de más de 30 años

d. **Marco muestral**

Es la enumeración de todas las unidades de muestreo disponibles para la selección de una etapa cualquiera del proceso de muestreo.

En los casos mencionados, el marco muestral en la etapa 1 serían todas las ciudades de más de 600.000 habitantes, en la segunda etapa, el marco muestral estaría compuesto por los barrios de ciudades, etc.

Pasos que conforman el proceso de muestreo

Paso 1: definir la población

Esto incluye definir:

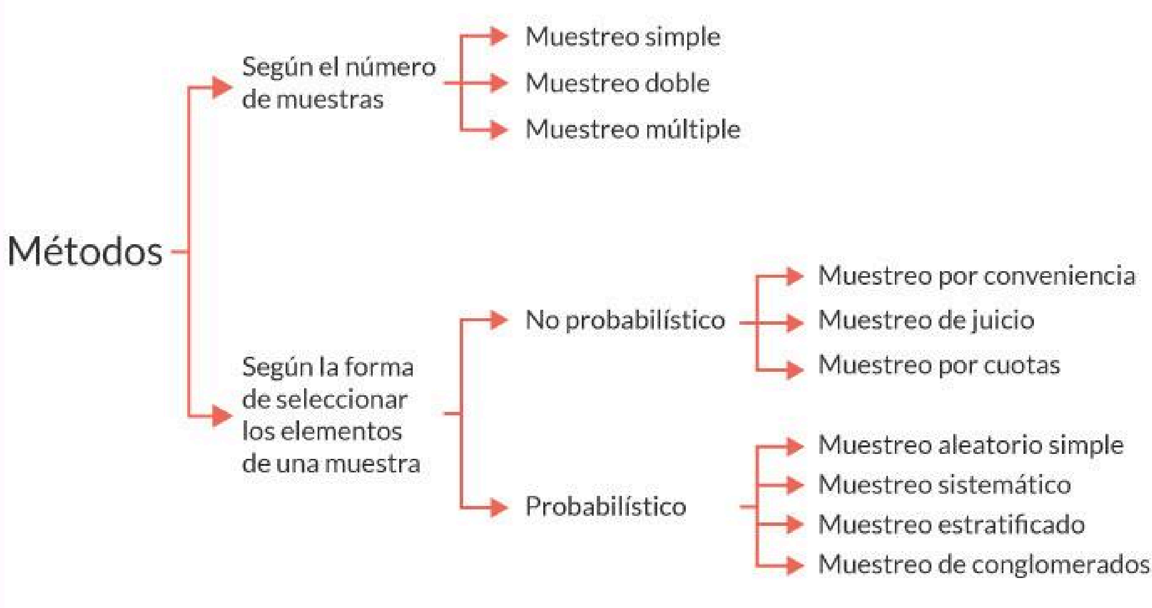
Elementos
Unidades de muestreo
Alcance
Tiempo

Paso 2: Identificar el marco muestral del cual se seleccionará la muestra.

Paso 3: Determinar el tamaño de la muestra.

Paso 4: Seleccionar la muestra.

Procedimiento de muestreo



Según el número de muestras:

a. **Muestreo simple:** se toma solamente una muestra.

b. **Muestreo doble:** cuando el resultado del estudio de la primera muestra no es todo lo satisfactorio que uno espera y, por consiguiente, no decisivo, por lo tanto se hace necesario considerar una segunda muestra.

Link del video: https://www.youtube.com/watch?v=8x38vWSMAZA

Texto del video: La calidad de un conjunto de productos determinado, si la primera muestra resulta una calidad muy alta en el conjunto, es aceptado; si resulta una calidad pobre, es rechazada.
Pero si resulta una calidad intermedia, entonces hace falta una segunda muestra. Después se combinan los resultados obtenidos de ambas muestras para tomar una determinación.

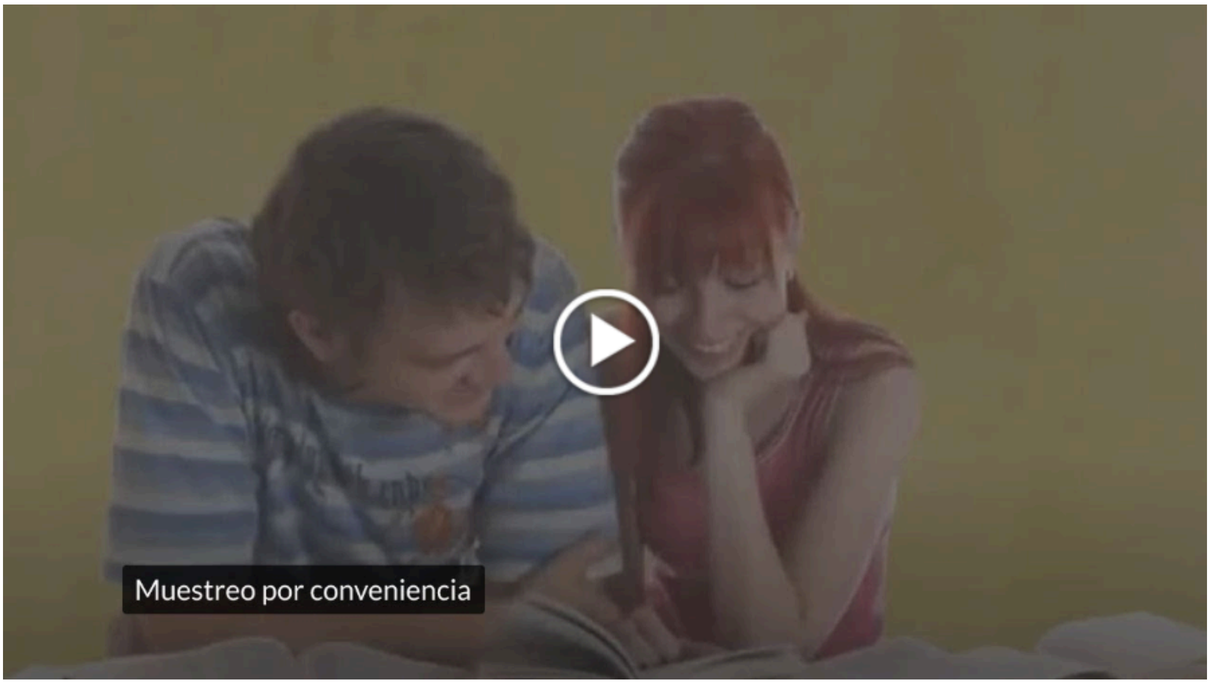
c. **Muestreo múltiple:** el procedimiento es similar al anterior, pero con más de dos muestras.

Según la forma de seleccionar los elementos de una muestra
1. No probabilístico

a. Muestreo por conveniencia: en este caso los elementos son seleccionados de acuerdo con la conveniencia del investigador. Se puede lograr información rápida y con bajo costo.

Este tipo de muestreo es utilizado, por lo general, para ensayar un cuestionario y al comienzo de una investigación.

Por ejemplo:



Link del video: https://www.youtube.com/watch?v=111TLSFjf9I
Texto del video:

- Solicitar la opinión de personas en un supermercado.

- Utilizar grupos de alumnos para comprobar una hipótesis de un experimento.
- Realizar entrevistas en la calle para la radio.

b. Muestreo de Juicio: en este caso se seleccionan los elementos a juicio de un experto, quien en función de su idoneidad, determina la muestra que considere representativa. Es importante tener en cuenta que en este tipo de muestreo es decisivo el juicio por parte del experto, ya que se pueden obtener sesgos subjetivos debido a las preferencias del investigador.

Por ejemplo: decidir qué barrios de Córdoba podrían considerarse los más convenientes para poner a prueba un nuevo producto recién elaborado, decisión de entrevistar a un visitador médico con relación a un medicamento o un laboratorio determinado, etc.

Un muestreo de juicio puede ser necesario cuando la probabilística no se puede implementar, o cuando el implementarlo elevaría enormemente los costos.

En caso de que el tamaño de la muestra sea pequeña, podría este tipo de muestreo ser más conveniente que la probabilística.

Por ejemplo: si es necesario elegir a dos estadios de básquet para realizar una determinada investigación, en este caso será necesario definir los estadios a muestrear en función de las características y el objetivo del estudio.

c. Muestreo por Cuotas: se alcanzan muestras con alguna característica de control. Se pretende diseñar previamente un modelo reducido de la población a muestrear y, de esta manera, se pueden eliminar los sesgos, que se generan en un muestreo de juicio.

Por ejemplo: supongamos que tenemos que observar una población de un barrio de la Capital de Córdoba y de la cual se sabe, sobre la base de datos estadísticos, que el 70 % de la población de más de 15 años se agrupa en el radio central y el resto en la periferia, con un pequeño predominio de mujeres sobre los hombres. Se pretende mantener en la muestra las mismas características obtenidas por los datos estadísticos. Si consideramos un tamaño de muestra n=25 se tendrá:

Hombres 10

Mujeres 15

En lo referido a la clasificación por edades, estará dada por la distribución que la misma tiene.

Este tipo de muestreo obedece en gran medida a la capacitación de los encuestadores.

2. Métodos Probabilísticos (muestreo aleatorio)

Los elementos que forman la muestra son seleccionados al azar, por consiguiente, todos tienen la misma probabilidad de ser elegidos.

Depende el error de muestreo, en gran medida, del criterio de quien lleva adelante el estudio.

Tipos de muestreo aleatorio

1. Muestreo aleatorio simple

Características

a. Todas las muestras tienen la misma probabilidad de ser elegidos.

b. Todos los elementos de la población se codifican y después se seleccionan mediante extracciones al azar o con tablas de números aleatorios.

En el caso que la población sea grande, es casi imposible numerar a todos los elementos. Por consiguiente se deben hacer modificaciones al muestreo aleatorio simple. Los tipos más comunes de muestreo aleatorio modificado son: sistemáticos, estratificados y de conglomerado.

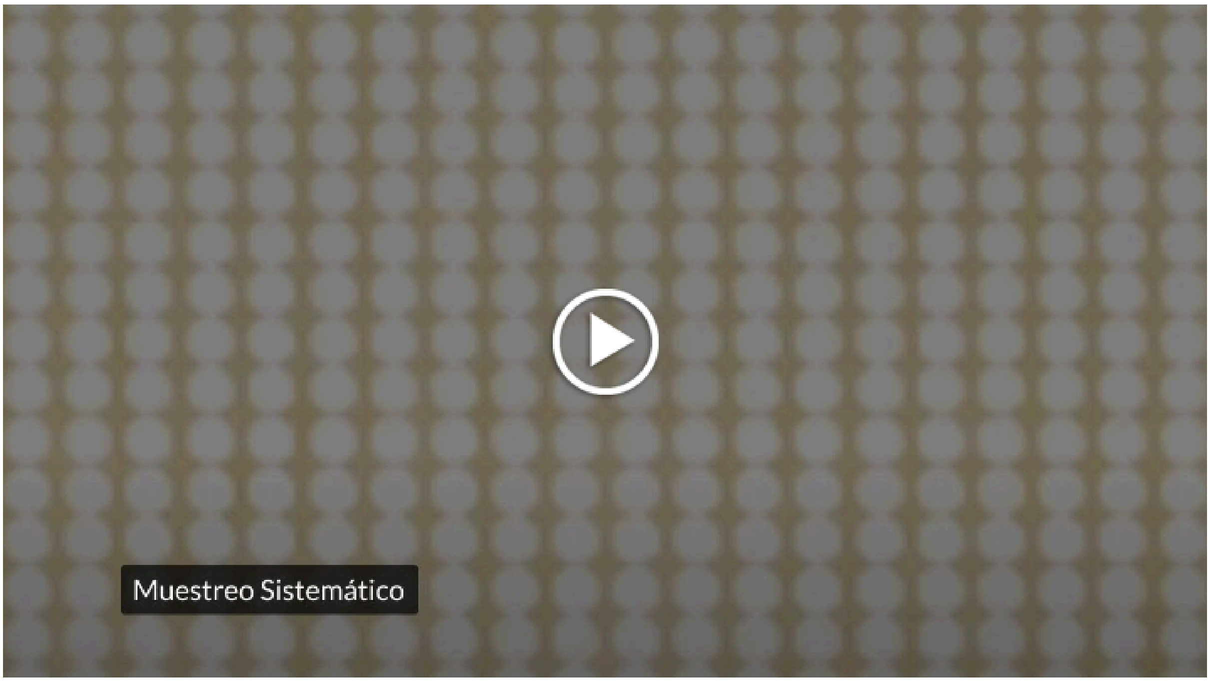
2. Muestreo sistemático

Características

a. Los elementos son seleccionados al azar, pero en forma ordenada.

b. El número de elementos de la población se divide por el número de la muestra. El cociente indicará si cada décimo, onceavo o centésimo elemento de la población va a ser seleccionado.

Por ejemplo:



Link del video: https://www.youtube.com/watch?v=k9PeKIHFBtc

Texto del video: Si tenemos una población de 10000, de la cual hay que tomar una muestra n=1000, el primer elemento de la muestra se escoge al azar; si es el número 200 de la lista, luego se escoge el número 300, luego el 400, el 500 y así sucesivamente hasta completar los 1000 elementos que formarán la muestra.

3. Muestreo estratificado

Características

a. Se divide la población en grupos homogéneos.

b. Los elementos son seleccionados al azar dentro de cada grupo.

c. El número de elementos que se selecciona de cada grupo es proporcional al tamaño del grupo con respecto a la población.

d. Si tomamos el mismo número de elementos de cada grupo, se le debe otorgar a los resultados obtenidos un peso, de acuerdo con la porción del grupo con respecto a la población total.

4. **Muestreo de conglomerados**

Características

a. Se divide a la población en grupos heterogéneos.

b. Se seleccionan al azar algunos de esos grupos.

c. Se seleccionan al azar elementos dentro de los grupos

Indique la opción correcta

1- La inferencia estadística establece juicios acerca de algún aspecto de una población, examinando una parte de ella llamada ______.

- ☐ población estándar
- ☐ proporción
- ☐ muestra

Indique la opción correcta

2- Las ventajas de la muestra respecto del censo son:

- ☐ es más caro, consume menos tiempo, una muestra puede ser más precisa.
- ☐ es más barato, consume menos tiempo, una muestra puede ser más precisa.
- ☐ es más barato, consume mucho tiempo, una muestra puede ser más precisa.

Indique la opción correcta

3- Los conceptos fundamentales para el muestreo son ____.

- ☐ elementos, población, unidad de muestreo, marco muestral.
- ☐ elementos, población, unidad de muestreo, media.
- ☐ elementos, población, unidad de muestreo, desviación estándar.

Indique la opción correcta

4- En el procedimiento de muestreo, según el número de muestras, una de ellas dice: "Cuando el resultado del estudio de la primera muestra no es todo lo satisfactorio que uno espera y, por lo tanto, no decisivo, se hace necesario considerar una segunda muestra. ¿A qué tipo de muestreo se refiere?

- ☐ muestreo simple.
- ☐ muestreo doble.
- ☐ muestreo múltiple.

Indique la opción correcta

5- En el procedimiento de muestreo, la forma de seleccionar los elementos de una variable no probabilística, una de ellas dice: se seleccionan los elementos a juicio de un experto, quien en función de su capacidad, determina la muestra que considere representativa. ¿A qué tipo de muestreo se refiere?

- ☐ muestreo por conveniencia.
- ☐ muestreo por cuotas.
- ☐ muestreo por juicio

Indique la opción correcta

6- En el procedimiento de muestreo, la forma de seleccionar los elementos de una variable probabilística, una de ellas indica las siguientes características: los elementos son seleccionados al azar, pero en forma ordenada, el número de elementos de la población se divide por el número de la muestra. El cociente indicará si cada décimo, onceavo o centésimo elemento de la selección va a ser seleccionado. ¿A qué tipo de muestreo se refiere?

- ☐ muestreo aleatorio.
- ☐ muestreo sistemático.
- ☐ muestreo estratificado.

Respuestas correctas[32]

[32] 1) muestra. 2) es más barato, consume menos tiempo, una muestra puede ser más precisa. 3) elementos, población, unidad de muestreo, marco muestral. 4) muestreo doble. 5) muestreo por juicio. 6) muestreo sistemático.

SP13/ Ejercicio resuelto

Según la situación profesional planteada, observamos lo siguiente:

P (media de la muestra <122 CPU) = P (media muestral estandarizada < 1.14) = 0,87

Luego, la probabilidad de que menos de 122 CPU en promedio sean devueltas es del 87%. Por lo tanto, debería hacer una auditoría ya que supera el 80%.

SP13/ Ejercicio por resolver

Una cadena de 150 supermercados, llamada "MEDITERRÁNEOS", ha sido puesta en venta y un grupo de inversionistas, que opera a nivel nacional, está interesado en adquirirla. Antes de hacerlo, la misma quiere asegurarse que sea redituable. El grupo ha solicitado los registros financieros de 35 de los supermercados. El Gerente General le afirma a los posibles compradores que las ganancias de cada supermercado es similar, con el mismo promedio y una variación de $1100. Si el Gerente está en lo cierto ¿cuál es la probabilidad de que la media de la muestra de los 35 supermercados se encuentre aproximadamente a los $100 de la media real?

Indique la opción correcta

1- La variabilidad producida en los estadísticos de las distintas muestras está relacionada con las características de la población, el tamaño de la muestra y el tipo de parámetro en estudio.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- En el Teorema del límite central el error estándar es igual a:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- La inferencia estadística establece juicios acerca de algún aspecto de una muestra, examinando una parte de ella llamada población.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Los pasos que conforman el proceso de muestreo son: definir la población, identificar el marco muestral del cual se seleccionará la muestra, determinar el tamaño de la muestra, además de seleccionar un procedimiento muestral y seleccionar la muestra.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- En el muestreo simple según el número de muestras, si el resultado del estudio de la primera muestra no es todo lo satisfactorio que uno espera y, por lo tanto, no decisivo, se hace necesario considerar una segunda muestra.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

6- En el procedimiento de muestreo no probabilístico, el muestreo por conveniencia se refiere a que los elementos son seleccionados de acuerdo con la conveniencia del investigador.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

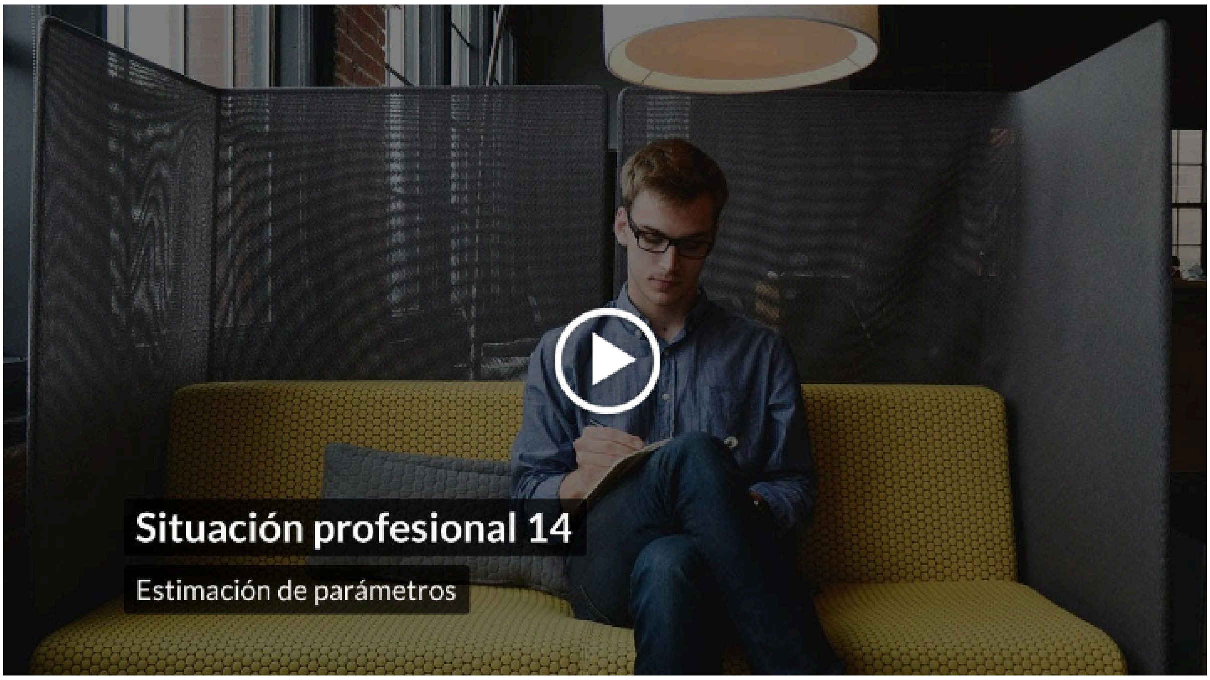
7- En el muestreo aleatorio simple según la forma de seleccionar los elementos de una variable (probabilística) sus características son: todas las muestras tienen la misma probabilidad de ser elegidas, todos los elementos de la población se codifican, después se seleccionan mediante extracciones al azar y, finalmente, cuando la población es grande, es casi imposible numerar todos sus elementos. Por lo tanto, hay que hacer modificaciones al muestreo aleatorio simple, siendo los tipos más comunes de muestreo aleatorio modificado los siguientes: sistemáticos, estratificados y de conglomerado.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[33]

[33] 1)Verdadero. 2)Verdadero. 3)Falso. 4)Verdadero. 5)Falso. 6)Verdadero. 7)Verdadero.

Situación profesional 14: Estimación de parámetros



Link del video: https://www.youtube.com/watch?v=caEnrREgPiw

Texto del video: Usted es parte del departamento de Recursos Humanos de la empresa "Limpiolín S.A.", que tiene una fábrica de productos de limpieza en el interior de Córdoba. Se desea determinar, a partir de un estudio sobre la población de empleados, cuánto influye el descanso diario en los niveles de producción de los mismos, ya que en estos últimos meses se notó una baja. Las autoridades, creen que esta situación, que no favorece a la empresa, puede estar provocada por una posible falta de descanso de sus empleados.

Para ello, se determina una muestra aleatoria de 80 personas que proporcionan la siguiente información:

1- Tiempo que descansa por día una persona:

x= 8,2 hs por día

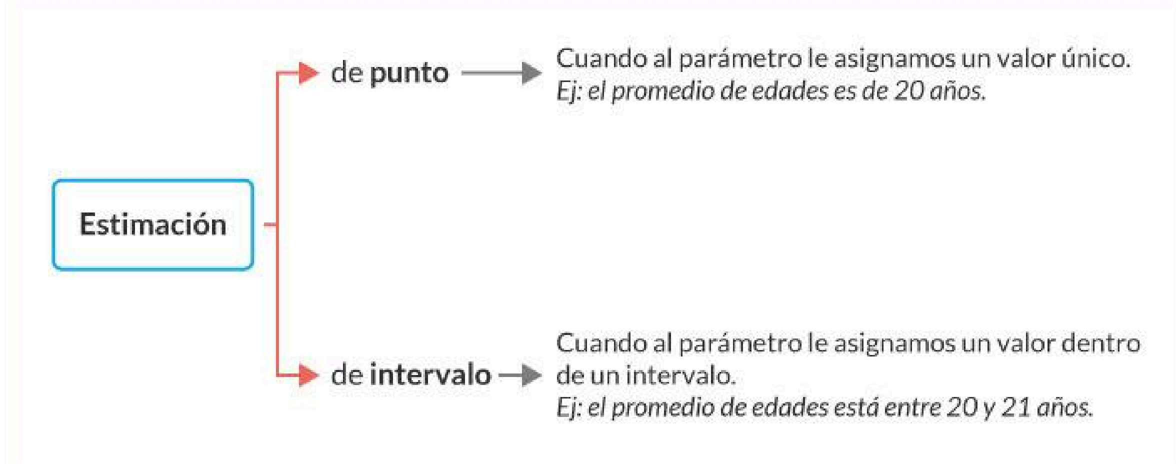
σ=2,8 hs por día

2- Se considera que el día franco cuenta como descanso también.

Sobre esa base, se determina que 60 personas descansan un día entre lunes y viernes.

SP14/H1: Estimación de punto y de intervalo de la media y proporción poblacional

Estimar es el proceso de usar un estadístico muestral para calcular (estimar) el correspondiente valor del parámetro poblacional desconocido.



La estimación de la media poblacional (μ) a partir de la media muestral (x) se hace a través de:

1. Estimación de punto: se asigna un valor único.

2. Estimación de intervalos: el procedimiento a seguir se detalla a continuación:

a. Encuentre la media muestral, lo cual es una estimación de punto.

b. Calcule la desviación estándar de la media σ:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

cuando la población es grande $\frac{n}{N} \leq 0{,}05$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

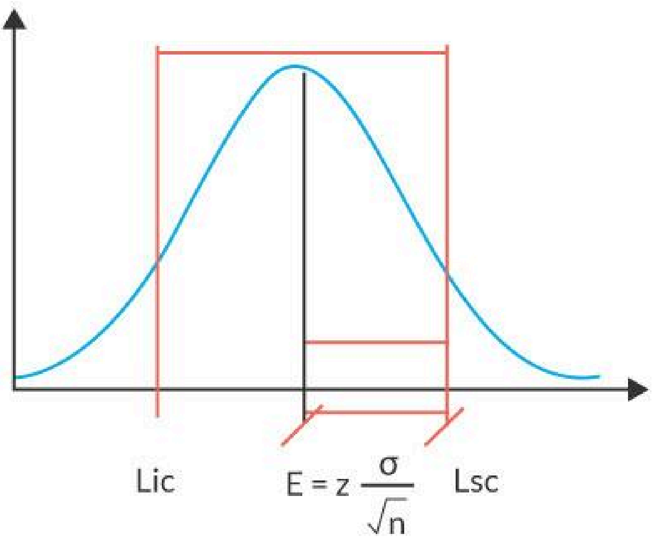
cuando n/N>0,05

La desviación estándar de la población σ puede calcularse:

- mediante una muestra
- subjetivamente
- por recorrido de los datos: 4%, 6% u 8% en la medida que los datos se distribuyan asimétricamente.

Para calcular los límites de confianza, tenga en cuenta que $\sigma\bar{x} = \frac{\sigma}{\sqrt{n}}$

Error permitido



A partir de la media, tendrá un error permitido, que estará dado por las siguientes expresiones:

El límite superior del intervalo, que llamaremos LSC, está dado por:

$$LSC = \bar{x} \; -+ \; z \,.\, \sigma\bar{x}$$

donde el error permitido es

$$E = z \,.\, \sigma\bar{x}$$

El límite inferior del intervalo, que llamaremos LIC, estará dado por:

$$LIC = \bar{x} - z \cdot \sigma\bar{x}$$

Es muy importante que usted recuerde que:

$$E = z \cdot \sigma\bar{x} \longrightarrow \text{Error permitido}$$

$$\sigma\bar{x} = \frac{\sigma}{\sqrt{n}} \longrightarrow \text{Error estándar muestral}$$

Estimación de una proporción (P) a partir de una proporción muestral (p)

El procedimiento es análogo al de estimar una media poblacional. El mismo consiste en:

1. Encontrar la proporción de la muestra (p)

2. Calcular la desviación estándar de la proporción (σp)

$$\sigma p = \sqrt{\frac{P.Q}{n}}$$

cuando la población es grande:

$$\frac{n}{N} \leq 0{,}05$$

La proporción P puede calcularse:

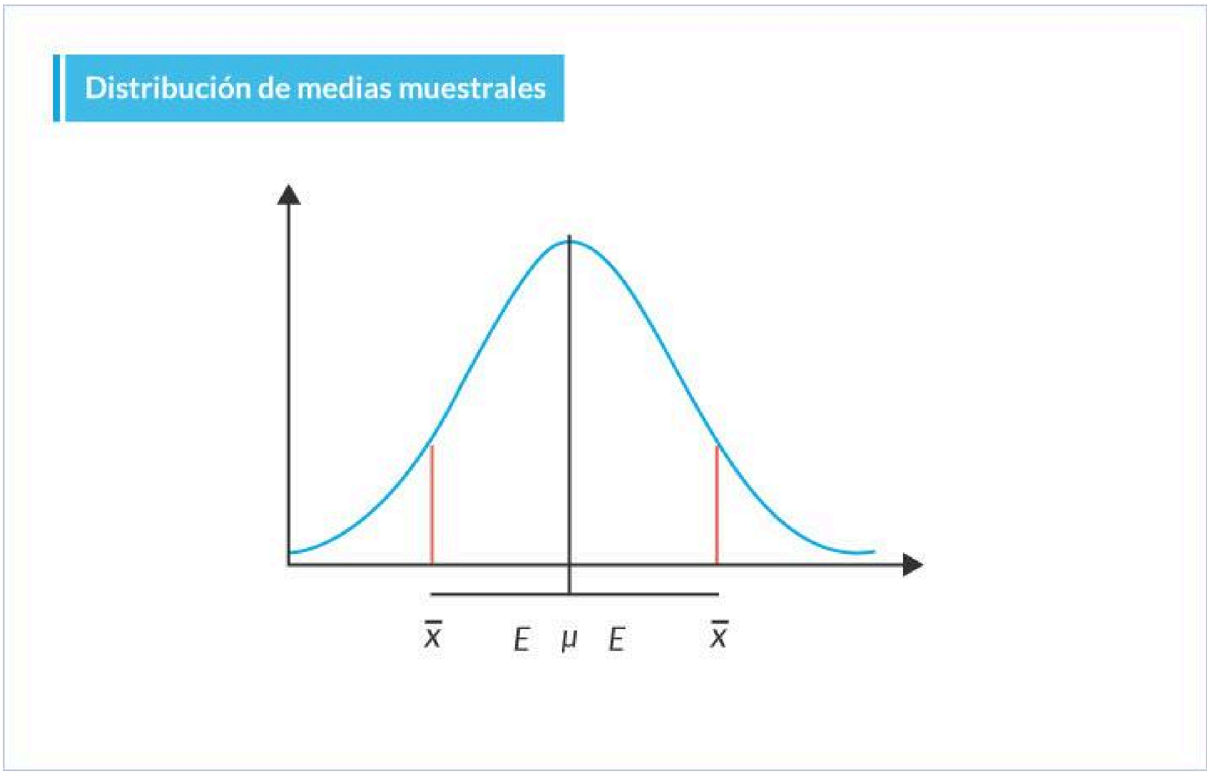
- mediante una muestra
- subjetivamente
- si no hay información previa P=0,5

3. Calcular los límites de confianza

límite superior=p+ zσp

límite inferior=p- zσp

Para estimar una media poblacional



El gráfico corresponde a la distribución de las medias muestrales, que de acuerdo al Teorema del Límite Central, sabemos que la distribución de todas las muestras posibles que se pueden tomar de la población, de n elementos cada una de ellas, y obteniendo de cada una la media, la distribución de esas medias es una distribución normal con una media igual a la media poblacional μ. Definiremos como error de muestreo E, a la diferencia entre la media de la distribución de las medias y la media correspondiente a una de las muestras x-. Ahora bien, como ese segmento comprendido entre la media de la distribución y la media de una muestra se mide en unidades de desviación estándar, y tendremos:

$$E = \bar{x} - \mu = z \cdot \sigma\bar{x}$$

Este error E también suele denominarse " $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ Error Permitido". Si tenemos en cuenta que, luego la expresión del error de estimación será:

Desvío estándar poblacional.

$$E = z \cdot \frac{\sigma}{\sqrt{n}}$$

Número de elementos de la muestra.

Que nos entregará la probabilidad de ocurrencia.

El error de estimación es un valor al cual lo podemos estimar a priori como el máximo error permitido entre el valor de la media de una muestra de número de elementos n, y el valor de la media poblacional, siempre sujeto a una determinada probabilidad de ocurrencia.

SP14/Autoevaluación 1

Indique la opción correcta

1- La estimación de la media poblacional a partir de la media muestral (x) se hace a través de: Estimación de punto y Estimación de intervalos.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- El procedimiento de la Estimación de intervalos es el siguiente: Encontrar la media muestral, lo cual es una estimación de Punto y Calcular la desviación estándar.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- La desviación estándar de la población puede calcularse solamente de manera subjetiva.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- La desviación estándar de la población puede calcularse mediante una muestra, subjetivamente y por recorrido de los datos.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- La Estimación de una proporción (P) a partir de una proporción muestral (p) consiste en: Encontrar la proporción de la muestra, Calcular la desviación estándar de la proporción y Calcular los límites de confianza.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

6- El error de estimación es un valor al cual lo podemos estimar a priori como el máximo error permitido entre el valor de la media de una muestra de número de elementos n, y el valor de la media poblacional, siempre sujeto a una determinada probabilidad de ocurrencia.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[34]

[34] 1)Verdadero. 2)Verdadero. 3)Falso. 4)Verdadero. 5)Verdadero. 6)Verdadero.

SP14/H2: Determinación del tamaño de una muestra

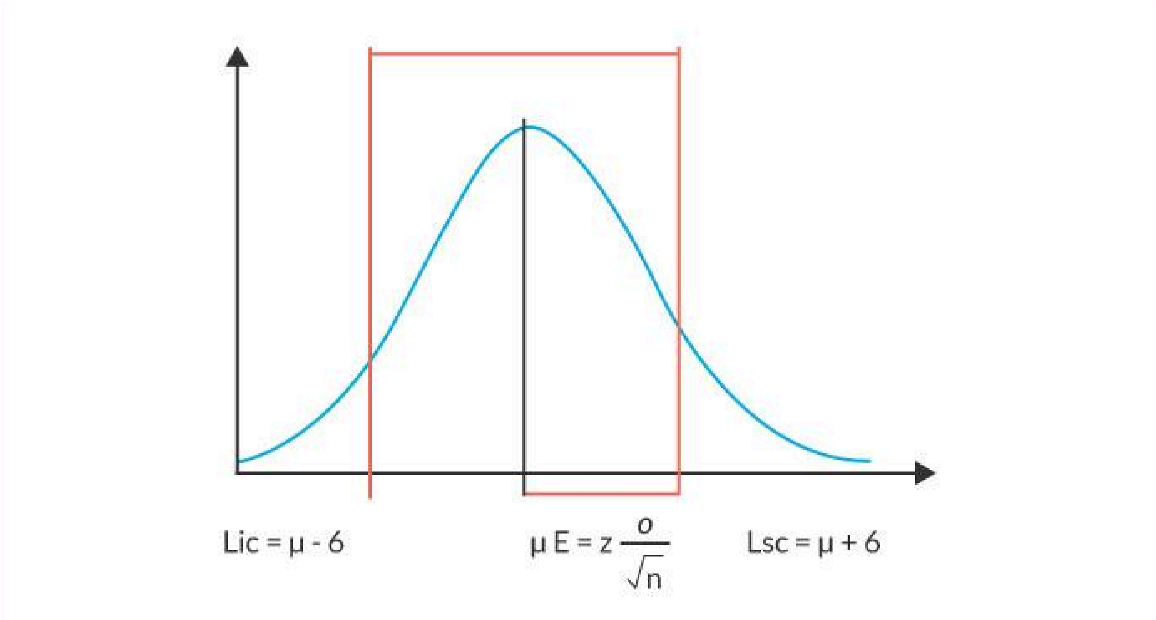
En muchas oportunidades, nos resulta necesario determinar cuál debe ser el tamaño de la muestra a tomar, a fin de que el estadístico entregado por la misma se comporte como un muy buen estimador. Analizaremos la determinación del tamaño de una muestra, tanto para la media como para el porcentaje poblacional.

a. Para estimar la media poblacional

La expresión del error de estimación nos permite determinar el tamaño que debe tener una muestra para la determinación de la media poblacional con un determinado error, y la probabilidad de que ello ocurra.

$$n = \left(\frac{z \, . \, \sigma}{E} \right)^2$$

Por ejemplo, suponga que se pretende estimar el sueldo medio de los operarios de una fábrica, pero especificando que el valor de la media de la muestra no difiera más de \$6 del valor medio poblacional μ con una determinada bondad. Esto implicaría que el error de estimación es de \$6, y el intervalo de confianza está dado por x ±\$6.



Z está establecida según el grado de bondad o nivel de confianza del intervalo.

Nivel de confianza	50%	68%	90%	95%	95,5%	99%	99,73%
Valor de Z	0,67	1	1,645	1,96	2	2,58	3

Los niveles de confianza más comunes son los del 95% y 99%.

Cómo estimar cuando se desconoce el desvío estándar poblacional.

En la mayoría de los casos que se presentan, el valor del desvío estándar será desconocido pudiendo, entonces, estimar su valor mediante las siguientes consideraciones:

1. Tomando como valor de desvío estándar poblacional al desvío correspondiente a una muestra S.

2. Aproximando en forma subjetiva, basados en experiencias pasadas.

3. Teniendo en cuenta la asimetría de la distribución, el recorrido de la misma podrá ser dividido por 4,6 u 8=recorrido/ω, donde ω tomará alguno de los valores mencionados, teniendo en caso de una distribución simétrica, ω=4.

b. Estimación de una proporción poblacional

El análisis es semejante al de la media; seguiremos el mismo proceso teniendo en cuenta que en este caso el error muestral para proporciones es:

$$E = z \sqrt{\frac{P.Q}{n}}$$

de donde, despejando a n de manera algebraica, tendremos que:

$$n = \frac{Z^2}{E^2} . P . Q$$

Teniendo en cuenta que:

E: se especifica en forma de porcentajes. Si se fijara, por ejemplo, un valor de E=3%, entonces el intervalo de confianza estará dado por: p±$3 (donde p es el porcentaje muestral).

p: es el porcentaje poblacional, el cual análogamente a la media podrá:

1. Ser estimado por la experiencia o por estudios muestrales anteriores.
2. Cuando no existen antecedentes que permitan la estimación de P se le otorga el valor del 50%

$$Q := (1 - P)$$

Observe que el producto P.Q adquiere su máximo valor cuando P=Q=0,5; ésta es la razón por la cual, no conociendo el valor poblacional P y no teniendo antecedentes que permitan definirlo, nos posicionamos en la situación más desfavorable asignándole el valor 0,5.

Indique la opción correcta

1- El gerente de un supermercado desea estimar el promedio de compras mensuales, de los clientes que usan tarjetas de crédito, con un error no mayor a 2,50 y con una aproximación del 95%. Se sabe además que, el desvío estándar es de 7,50.Luego, el número de cuentas que deben seleccionarse para el muestreo son:

- ☐ 30 cuentas
- ☐ 35 cuentas
- ☐ 68 cuentas

Indique la opción correcta

2- Una empresa de camiones desea conocer el kilometraje promedio recorrido semanalmente por sus unidades, bajo las siguientes condiciones:

- ☐ 246 camiones
- ☐ 208 camiones
- ☐ 272 camiones

Indique la opción correcta

3- Para estimar una proporción a partir de una proporción muestral primero debemos calcular la desviación estandar de la proporción.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- La expresión del error de estimación nos permite determinar el tamaño que debe tener una muestra para la determinación de la media poblacional con un determinado error, y la probabilidad de que ello ocurra.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- En la mayoría de los casos que se presentan, el valor del desvío estándar será desconocido pudiendo, entonces, estimar su valor mediante las siguientes consideraciones:

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[35]

[35] 1) 35 cuentas. 2) 246 camiones. 3) Falso. 4) Verdadero. 5) Verdadero.

De acuerdo a lo planteado en el enunciado de la situación profesional:

a. Observemos que se tienen los siguientes datos:

$n=80$

$x= 8,2$ hs.

$\sigma=2,8$ hs.

$z=1,645$

$\sigma x =2,8/\sqrt{80} =0,31$

El intervalo resulta $8,2 \pm 1,645(0,31)=8,2 \pm 0,50$

Por lo tanto, podemos decir que la media poblacional se encuentra entre 7,7 y 8,7 horas por día, con una certeza del 90%.

b. $p=60/80=0,75$ o un 75%

$Q= 1- 0,75= 0,25$ o 25%

$z= 1,645$

$\sigma p=\sqrt{(75.25)}/80 =4,84\%$

Por tanto, el porcentaje de gente de la muestra que descansa un día entre lunes y viernes, con una certeza del 90%, es de 4,84%.

SP14/ Ejercicio por resolver

Si este equipo deseara hacer el mismo estudio que se hizo sobre lo planteado en la situación profesional, pero sobre otra población:

a, ¿Qué tamaño de muestra se requiere si se desea tener un nivel de confianza del 90% con respecto a la cantidad de horas por día de descanso, con un error máximo de 1 hora y una σ=2 hs?

b, ¿Qué tamaño de muestra se necesita, si desea tener un nivel de confianza del 90% con un error máximo de 3,5%, para estimar correctamente el porcentaje de personas que descansan un día entre lunes y viernes, si no se tiene información previa?

Indique la opción correcta

1- La estimación se presenta cuando al parámetro le asignamos un valor dentro de un intervalo.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- El error de muestreo E es igual a la diferencia entre la media de la distribución de las medias y la media correspondiente a una de las muestras x.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- En la mayoría de los casos que se presentan, el valor del desvío estándar será desconocido pudiendo, entonces, estimarlo por un lado, tomando como valor de desvío estándar poblacional al desvío correspondiente a una muestra S y, por el otro, aproximando en forma subjetiva, basados en experiencias pasadas.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Estimar es el proceso de usar un estadístico muestral para calcular (estimar) el correspondiente valor del __________ desconocido.

- ☐ Desvío
- ☐ Parámetro poblacional

Indique la opción correcta

5- Se dice que se tiene una estimación de punto cuando al parámetro le asignamos __________.

- ☐ Dos posibles valores
- ☐ Un valor único

Indique la opción correcta

6- Se dice que se tiene estimación de intervalo cuando al parámetro le asignamos _______________.

- ☐ Un valor dentro de un intervalo
- ☐ Valores infinitos

Indique la opción correcta

7- Según el Teorema central del límite, sabemos que si consideramos la distribución de todas las muestras posibles que se pueden tomar de la población, de la misma cantidad de elementos cada una de ellas, y obteniendo de cada una la media, entonces la distribución de esas medias es una distribución normal con una media igual a ______.

- ☐ Uno
- ☐ La media poblacional μ

Indique la opción correcta

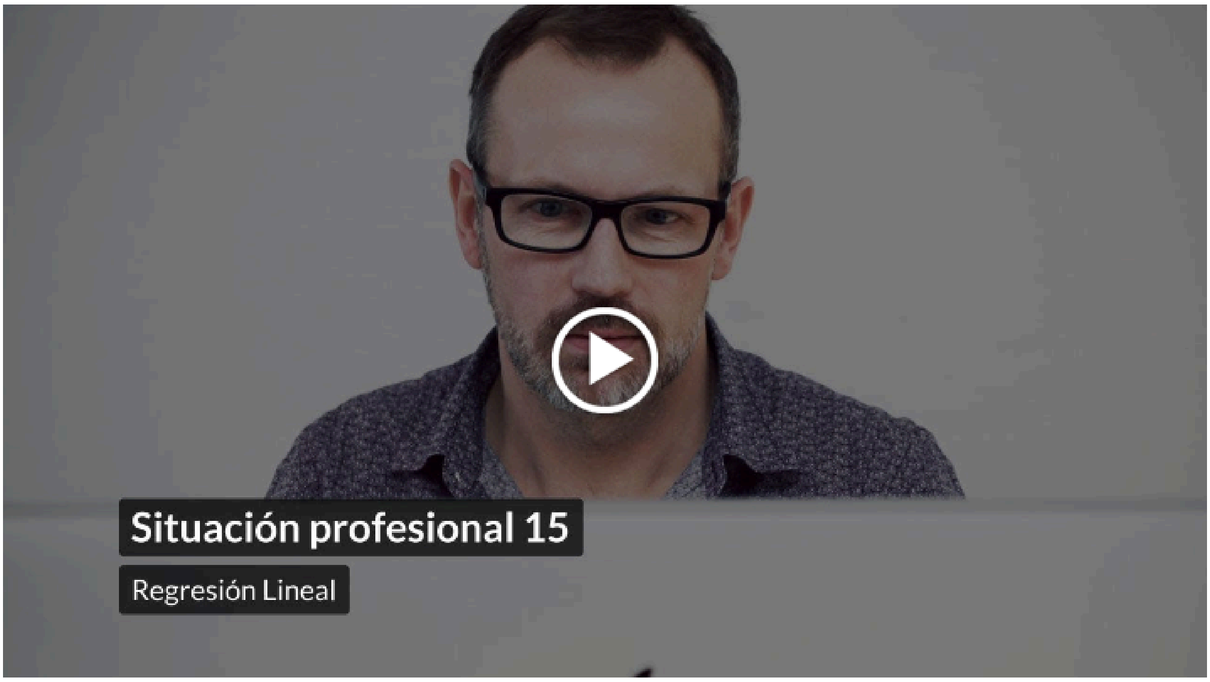
8- Se define como error de muestreo E a la diferencia entre __________.

- ☐ la media de la distribución de las medias y la media correspondiente a una de las muestras
- ☐ la media de la población

Respuestas correctas[36]

[36] 1) Falso. 2) Verdadero. 3) Falso. 4) Parámetro poblacional. 5) Un valor único. 6) Un valor dentro de un intervalo. 7) La media poblacional μ. 8) la media de la distribución de las medias y la media correspondiente a una de las muestras.

Situación profesional 15: Regresión Lineal



Link del video: https://www.youtube.com/watch?v=fl06oz1T_iE

Texto del video: Exequiel Juárez es el propietario de un nuevo emprendimiento en un pueblo pequeño, elabora y vende dulce de arándanos artesanal.

Dado que su empresa es bastante nueva, no conoce formalmente las cifras de sus costos, ingresos y beneficios. Usted lo

ayudará a determinar la cantidad de Kg. que debe vender por mes, si quiere tener un ingreso aproximado de $6.500. El Sr. Juárez registró los montos de sus ventas y los ingresos correspondientes, y obtuvo los siguientes resultados:

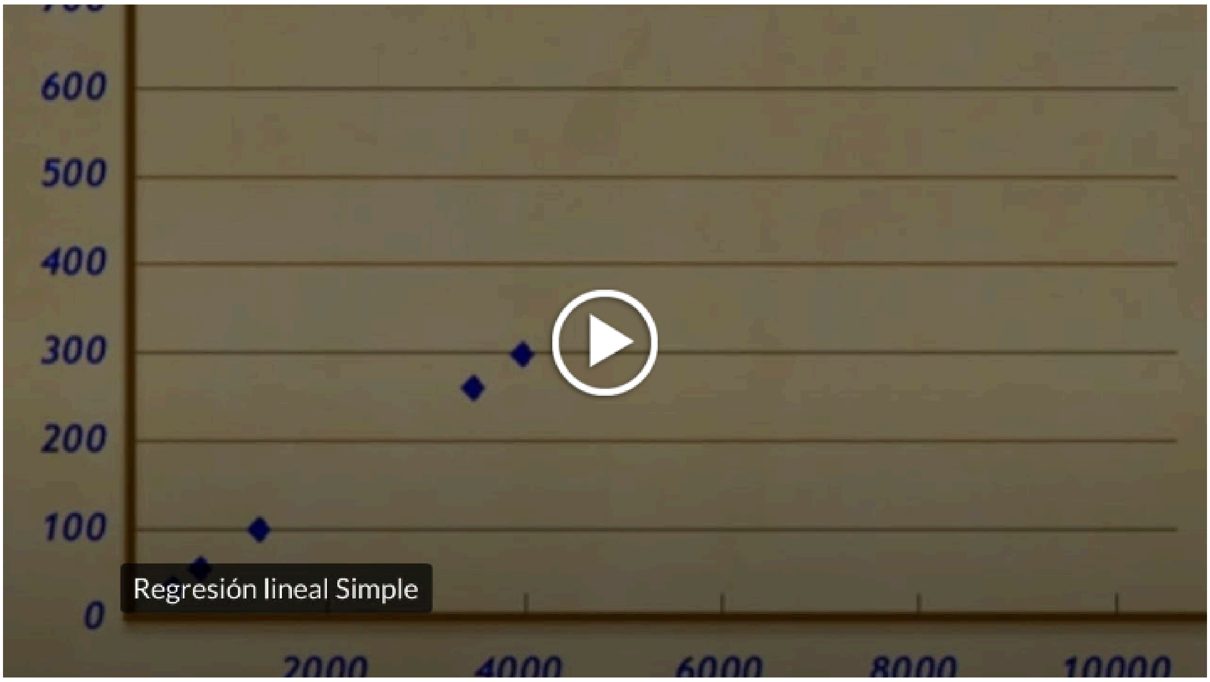
Ingreso (x)	Kg vendidos (y)
255	20
500	40
1253	100
3120	250
3745	300
5625	450
8750	700

SP15/H1: Análisis de regresión lineal simple

En esta situación profesional presentaremos una introducción a la regresión lineal simple que resulta, en casos muy puntuales, una herramienta interesante.

Hasta ahora hemos trabajado con situaciones en las cuales las variables analizadas son de tipo unidimensional (variables aleatorias únicas). La primera diferencia que se presenta aquí es el estudio de variables aleatorias bidimensionales, es decir, variables definidas por dos características, que pueden graficarse en un plano coordenado.

Para comenzar, debemos tener en cuenta que no siempre podrá utilizarse este método, ya que dependerá su uso de la forma en la cual se presentan los datos. Tendremos como objetivo poder ajustar a los datos una recta, que llamaremos Recta de Regresión Lineal Simple, lo cual sólo podrá realizarse cuando al presentar los datos, se sabe que los mismos tienen un comportamiento lineal. Esta recta se usa para estimar valores de una de las variables (la dependiente, que es motivo de estudio), en función de la otra (variable independiente, es la que condiciona a la variable dependiente).



Link del video: https://www.youtube.com/watch?v=I9weVPO-u-I

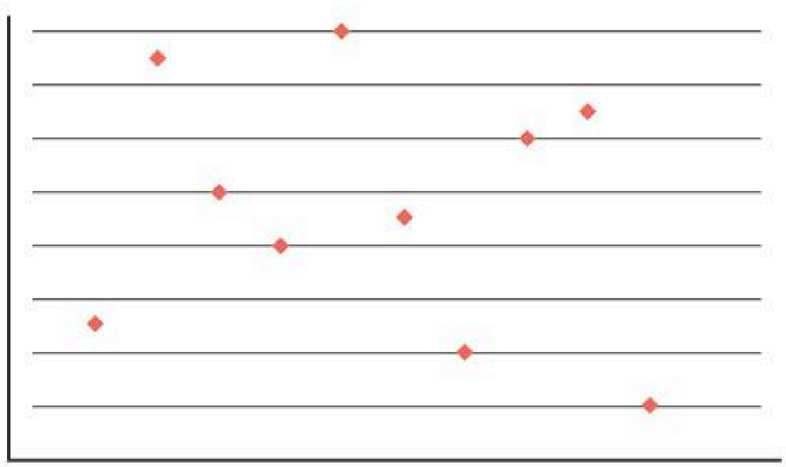
Texto del video: Tomaremos ahora el ejemplo de la situación profesional. Si observamos el gráfico de los kilogramos vendidos en función del ingreso conseguido, veremos que los puntos se encuentran distribuidos de manera creciente y de manera tal, que una recta podría ajustarse aproximadamente a los datos. Es decir, en este caso particular, se observa que cuanto más ingreso se quiere obtener, mayor será la cantidad del producto que debe venderse.

Volvamos sobre la noción de "ajustar" una recta; como dijimos, del gráfico se observa que una recta podría pasar muy próxima a los puntos dados. Esta condición, de que los puntos puedan ser aproximados por una recta, es condición necesaria para el análisis de regresión lineal simple. Cabe aclarar que deben cumplirse otras hipótesis para realizar un análisis completo, pero en este texto no profundizaremos dichas hipótesis.

El análisis de regresión lineal simple proviene de un método de optimización que se llama "Método de Cuadrados Mínimos", cuyo principio ya fue aplicado al calcular los desvíos cuadráticos medios, para calcular el desvío estándar. Se aplica este método ya que se pretende que la distancia de los puntos dados a la recta a conseguir, sea la mínima posible para que, justamente, la recta se ajuste "lo más posible".

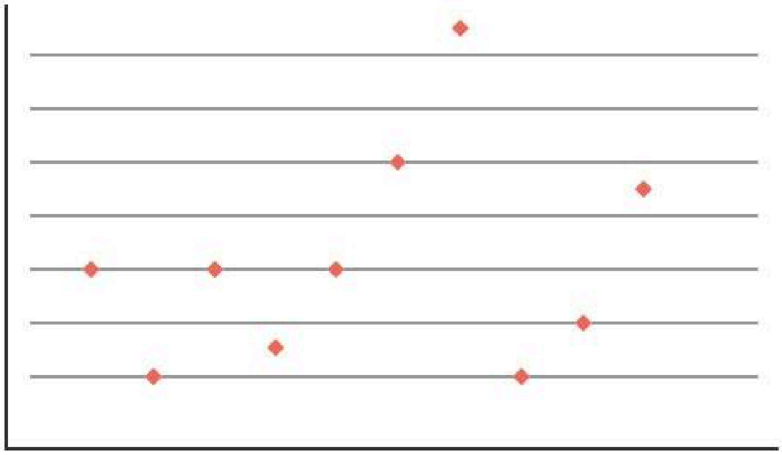
En los gráficos que se presentan a continuación, el lector podrá ver en qué tipos de casos no se puede utilizar un análisis de regresión lineal simple.

Gráfico A



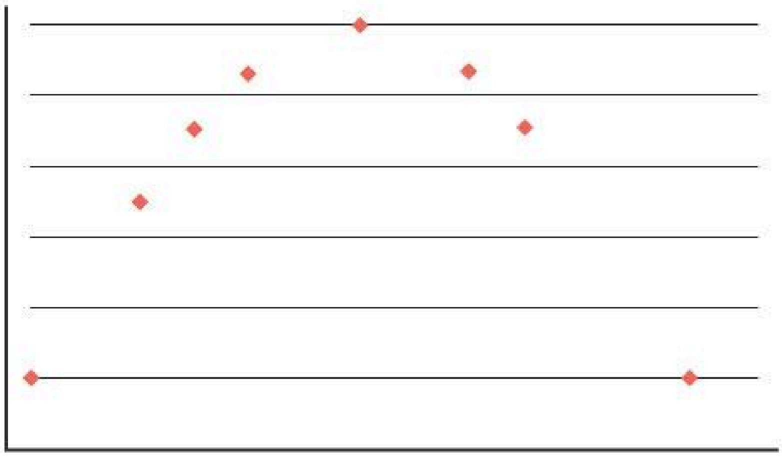
Los puntos están distribuidos de manera que una recta no podría aproximarlos.

Gráfico B



Los puntos se distribuyen de manera ondulante, y una recta no podría ajustarse.

Gráfico C



Los puntos describen una curva; por tanto una recta no podría aproximarlos.

La razón es sencilla: observe que en ninguno de estos casos es posible que la gráfica se asemeje a los puntos dispersos de una recta.

Lo que haremos a continuación, será encontrar la recta que mejor se aproxime a los puntos de nuestro ejemplo original. Tengamos en cuenta que deseamos estimar la cantidad de Kg. de producto que deben venderse para lograr un ingreso deseado. Si, sobre los puntos dados al principio, generamos una recta que se aproxime lo más posible a todos ellos, podremos, de alguna manera, estimar lo buscado en el problema. Obviamente, el valor encontrado, será, como dijimos, una "estimación", lo que quiere decir que el valor hallado será una estimación del valor real, que proviene, por supuesto, de estimar la recta que más se ajuste; pero la idea es que bajo las hipótesis del comienzo, esta estimación sea bastante razonable. La mayor o menor coincidencia entre el valor real y el estimado dependerá en gran medida de la mejor o peor alineación de los puntos sobre la recta.

Entonces, dados los puntos, y habiendo observado que una recta podría ajustarse aproximadamente al gráfico, intentamos determinar analíticamente dicha recta.

Recordemos que la ecuación de una recta se expresa: y= ax + b, donde a y b son constantes llamadas pendiente y ordenada al origen, respectivamente. La pendiente representa la inclinación de la recta, y la ordenada al origen es el punto donde la recta corta al eje y. Volvamos sobre los posibles valores de la pendiente; como dijimos se refiere a la inclinación de la recta; si el valor de a es positivo, se tendrá que la relación entre las dos variables x e y es creciente, y en caso de ser negativo, la relación será decreciente.

En nuestro caso, la variable x son los ingresos, y la variable y son los Kg vendidos.

¿Cómo encontraremos los coeficientes a y b de la recta?

El cálculo de los coeficientes se realiza de la siguiente forma:

$$a = \frac{\left(\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}\right)}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}$$

$$b = \overline{Y} - a\,.\,\overline{X}$$

Como queremos determinar la recta en nuestro ejemplo, necesitamos calcular los coeficientes a y b que determinarán la misma; por ello, construiremos una tabla con los valores necesarios para el cálculo. Esto es totalmente razonable, si se quiere organizar el trabajo, para construir las sumatorias.

Una vez confeccionada la tabla, podremos obtener con facilidad todos los elementos que componen el cálculo de a y b. Note que habiendo calculado el valor de a, el cálculo de b es muy sencillo.

La tabla se muestra a continuación:

Ingreso en pesos (x)	Kilogramos vendidos (y)	X^2	X.Y
255	20	65025	5100
500	40	250000	20000
1253	100	1570009	125300
3120	250	9734400	780000
3745	300	14025025	1123500
5625	450	31640625	2531250
8750	700	76562500	6125000

Además, tenemos que:

$$\overline{Y} = 1860 / 7 = 265{,}7$$

$$\overline{X} = 23248 / 7 = 3321{,}1$$

Por lo tanto, reemplazando las expresiones de a y b, se obtiene que

a = (10710150 - 7 . 265,7 . 3321,1) / (56639647,53)

= (4533236,11) / (56639647,53)

= 0,08

b = 265,7 - 0,08 . 3321

= 0,02

Estimación vs. Predicción

Existe una diferencia muy marcada entre los conceptos de estimación y predicción. Cuando determinamos una aproximación, a través del modelo, de valores que se encuentran dentro del intervalo en estudio, estaremos estimando el valor en cuestión; a diferencia de lo que se haría calculando para un valor fuera del intervalo en estudio, que sería predecir. Este modelo lineal resulta una buena aproximación, y arroja resultados razonables dentro del intervalo en estudio, no podemos asegurar lo mismo para valores fuera de él, ya que el comportamiento de los datos podría ser muy diferente fuera del mismo.

Exequiel Juárez es el propietario de un nuevo emprendimiento en un pueblo pequeño, elabora y vende dulce de arándanos artesanal. Dado que su empresa es bastante nueva, no conoce formalmente las cifras de sus costos, ingresos y beneficios. Usted lo ayudará a determinar la cantidad de Kg. que debe vender por mes, si quiere tener un ingreso aproximado de $6.500. El Sr. Juárez registró los montos de sus ventas y los ingresos correspondientes y obtuvo los siguientes resultados:

Ingreso (x)	Kg vendidos (y)
255	20
500	40
1253	100
3120	250
3745	300
5625	450
8750	700

Indique la opción correcta

1- Sobre la base del ejemplo de la situación profesional, estime el valor de Kg vendido para el siguiente ingreso: 300

- ☐ 24,02
- ☐ 24,04
- ☐ 24
- ☐ 24,03
- ☐ 24,05

Indique la opción correcta

2- Sobre la base del ejemplo de la situación profesional, estime el valor de Kg vendido para el siguiente ingreso: 750

- ☐ 70,02
- ☐ 60,02
- ☐ 50
- ☐ 60,5
- ☐ 61,5

Indique la opción correcta

3- Sobre la base del ejemplo de la situación profesional, estime el valor de Kg vendido para el siguiente ingreso: 3500

- ☐ 280,05
- ☐ 208,05
- ☐ 208,20
- ☐ 280,04
- ☐ 280,02

Indique la opción correcta

4- Sobre la base del ejemplo de la situación profesional, estime el valor de Kg vendido para el siguiente ingreso: 4000

- ☐ 332,07
- ☐ 342,04
- ☐ 320,02
- ☐ 300
- ☐ 308,09

Indique la opción correcta

5- Sobre la base del mismo situación profesional, calcule los ingresos conseguidos para las siguientes cantidades de Kg de producto vendido: 30

- ☐ 374,75
- ☐ 375
- ☐ 374,3
- ☐ 373,4
- ☐ 374,65

Indique la opción correcta

6- Sobre la base del mismo situación profesional, calcule los ingresos conseguidos para las siguientes cantidades de Kg de producto vendido: 50

- ☐ 624,75
- ☐ 624,74

- ☐ 624,73
- ☐ 624,76
- ☐ 624

Indique la opción correcta

7- Sobre la base del mismo ejemplo, calcule los ingresos conseguidos para las siguientes cantidades de Kg de producto vendido: 270

- ☐ 3271,20
- ☐ 3271,2
- ☐ 3374,39
- ☐ 3374,75
- ☐ 3374,5

Indique la opción correcta

8- Sobre la base del mismo ejemplo, calcule los ingresos conseguidos para las siguientes cantidades de Kg de producto vendido: 600

- ☐ 7499,19
- ☐ 7490,25
- ☐ 7499,33
- ☐ 7479,94
- ☐ 7499,75

Respuestas correctas[37]

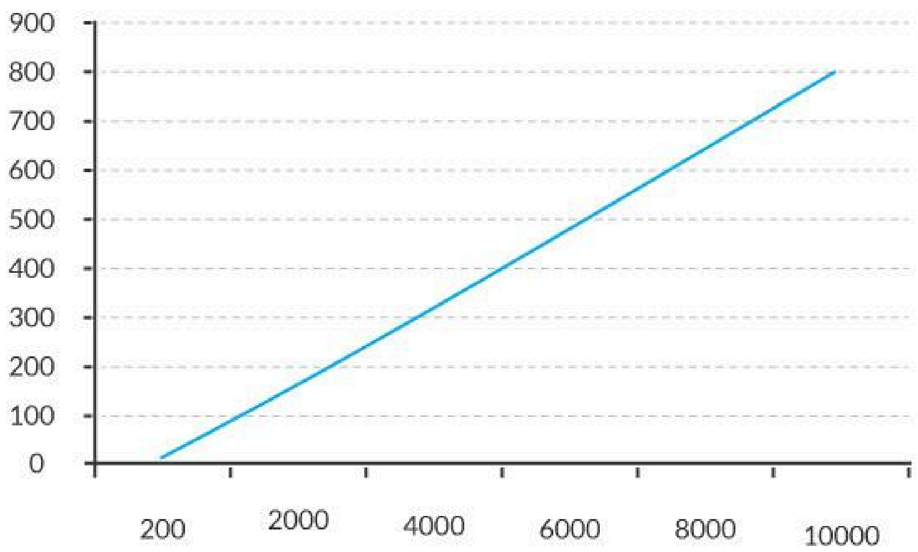
[37] 1) 24,02. 2) 60,02. 3) 280,02. 4) 320,02. 5) 374,75. 6) 624,75. 7) 3374,75. 8) 7499,75.

Pues bien, ya tenemos los coeficientes que nos determinan la ecuación de nuestra recta. Lo único que queda por hacer es reemplazar los valores que nos interesa conocer. Tomemos la incógnita de la situación profesional y supongamos que se pretendía tener como ingreso la suma de $6500.

La ecuación de la recta es Y = 0,08 X + 0,02

Por tanto, si reemplazamos en X el valor 6500, tendremos que los Kilogramos que deben venderse para tener el ingreso mencionado es Y = 0,08. 6500 + 0,02= 520,02 Kilogramos de dulce de arándanos

Así, podremos decir que la cantidad aproximada de Kilogramos de dulce que este productor debe vender para tener los ingresos deseados, es de aproximadamente 500 Kg. Veamos la situación en un gráfico que se presenta a continuación:



Analizaremos ahora el coeficiente a de la recta. Por lo que mencionamos antes, dado que el valor de la pendiente es positivo, se tiene un comportamiento creciente; es decir que cuanto más se vende, mayores serán los ingresos. Si la pendiente hubiera resultado negativa, podríamos haber asumido que esta producción no estaba arrojando en realidad ganancias para que el negocio se considere, a mayor producción, un negocio rentable.

SP15/ Ejercicio por resolver

Una empresa estadounidense realiza un estudio sobre adolescentes en riesgo de obesidad; para ello se registra a partir de un grupo de 8 adolescentes, el número de comidas rápidas que consumieron en la última semana cada uno de ellos, ya que se considera como factor fundamental en la obesidad de adolescentes, la cantidad de comida chatarra consumida. Usted trabaja en el equipo interdisciplinario que se encarga de realizar dicho estudio. Se supone que los datos se ajustan a un modelo lineal, y están expresados a continuación, medidos en libras (1 lb = 0.453) kg:

Sujeto	Número de comidas rápidas consumidas en la última semana	Peso (libras)			
	X	Y	X^2	Y^2	X . Y
1	2	112	4	12544	224
2	2	131	4	17161	262
3	5	171	25	29241	855
4	2	160	4	25600	320
5	4	182	16	33124	728
6	3	165	9	27225	495
7	3	149	9	22201	447
8	2	137	4	18769	274
	23	1207	75	185865	3605

A partir de estos datos:

a. Determine explícitamente la recta que se ajusta a este modelo.
b. Estime, a partir de lo encontrado, cuántos Kg. pesará un adolescente que consume 4 comidas rápidas semanales.

SP15/Evaluación de paso

Indique la opción correcta

1- El análisis de regresión lineal simple proviene de un método de optimización que se llama “Método de Cuadrados Mínimos”.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

2- La ecuación de una recta se expresa y= ax + b, donde a y b son constantes llamadas pendiente y ordenada al origen, respectivamente.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

3- Estimación y Predicción se refieren a lo mismo.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

4- Cuando determinamos una aproximación, a través del modelo, de valores que se encuentran dentro del intervalo en estudio, estamos haciendo una Predicción del valor en cuestión.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

5- Las variables aleatorias bidimensionales son variables definidas por dos características, que pueden graficarse en varios planos coordenados.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

6- En la regresión lineal, los datos analizados son unidimensionales.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

7- La regresión lineal intenta ajustar a los datos alguna curva .

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

8- Siempre que se quiera estimar puede usarse regresión lineal.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

9- Estimar y predecir son conceptos muy diferentes.

- ☐ Verdadero
- ☐ Falso

Indique la opción correcta

10- Los datos están dados por datos bidimensionales, donde las 2 variables son independientes.

- ☐ Verdadero
- ☐ Falso

Respuestas correctas[38]

[38] 1) Verdadero. 2) Verdadero. 3) Falso. 4) Falso. 5) Falso. 6) Falso. 7) Falso. 8) Falso. 9) Verdadero. 10) Falso.

Cierre

Este texto ha sido elaborado con mucha dedicación y esmero, a fin de cubrir ampliamente la necesidad de los lectores.

El mismo cuenta con quince situaciones profesionales que reflejan posibles situaciones de la vida real profesional y, por tanto, también las herramientas necesarias a fin de comprenderlas, y analizarlas desde la perspectiva de la probabilidad y la estadística. El objetivo es que ustedes sean capaces de resolver las situaciones planteadas de manera concreta. Además, este texto contiene gran cantidad de ejercicios en cada situación profesional, a los fines tanto de fijar conceptos, como repasar e integrar los mismos.

Esperamos que este trabajo resulte una herramienta de gran utilidad, y sea un apoyo en este proceso de aprendizaje.

Los autores

Bibliografía

- BERENSON–LEVINE, David: "Estadística Básica en Administración". Ed. Prentice Hall. Ed. 1996.
- CHOU: "Análisis Estadístico". Ed. Mc Graw Hill. Ed. 1990.
- SPIEGEL Murray: "Teoría y Problemas de Probabilidad y Estadística". Ed. Mc Graw Hill. Ed. 1991.
- YAMANE TARO: "Estadística". Ed. Harla Harper Row Latinoamericana Ed. 1979 Cuarta Edición.