Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO

Estadística Inferencial / Nelly Yolanda Cespedes Guevara, / Bogotá D.C., Fundación Universitaria del Área Andina. 2017

978-958-5459-83-0

Catalogación en la fuente Fundación Universitaria del Área Andina (Bogotá).

Edición:
Fondo editorial Areandino
Fundación Universitaria del Área Andina
Calle 71 11-14, Bogotá D.C., Colombia
Tel.: (57-1) 7 42 19 64 ext. 1228
E-mail: publicaciones@areandina.edu.co
http://www.areandina.edu.co

Primera edición: noviembre de 2017

Corrección de estilo, diagramación y edición: Dirección Nacional de Operaciones virtuales
Diseño y compilación electrónica: Dirección Nacional de Investigación

Hecho en Colombia
Made in Colombia

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina
MIEMBRO DE LA RED
ILUMNO

Índice

i

Índice

1

Unidad 1

Probabilidad

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO

Introducción

En esta cartilla se aborda el conocimiento teórico y práctico sobre las bases de la estadística inferencial a partir de la conformación de los enfoques de probabilidad.

Es importante leer cuidadosamente el contenido de la cartilla. Al interpretar las expresiones matemáticas, permitirá comprender los elementos expuestos para así aplicar los contenidos y asimilarlos.

Frecuentemente, para solucionar situaciones cotidianas, se requiere diferenciar los procedimientos de cálculo para resolver los diferentes contextos matemáticos.

U1

Metodología

El proceso de aprendizaje académico por medio de esta cartilla, se desarrolla en tres etapas. La primera etapa es la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido por medio de la revisión de los ejercicios resueltos con el fin de ganar confianza en la solución de estos propuestos. La segunda etapa es el análisis del video y el contenido del mismo y la tercera etapa, es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

Cabe anotar que la segunda etapa puede ser retomada cuantas veces quiera, pero tiene un tiempo limitado para adquirir los conocimientos. Recuerde que debe avanzar con las siguientes cartillas y tener en cuenta las recomendaciones dadas a lo largo del avance del curso.

U1 Desarrollo temático

Componente motivacional

Este tema nos sirve para solucionar situaciones reales o ideales de manera eficiente. Nos aporta herramientas teóricas para identificar apropiadamente los cálculos de probabilidad con el fin de entender los aspectos básicos de la estadística inferencial.

Por ejemplo, identificar las diferencias entre distribuciones muestrales con el fin de establecer los conceptos prácticos de la solución de ejercicios.

Recomendaciones académicas

La presente cartilla contiene la fundamentación teórica de los aspectos teóricos y prácticos sobre las distribuciones muestrales que permite evidenciar la solución de situaciones problémicas que conduzcan a la comprensión de ejercicios sobre los aspectos esenciales de la estadística inferencial.

La matemática financiera es más sencilla de lo que nos parece. Es la utilización y aplicación de normas o leyes denominadas propiedades. Cuando se aprende apropiadamente estas propiedades, se puede afrontar de manera correcta cada una de las situaciones o ejercicios que nos solicitan, por esto, la estadística no es difícil, lo difícil es comprender y aplicar las propiedades que orientan el desarrollo de la misma asignatura. Todos, de una manera u otra podemos llegar a manejar cada día mejor la matemática.

Desarrollo de cada una de las unidades temáticas

Probabilidad

La probabilidad es una de las ramas de la estadística que se convierte en una de las aplicaciones reales de contexto, en donde se busca trabajar con el tema de predicción para generar procesos de abstracción en el concepto de medida de posibilidades y ocurrencia de eventos.

Experimentos aleatorios

En cualquier situación cotidiana, la probabilidad se encuentra presente en cualquier esquema de conceptualización de procesos matemáticos de la vida real. Por ejemplo:

1. Lanzar un dado y verificar el número de veces que cae el número 4.
2. Lanzar una moneda y verificar si cae cada o sello.
3. Dejar caer un objeto desde una altura particular.

En los anteriores casos se puede predecir cuál va a ser el resultado, por lo tanto, este tipo de casos se conocen como deterministas. En un ejemplo de lanzamiento de dados se puede notar que no se conoce con certeza el número de cara que va a caer sino se hace un proceso de verificación después del lanzamiento.

Los experimentos que no se pueden predecir se conocen como experimentos aleatorios.

Definiciones básicas

1. Un espacio muestral se define como un conjunto de todos los posibles resultados de un experimento aleatorio cualquiera. Se designa con E (o bien por la letra griega omega Ω). A cada elemento que forma parte del espacio muestral se le denomina suceso elemental. (Estadística aplicada. Recuperado de http://www.iesramonolleros.es/matematicas/estapli/pdf/t5.pdf)

Ejemplo:

1. ¿Cuál es el espacio muestral asociado al experimento de lanzar un dado y observar cuál de sus caras queda hacia arriba? En este caso, hay 6 posibles resultados E = {1, 2, 3, 4, 5, 6}.

Tipos de sucesos o eventos

2. Suceso imposible: no tiene ningún elemento y lo representaremos por Ø.
3. Suceso seguro: formado por todos los posibles resultados (es decir, el espacio muestral).
4. Espacio muestral: se representa por S, y se refiere al conjunto de todos los sucesos aleatorios.

Ejemplo:

1. En el lanzamiento de la moneda tenemos que el espacio muestral era E= {A, B}, el espacio muestral es:

- Eventos con 0 elementos: Ø
- Eventos con 1 elemento: {A}, {B}
- Eventos con 2 elementos:{A, B}

S = {Ø, {C}, {X}, {C, X}}.

2. En el caso del lanzamiento de dos monedas, el espacio muestral es: (A es cara y B es sello).

E = {(A, A), (A, B), (B, A), (B,B}

Si se describen los demás elementos del espacio muestral tenemos:

a. Al menos una cara= {(A, A), (A, B), (B, A)}
b. Salen más caras que sellos = {(A, A)}
c. La moneda no cae en ninguna posición = Ø
d. No sale ningún sello = {(A, A)}

Propiedad:

Si el espacio muestral tiene n elementos, el espacio muestral tiene 2n elementos por cada evento.

Operaciones con eventos

1. Eventos iguales: dos eventos A y B son iguales si están compuestos por los mismos elementos, es decir, A = B.
2. Intersección de eventos: la intersección de los eventos A y B, se representa por A B, es decir, "ocurren A y B a la vez".

Ejemplo: al lanzar un dado, ya sabemos que el espacio muestral es E= {1, 2, 3, 4, 5, 6}. Sean los eventos A ="un nº par"= {2, 4, 6}, y B ="un número entre 2 y 4 (inclusive)"= {2, 3, 4}.

El evento A ∩ B es tal que ocurren A y B a la vez, es decir:

$A \cap B$ = "sacar un nº par y que esté entre 2 y 4"= {2,4}.

El evento $A \cap B$ son los elementos comunes a los conjuntos A y B (elementos que están en los dos conjuntos).

Unión de eventos: la unión de eventos A y B se representa por $A \cup B$ al suceso "ocurre A o bien ocurre B o bien ocurren ambos a la vez" (también podemos decir que "ocurre alguno"). Es decir $A \cup B$ son los elementos que están en ambos conjuntos (aunque no necesariamente en los dos a la vez).

La expresión matemática es:

$A \cap B \subset A \cup B$

(El símbolo $\subset$ significa "contenido", es decir, el primer conjunto es un subconjunto del segundo).

Evento contrario: dado un evento A, denominaremos evento contrario de A y se representaría por $A \subset$ (o bien $A \subset$ o bien Ac) al suceso que tiene por elementos a todos aquellos que no pertenecen a A.

Diferencia de eventos: si A y B son dos eventos, A – B y B–A, que consta de los elementos que están en B pero no están en A. Por ejemplo, si A = {2, 4, 6}, B = {2, 3, 4}, tenemos que B – A= {3}. Se cumple que $B - A = B - A \cap B$, y también que $B - A = A \cap B$.

Propiedades de las operaciones con eventos

Las operaciones con sucesos tienen las siguientes propiedades, la mayoría de ellas bien conocidas:

Conmutativa $A \cap B = B \cap A$ $A \cup B = B \cup A$

Asociativa $A \cap (B \cap C) = (A \cap B) \cap C$ $A \cup (B \cup C) = (A \cup B) \cup C$

Idempotente $A \cap A = A$ $A \cup A = A$

Simplificación $A \cap (A \cup B) = A$ $A \cup (A \cap B) = A$

Distributiva $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Elemento neutro $A \cap E = A$ $A \cup \emptyset = A$

Absorción $A \cup \emptyset = A$ $A \cup E = A$ (no se puede cambiar teoría de probabilidad)

Leyes de De Morgan

Si A y B son dos sucesos, se verifican:

$(A \cup B) = A \cap B$

$(A \cap B) = A \cup B$

Regla de Laplace

Si realizamos un experimento aleatorio en el que hay n eventos elementales, todos igualmente probables, entonces si A es un evento, la probabilidad de que ocurra el evento A es: p(A) = número de casos favorables al evento A número de casos posibles.

Definición axiomática de probabilidad

Una probabilidad p es una función que asocia a cada evento A del espacio muestral S, un número real p(A), es decir: $p: S \longrightarrow R$, y que cumple las propiedades:

1. $0 \leq p(A) \leq 1$, (es decir, cualquier suceso tiene probabilidad positiva y menor o igual que 1).
2. $p(E) = 1$ (la probabilidad del suceso seguro es 1).
3. Si A y B son incompatibles, es decir A

$\cap$ B = ∅, entonces p(A ∪ B) = p(A) + p (B). (Es decir la probabilidad de la unión es la suma de las probabilidades si los eventos tienen intersección vacía).

Propiedad:

Si w1, *w2,. . ., wn* son los n eventos elementales de un evento aleatorio cualquiera, p una función p: S —→ R de modo que cumple las propiedades:

1. $0 \leq p\ (w) \leq 1\ \forall\ i \in \{1, 2, \ldots, n\}$
2. *p* (w1) + *p*(w2) + . . .+ p(wn) = 1

Entonces p es una probabilidad. (Estadística aplicada. Recuperado de http://www.iesramonolleros.es/matematicas/estapli/pdf/t5.pdf)

Ejemplo: Comprobar si las siguientes funciones mostradas en los eventos son probabilidad, siendo E = {a, b, c, d} el espacio muestral del experimento aleatorio:

a. *p*(a) = 1/2

b. *p*(b) = 1/3

c. *p*(c) = 1/4

d. *p*(d) = 1/5

Se cumple la probababilidad ya que cada probabilidad es inferior a 1.

La segunda propiedad no se cumple, porque la suma es diferente de 1

½+1/3+1/4+1/5 = 77/60

Definición de probabilidad

1. $p(\overline{A})=1-p(A)$

   En efecto, puesto que $E = A \cup \overline{A}$ y además A y $\overline{A}$ son incompatibles, resulta por la propiedad 3) de la definición que $p(E) = p(A \cup \overline{A}) = p(A) + p(\overline{A})$ Y por la propiedad 2), p(E)=1, luego $1 = p(A) + p(\overline{A})$ y por tanto $p(\overline{A}) = 1 - p(A)$.

2. $p\ (\overline{A}) = 0$

   Como E = ∅, resulta que:

   $p(\overline{E}) = p\ (\emptyset) = 1 - p\ (E) = 1 - 1 = 0$

3. Si A y B son dos eventos cualesquiera,

   $p\ (A \cup B) = p(A) + p(B) - p(A \cap B)$

4. Si A, B y C son tres eventos cualesquiera,

   $p\ (A \cup B \cup C) = p(A) + p(B) + p(C) - p(A \cap B) - p(A \cap C) - p(B \cap C) + p(A \cap B \cap C)$

Probabilidad condicionada

Definición:

Sea A un evento aleatorio asociado a un experimento aleatorio, y sea B otro evento que sabemos que se ha realizado. La probabilidad de A condicionada a B y se expresa por p(A/B) a la expresión: p(A/B) = p(A ∩ B) p (B) (de idéntico modo se define p(B/A),).

Teorema de la probabilidad total

Un experimento compuesto es aquel que consta de dos o más experimentos aleatorios simples. Es decir, si tiramos un dado, o una moneda, son experimentos aleatorios simples, pero si realizamos el experimento de tirar un dado y posteriormente una moneda, es un experimento compuesto.

Propiedad:

De la fórmula para calcular la probabilidad condicionada se deduce inmediatamente que:

p(A ∩ B) = p (B)• p(A/B) y p(A ∩ B) = p(A)• p(B/A)

Teorema de la probabilidad total

Si *A1, A2, . . .,* An son eventos incompatibles 2 a 2, y cuya unión es el espacio muestral (A1 ∪ A2 ∪ . . . ∪ An = E), y B es otro evento, resulta que: *p(B) = p(A1) • p(B/A1) + p(A2) • p(B/A2) + . . .+ p(An) • p(B/An).*

Teorema de Bayes

Si *A1, A2*. . ., An son sucesos incompatibles 2 a 2, y cuya unión es el espacio muestral (A1 ∪ A2 ∪. . .∪ An = E), y B es otro suceso, resulta que:

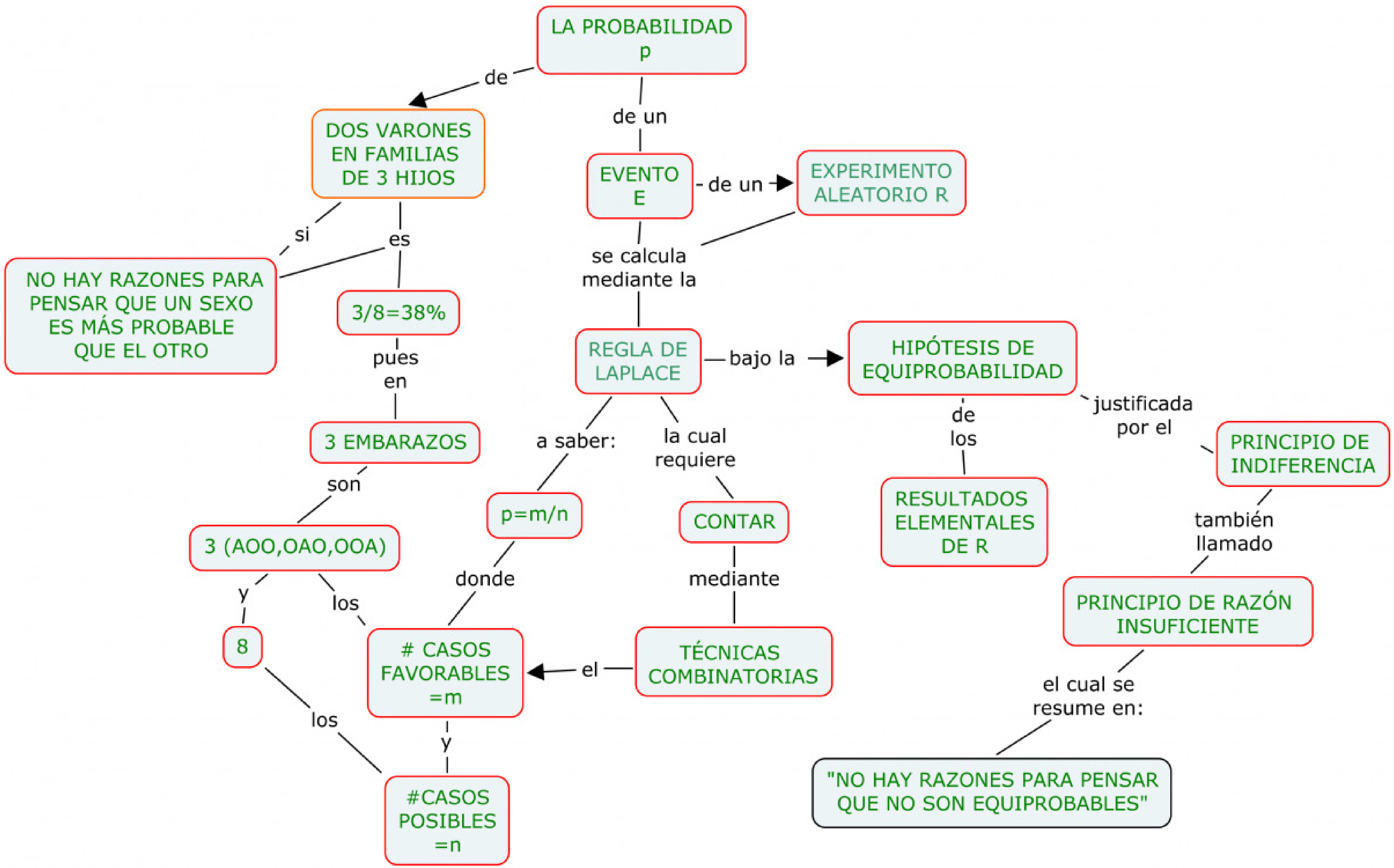
P (Ai/B) = p (Ai) •p (B/Ai)

P (A1) •p(B/A1) + p(A2)• p(B/A2) + . . .+ p(An) •p(B/An). (Probabilidad. Recuperado de http://sauce.pntic.mec.es/~jpeo0002/Archivos/PDF/T02.pdf)

Ejemplos, ejercicios o casos de aplicación práctica

1. Consideremos el experimento lanzar dos monedas al aire. Calcular la probabilidad del suceso sacar una cara y una sello
2. Calcula la probabilidad de obtener dos veces 4 al lanzar dos dados.
3. Se extrae una bola de una bolsa que contiene 4 bolas blancas, 5 rojas y 2 negras. ¿Cuál es la probabilidad de que no sea negra?
4. En una baraja hemos suprimido varias cartas. Entre las cartas que nos quedan se dan las siguientes probabilidades de ser extraídas: p(R) = 0,15, p (B) = 0,3, p (carta que no sea ni rey ni corazón) = 0,6. ¿Está entre ellas el rey de corazón? En caso afirmativo calcula su probabilidad.
5. En una determinada población, el 80% son aficionados al fútbol, el 50% al tenis y el 55% al baloncesto. El 35% lo son al fútbol y al tenis, el 50% al tenis y al baloncesto y el 60% al futbol y al baloncesto, mientras que el 20% lo son a los tres deportes. ¿Cuál es la probabilidad de que un individuo escogido al azar no sea aficionado a ninguno de los tres deportes?
6. Lanzamos una moneda hasta observar la segunda cara. ¿Cuál es la probabilidad de observar dos cruces antes de que se observe la segunda cara?
7. Calcular la probabilidad de obtener un 4 de diamante ó un rey de trébol al extraer una carta de una baraja.
8. Una urna contiene tres bolas rojas y dos verdes y otra contiene dos bolas rojas y tres verdes. Se toma, al azar, una bola de cada urna. ¿Cuál es la probabilidad de que ambas sean del mismo color?, ¿y la de que sean de distinto color?
9. Supongamos un dado cuyas caras pares son de color negro y las impares de color rojo. Se lanza el dado y la cara obtenida es de color negro. ¿Cuál es la probabilidad de que salga un 5?, ¿y de que salga un 4? Jusitifca la respuesta.
10. En una clase estudian bastante el 70%, y el resto estudian muy poco. De los alumnos que estudian bastante aprueba el 50%, y de los que estudian muy poco sólo aprueba el 3%. Después de hacer el examen se eligió al azar un alumno y resultó que había suspendido. Determinar la probabilidad de que hubiera estudiado bastante. (Ejercicios de probabilidad. Recuperado de http://actividadesinfor.webcindario.com/probabilidad.htm).

Síntesis de cierre del tema



(Competencias matemáticas estocásticas. Recuperado de http://www.matetam.com/blog/entradas-jmd/competencias-matematicas-estocasticas?page=3)

Actividades auto-evaluativas propuestas al estudiante

(Problemas de probabilidad condicionada. Recuperado de http://www.vitutor.com/pro/2/a_h.html)

1. En un colegio los estudiantes pueden optar por cursar como lengua extranjera inglés o francés. En un determinado curso, el 80% de los alumnos estudia inglés y el resto francés. El 20% de los que estudian inglés son chicos y de los que estudian francés son chicos el 50%. El elegido un alumno al azar, ¿cuál es la probabilidad de que sea chica?
2. Para la preparación de un examen un estudiante sólo ha estudiado 16 de los 30 temas correspondientes a la materia del mismo. Esta prueba se realiza extrayendo al azar dos temas y dejando que el estudiante escoja uno de los dos para ser examinado del mismo. Hallar la probabilidad de que el estudiante pueda elegir en el examen uno de los temas estudiados.
3. Una bolsa contiene tres monedas. Una moneda es normal, otra tiene dos caras y la otra está cargada de modo que la probabilidad de obtener cara es de 1/3. Se selecciona una moneda lanzar y se lanza al aire. Hallar la probabilidad de que salga sello.
4. Para levantarse temprano un estudiante cuenta, con la ayuda de un despertador, el cual consigue despertarlo en un 70% de los casos. Si oye el despertador, la probabilidad de que se levante temprano es 0.9 y, en caso contrario, de 0.5.
5. Si va a levantarse temprano, ¿cuál es la probabilidad de que haya oído el despertador?
6. Si no se levanta temprano, ¿cuál es la probabilidad de que no haya oído el despertador?
7. En una bolsa hay 4 fichas rosadas y 8 verdes. ¿Cuál es la probabilidad de que al sacar 2 fichas con reposición, estas sean amarillas?
8. Un estudiante contesta al azar 80 preguntas de un examen, si cada pregunta tiene 4 alternativas y solo una de ellas es correcta. ¿cuál es la probabilidad de que saque el puntaje máximo?
9. En una empresa trabajan hombres y mujeres, además se sabe que un 25% de los empleados se han perfeccionado en el extranjero. Si el 45% de las son mujeres, ¿cuál es la probabilidad de que al escoger una persona de la empresa, esta sea mujer y se haya perfeccionado en el extranjero?
10. Un restaurante ofrece un almuerzo en que se pueden elegir 3 entradas, 4 platos fuerte y 5 postres. Si no me gustan 2 de los platos fuertes y 3 de los postres. ¿Cuál es la probabilidad de que el menú sea de mi agrado si la elección es al azar?
11. Un estudiante responde al azar 4 preguntas de verdadero y falso en un examen. ¿Cuál es la probabilidad de que acierte todas las preguntas?
12. Calcular la probabilidad de, en una carrera de 23 caballos, acertar los 4 que quedan primeros (sin importar cuál de ellos queda primero, cual segundo y cual tercero).

Unidad 1

Distribuciones de probabilidad

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO

Introducción

En esta cartilla se abordan las distribuciones muestrales para grupos de datos, teniendo en cuenta los cálculos de la distribución para la media, la proporción y las diferencias aplicando el método de la estadística inferencial.

Es importante leer cuidadosamente el contenido de la cartilla. Al interpretar cada una de las expresiones, le permite comprender los elementos expuestos para así aplicar los contenidos y asimilarlos.

Frecuentemente, para solucionar situaciones cotidianas, se requiere diferenciar los procedimientos de cálculo para resolver las diferentes contextos matemáticos.

U1

Metodología

El proceso de aprendizaje académico por medio de esta cartilla se desarrolla en tres etapas. La primera etapa es la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido por medio de la revisión de los ejercicios resueltos con el fin de solucionar los ejercicios propuestos. La segunda etapa es el análisis del video y el contenido del mismo. La tercera etapa es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

La segunda etapa puede ser retomada cuantas veces quiera, pero tiene un tiempo limitado para adquirir los conocimientos. Recuerde que debe avanzar con las siguientes cartillas y tener en cuenta las recomendaciones dadas a lo largo del avance del curso.

U1

Desarrollo temático

Componente motivacional

Este tema sirve para solucionar situaciones reales o ideales de manera eficiente, aporta herramientas teóricas para identificar apropiadamente los cálculos de la distribución de la probabilidad en la media, la varianza y la desviación estándar.

Por ejemplo, identificar las diferencias entre las distribuciones de probabilidad con el fin de establecer los conceptos prácticos de la solución de ejercicios.

Recomendaciones académicas

La presente cartilla contiene la fundamentación teórica de los aspectos teóricos y prácticos sobre la distribución de probabilidad que permite evidenciar la solución de situaciones problémicas que conduzcan a la comprensión de ejercicios sobre los aspectos esenciales de la estadística inferencial.

La estadística inferencial es más sencilla de lo que nos parece. Es la utilización y aplicación de normas o leyes denominadas propiedades. Cuando se aprende apropiadamente estas propiedades, se puede afrontar de manera correcta cada una de las situaciones o ejercicios que se pueden presentar, por esto la matemática no es difícil lo difícil es comprender y aplicar las propiedades que orientan el desarrollo de la misma asignatura. Todos, de una manera u otra podemos llegar a manejar, cada día mejor, la matemática.

Desarrollo de cada una de las unidades temáticas

Competencia general	Competencia específica
El estudiante reconoce las características de las relaciones de las distribuciones de probabilidad en la estadística inferencial.	El estudiante diferencia las características de la distribución de probabilidad en diferentes contextos.
El estudiante relaciona las características que contribuyen a la planeación, organización, dirección, coordinación, control y evaluación de la estadística inferencial.	El estudiante calcula la distribución de probabilidad a partir de los promedios, la proporción y las diferencias estadísticas.

Distribuciones de probabilidad

Variables aleatorias

Una variable aleatoria se reconoce por ser parte de un experimento aleatorio. Usualmente se representa por las últimas letras del alfabeto: X, Y o Z.

Una variable aleatoria X es una función cuyo dominio es la colección de eventos del espacio muestral S y tiene por rango R, el subconjunto de los números reales.

Algunos ejemplos de variables aleatorias son:

X: el número de caras que se pueden obtener al lanzar los dados.

Y: el número de veces que se puede obtener cara o sello al lanzar una moneda.

Z: el número de palabras subrayadas en un libro.

Ejemplo

De una caja que contiene 5 pelotas numeradas del 1 al 5 se extraen 3 bolas una por una. Entonces **Z:** es el número mayor que se puede obtener después de sacar 3 veces una pelota.

Aquí el espacio muestral S es:

S = {(1,2,3), (1,2,4), (1,2,5), (1,3,4), (1,3,5), (1,4,5), (2,3,4), (2,3,5), (2,4,5), (3,4,5)} y la variable aleatoria Z asume los valores: 3, 4 y 5.

Las variables aleatorias permiten introducir la notación matemática de probabilidad, con el fin de describir el número de elementos que tiene un evento.

Si se define el rango de valores R de cualquier variable aleatoria como finito o infinito enumerable se considera variable aleatoria discreta y si no se puede enumerar se dice que es variable aleatoria contínua.

Ejemplo:

En un experimento se lanzan tres monedas y el resultado obtenido es:

S = [(c,c,c); (c,c,s); (c,s,c); (s,c,c); (c,s,s); (s,c,s); (s,s,c); (s,s,s)]

Entonces, los resultados del espacio muestral, se representa con una variable aleatoria, que hace referencia al número de veces que una de las monedas cae cara o sello, siendo enumeradas por un número natural cualquiera 0, 1, 2, 3, etc.

Las variables aleatorias permiten establecer una correspondencia general con cada uno de los elementos del espacio muestral, así pues:

Z = variable aleatoria del espacio muestral S.

A = elemento de cada evento en el espacio muestral.

En este sentido, se puede definir la probabilidad de una variable discreta, como:

Entonces, P (A = a) representa la probabilidad que la variable A tome en el valor a.

La definición de la probabilidad para una variable aleatoria A, se puede reconocer por medio de las siguientes propiedades:

1. Los valores de probabilidad no pueden tomar valores negativos.
2. La suma de todos los valores de probabilidad deben ser igual a 1.

Ejemplo:

En el experimento aleatorio de lanzar tres monedas y obtener cara (c) o sello (s). Calcule la distribución de probabilidad a través de la representación de los datos en una tabla que muestre la variable aleatoria:

S = [(c,c,c); (c,c,s); (c,s,c); (s,c,c); (c,s,s); (s,c,s); (s,s,c); (s,s,s)]

e(S)	A	P(A)
c,c,c	0	1/8
C,c,s	1	3/8
C,s,c	1	
S,c,c	1	
C,s,s	2	3/8
S,c,s	2	
S,s,c	2	
S,s,s	3	1/8

Tabla 1. Variables aleatorias
Fuente: propia.

Función de probabilidad de una variable aleatoria discreta

La función de probabilidad para una variable aleatoria se puede escribir como (x) = P [X = x], con las propiedades. (Rodríguez, L. (2007). Probabilidad y estadística básica para ingenieros. Recuperado de http://archuto.files.wordpress.com/2011/02/probabilidad_y_estadistica_basica.pdf))

i) $p(x) > 0$ y

ii) $p(x) = 1$

Ejemplo

Hallar la función de probabilidad de la variable del ejemplo anterior

Solución:

La probabilidad $P(x)$ se puede escribir como:

X	P(X)
3	1/10
4	3/10
5	6/10

Función de distribución acumulativa

La función de distribución acumulativa se define por:

$F(t) = (P \leq t) = \sum p(x)$

Siendo t cualquier número real. En particular, t es un valor que está en Rx, el cual consiste de enteros no negativos, entonces:

$F(t) = p(0) + p(1) + p(2) + p(3) + \ldots + p(t)$

Propiedad

La función de distribución de probabilidad y la distribución acumulativa se pueden relacionar así:

$p(x) = F(x) - F(x-1)$

para todo valor de una variable aleatoria X.

Varianza de una variable discreta

La varianza es una medida estadística que determina el nivel de dispersión en torno a una variable teniendo en cuenta el valor del promedio.

La expresión matemática es:

σ2 = V(X) = E (X – μ)2 para una variable aleatoria X de un experimento aleatorio con n eventos.

Teniendo en cuenta, el ejemplo realizado con el experimento aleatorio podemos calcular la varianza así:

Si se lanzan tres monedas y se desea obtener el resultado de cara (c) o sello (s). Calcule la distribución de probabilidad a través de la representación de los datos en una tabla que muestre la variable aleatoria:

S = [(c,c,c); (c,c,s); (c,s,c); (s,c,c); (c,s,s); (s,c,s); (s,s,c); (s,s,s)]

e(S)	A	P(A)
c,c,c	0	1/8
C,c,s	1	3/8
C,s,c	1	
S,c,c	1	
C,s,s	2	3/8
S,c,s	2	
S,s,c	2	
S,s,s	3	1/8

Tabla 1. Variable discreta
Fuente: propia.

Propiedades de la varianza V(X)

Las propiedades de la varianza permiten establecer parametros de relación entre las varaibales discretas, de tal modo que se cumple:

V(X) = variable aleatoria X.

f(X) = función de distribución de probabilidad.

a, b = números reales.

V (aX + b) = a²V(X)

Distribución binomial

La distribución binomial hace referencia las variables p y q, p se denomina éxito y q es fracaso, se asocia con el experimento de Bernoulli, ya que las repeticiones de los experimentos son independientes entre sí.

La función de probabilidad de una binomial es de la forma:

$$p(x) = \left(\frac{n}{x}\right) p^{x} (1-p)^{n-x}$$

para x = 0, 1, …,*n*.

Propiedades de la distribución binomial

a. Se calcula la media: μ = *np*
b. Se puede hallar la varianza: $\sigma^2 = npq$
c. Para calcular el coeficiente de sesgo:

$$\sigma^3 = \frac{q-p}{\sqrt{npq}}$$

d. La desviación típica: $\sigma = \sqrt{npq}$
e. Si p es menor que *0.5,* la distribución binomial está sesgada hacia la derecha.
f. conforme p aumenta, el sesgo es menos notable.
g. Si *p = 0.5,* la distribución binomial es simétrica.

h. Si p es mayor que *0.5,* la distribución esta sesgada hacia la izquierda.

Distribución de Poisson

La distribución de Poisson se calcula cuando el número de intentos es grande ($n \geq 100$) y $\mu = np$ es pequeño ($\mu < 10$), entonces la formula de la distribución de Poisson se escribe como:

P(x) = λx e-λ / x!

Donde el parámetro $\lambda > 0$, $e = 2.71828$ y $x = 1, 2, 3$; en donde se pueden aproximar a las probabilidades binomiales.

Ejemplo:

Si la probabilidad de que un avión comercial choque durante el vuelo es de 0.0001. ¿Cuál es la probabilidad de que choquen 10 de los siguientes 30000 aviones?

P(x) = λx e-λ / x! = P (10) = (3)10 e-3 / 10! = 0.00081

Ejemplos, ejercicios o casos de aplicación práctica

1. La probabilidad de tener un objeto defectuoso en una línea de ensamblaje es de 0.05. Si el conjunto de objetos terminados constituye un conjunto de ensayos independientes. ¿Cuál es la probabilidad de que entre diez unidades dos se encuentren defectuosas?
2. Una empresa de componentes mecánicos comprueba que el número de componentes que fallan antes de cumplir 50 horas de funcionamiento es una variable aleatoria de Poisson. Si el número promedio de estos fallos es 7, ¿cuál es la probabilidad de que falle un componente en 30 horas?
3. Cada muestra de agua tiene 20% de posibilidades de contener una molécula rara particular. Suponga que las muestras son independientes con respecto a la presencia de la molécula rara. Encuentre la probabilidad de que en las siguientes 50muestras, exactamente 3 contengan la molécula rara.
4. Un lote contiene 400 piezas de un proveedor de tejas y 250 unidades de un proveedor de tubería del estado vecino. Si se seleccionan cinco piezas al azar y sin reemplazo, ¿cuál es la probabilidad de que todas sean del proveedor local?
5. Una fábrica tiene tres máquinas que fabrican cuchillos. La máquina A produce el 60% del total de tornillos, la máquina B el 40% y la C el 10%. De la máquina A salen un 15% de cuchillos defectuosos, de la B un 3% y de la C un 5%. Calcule la probabilidad de que un cuchillo elegido al azar sea defectuoso.
6. Se lanzan tres monedas y se cuenta el número de caras obtenidas. Presente la información en una una tabla con las probabilidades, represéntala gráficamente y calcula la media y la desviación estándar.
7. Una caja contiene 8 mesas, de los cuales están 2 están defectuosas. Se selecciona una mesa de la caja y se prueba. Si este sale defectuosa se selecciona y se prueba otra mesa, hasta que se escoja una mesa no defectuosa. Encuentre el número esperado E de mesas seleccionadas.

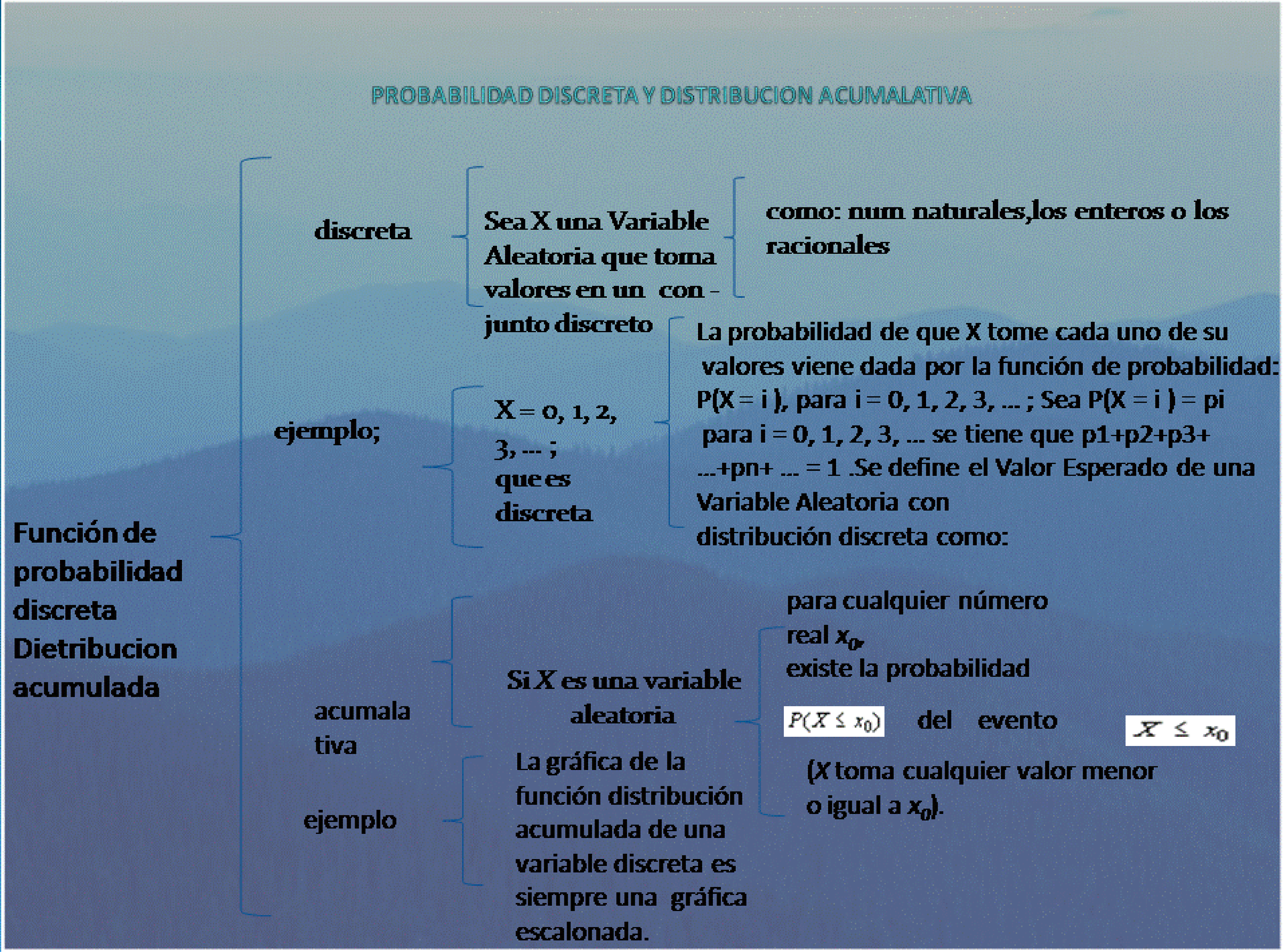


Imagen 1.Probabilidad discreta y distribución acumulativa. Recuperado de http://laobraperfectamz.blogspot.com/2010/03/probabilidad-discreta-y-distribucion.html

Actividades auto-evaluativas propuestas al estudiante

1. Una caja, H, contiene tres tarjetas con los números 1, 2 y 3, respectivamente. Otra caja B, tiene dos tarjetas, con los números 4 y 5. Elegimos una caja al azar, extraemos una tarjeta y miramos el número obtenido. Calcula la media y la desviación estándar.
2. Se sabe que el 40 % de la población de una determinada ciudad ve un concurso que hay en televisión. Se llama por teléfono a 15 personas de esa ciudad elegidas al azar. Calcula la probabilidad de que, entre esas 15 personas, estuvieran viendo el programa. Más de 7. Alguna de las 15.
3. La probabilidad de que una bolsa con huevos se rompa cuando es transportada es del 3%. Si se transportan 20 bolsas de huevos, calcula la probabilidad de que: Se rompan más de dos. No se rompa ninguno.
4. En un estudio realizado por una dependencia de transporte y vialidad, se encontró que el 75% de los vehículos de transporte colectivo, tienen alguna irregularidad de acuerdo al reglamento de tránsito. Si se someten 10 de estos vehículos a revisión, hallar la probabilidad. Exactamente 4 tengan alguna irregularidad.
5. Una máquina produce partes de automóviles de las cuales 0.6% son defectuosas. Si una muestra aleatoria de 20 partes producidas por ésta máquina contiene más de una pieza defectuosa, la máquina se apaga para repararse. Calcule la probabilidad de que la máquina sea apagada por reparaciones basadas en este plan de trabajo.
6. De acuerdo a una publicación Especializada los accidentes laborales causan un 58% de las ejecuciones hipotecarias. Dadas 15 ejecuciones auditadas por un banco, cual es la probabilidad de que menos de 6 se deban a accidentes laborales
7. Suponga que toma una prueba de 4 preguntas de múltiple opción adivinando. Cada pregunta tiene 5 opciones y sólo una es correcta. ¿Cuál es la probabilidad de que adivine más de la mitad correctamente? ¿Cuál es la probabilidad de que la segunda respuesta sea correcta?

Unidad 2

Distribución normal

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO

Introducción

En esta cartilla se aborda el conocimiento teórico y práctico sobre las bases de la matemática financiera a partir del cálculo de la distribución de probabilidad normal utilizadas en cualquier ejercicio o problema de carácter estadístico.

Es importante leer cuidadosamente el contenido de la cartilla. Al interpretar cada una de las expresiones, le permite comprender los elementos expuestos para así aplicar los contenidos y asimilarlos.

Frecuentemente, para solucionar situaciones cotidianas, se requiere diferenciar los procedimientos de cálculo para resolver los diferentes contextos matemáticos.

U2

Metodología

El proceso de aprendizaje académico por medio de esta cartilla se desarrolla en tres etapas. La primera etapa es la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido por medio de la revisión de los ejercicios resueltos con el fin de ganar confianza en su solución. La segunda etapa es el análisis del video y el contenido del mismo. La tercera etapa es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

Cabe anotar que la segunda etapa puede ser retomada cuantas veces quiera, pero tiene un tiempo limitado para adquirir los conocimientos. Recuerde que debe avanzar con las siguientes cartillas y tener en cuenta las recomendaciones dadas a lo largo del avance del curso.

U2

Desarrollo temático

Componente motivacional

Este tema sirve para solucionar situaciones reales o ideales de manera eficiente, aporta herramientas teóricas para identificar apropiadamente los cálculos de distribución de probabilidad normal con el fin de entender los aspectos básicos de la estadística inferencial.

Por ejemplo, identificar las diferencias entre la probabilidad normal y estándar con el fin de establecer los conceptos prácticos de la solución de ejercicios.

Recomendaciones académicas

La presente cartilla contiene la fundamentación teórica de los aspectos teóricos y prácticos sobre la probabilidad normal y estándar que permite evidenciar la solución de situaciones problémicas que conduzcan a la comprensión de ejercicios sobre los aspectos esenciales de la estadística inferencial.

La matemática financiera es más sencilla de lo que nos parece. Es la utilización y aplicación de normas o leyes denominadas propiedades. Cuando aprendo apropiadamente estas propiedades, se puede afrontar de manera correcta cada una de las situaciones o ejercicios que se presentan, por esto la matemática no es difícil lo difícil es comprender y aplicar las propiedades que orientan el desarrollo de la misma asignatura. Todos, de una manera u otra podemos llegar a manejar, cada día mejor, la matemática.

Desarrollo de cada una de las unidades temáticas

Competencia general	Competencia específica
El estudiante constituye el núcleo de la estadística inferencial, y su propósito es aportar técnicas, métodos y herramientas para el calculo de operaciones inferencia estadística	El estudiante diferencia las características de los cálculos distribución binomial y normal estándar.

El estudiante relaciona la características que contribuyen a la planeación, organización, dirección, coordinación, control y evaluación de cualquier proyecto de inversión.	El estudiante calcula distribuciones de probabilidad sencillas teniendo en cuenta los procesos de estadística inferencial.
El estudiante entiende que la estadística le permite hacer predicciones reales de cantidades expresadas con probabilidad.	El estudiante entiende las características de los tipos de probabilidad aplicándolas en la solución de ejercicios en contexto.
El estudiante constituyen el núcleo de la estadística inferencia, y su propósito es aportar técnicas, métodos y herramientas para el calculo de operaciones de inferencia estadística.	El estudiante diferencia las características de los cálculos distribución binomial y normal estándar.
El estudiante relaciona las características que contribuyen a la planeación, organización, dirección, coordinación, control y evaluación de cualquier proyecto de inversión.	El estudiante calcula distribuciones de probabilidad sencillas teniendo en cuenta los proceso de estadística inferencia.
El estudiante entiende que la estadística le permite hacer predicciones reales de cantidades expresadas con probabilidad.	El estudiante entiende las características de los tipos de probabilidad aplicándolas de la solución de ejercicios en contexto.

Distribución normal

La distribución normal fue reconocida por primera vez por el francés Abraham de Moivre (1667-1754).

Posteriormente, Carl Friedrich Gauss (1777-1855) elaboró desarrollos más profundos y formuló la ecuación de la curva; de ahí que también se la conozca, más comúnmente, como la "campana de Gauss". La distribución de una variable normal está completamente determinada por dos parámetros, su media y su desviación estándar, denotadas generalmente por μ y σ . Con esta notación, la densidad de

la normal viene dada por la ecuación:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\}; -\infty < x < \infty$$

determina la curva en forma de campana. Así, se dice que una característica X sigue una distribución normal de media μ y varianza σ2, y se denota como X~ N (μ, σ) , si su función de densidad viene dada por la ecuación anteriormente mencionada.

Al igual que ocurría con un histograma, en el que el área de cada rectángulo es proporcional al número de

datos en el rango de valores correspondiente si, en el eje horizontal se levantan perpendiculares en dos puntos a y b, el área bajo la curva delimitada por esas líneas indica la probabilidad de que la variable de interés, *X,* tome un valor cualquiera en ese intervalo. Puesto que la curva alcanza su mayor altura en torno a la media, mientras que sus "ramas" se extienden asintóticamente hacia los ejes, cuando una variable siga una distribución normal, será mucho más probable observar un dato cercano al valor medio que uno que se encuentre muy alejado de éste. (Díaz, S. (2001). La distribución normal. Recuperado de https://www.fisterra.com/mbe/investiga/distr_normal/distr_normal.asp)

Propiedades de la distribución normal

La distribución normal posee ciertas propiedades importantes que conviene destacar:

I. Tiene una única moda, que coincide con su media y su mediana.

II. La curva normal es asintótica al eje de abscisas. Por ello, cualquier valor entre $-\infty$ y $+\infty$ es teóricamente posible. El área total bajo la curva es, por tanto, igual a 1.

III. Es simétrica con respecto a su media μ. Según esto, para este tipo de variables existe una probabilidad de un 50% de observar un dato mayor que la media, y un 50% de observar un dato menor.

IV. La distancia entre la línea trazada en la media y el punto de inflexión de la curva es igual a una desviación típica (σ). Cuanto mayor sea σ, más aplanada será la curva de la densidad.

V. El área bajo la curva comprendida entre los valores situados aproximadamente a dos desviaciones estándar de la media es igual a 0.95. En concreto, existe un 95% de posibilidades de observar un valor comprendido en el intervalo (μ -1.96σ; μ + 1.96σ).

VI. La forma de la campana de Gauss depende de los parámetros μ y σ. La media indica la posición de la campana. Por otra parte, la desviación estándar determina el grado de apuntamiento de la curva. Cuanto mayor sea el valor de, más se dispersarán los datos en torno a la media y la curva será más plana. Un valor pequeño de este parámetro indica, por tanto, una gran probabilidad de obtener datos cercanos al valor medio de la distribución. Como se deduce de este último apartado, no existe una única distribución normal, sino una familia de distribuciones con una forma común, diferenciadas por los valores de su media y su varianza. De entre todas ellas, la más utilizada es la distribución normal estándar, que corresponde a una distribución de media 0 y varianza 1. Así, la expresión que define su densidad se puede expresar:

$$f(z)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-z^2}{2}\right\};-\infty<z<\infty$$

Es importante conocer que, a partir de cualquier variable X que siga una distribución N (μ, σ) , se puede obtener otra característica Z con una distribución normal estándar, sin más que efectuar la transformación:

$$Z=\frac{x-\mu}{\sigma}$$

Esta propiedad resulta especialmente interesante en la práctica, ya que para una distribución existen tablas publicadas a partir de las que se puede obtener de modo sencillo la probabilidad de observar un dato menor o igual a un cierto valor z, y que permitirán resolver preguntas de probabilidad acerca del comportamiento de variables de las que se sabe o se asume que siguen una distribución aproximadamente normal. (Díaz, S. (2001). La distribución normal. Recuperado de https://www.fisterra.com/mbe/investiga/distr_normal/distr_normal.asp)

Consideremos, por ejemplo, el siguiente problema: supongamos que se sabe que el peso de los sujetos de una determinada población sigue una distribución aproximadamente normal, con una media de 80 Kg y una desviación estándar de 10 Kg. ¿Podremos saber cuál es la probabilidad de que una persona, elegida al azar, tenga un peso superior a 100 Kg?

Denotando por *X* a la variable que representa el peso de los individuos en esa población, ésta sigue una distribución N (80, 10). Si su distribución fuese la de una normal estándar podríamos calcular la probabilidad que nos interesa. Como éste no es el caso, resultará entonces útil transformar esta característica según (2), y obtener la variable:

Para poder utilizar dicha tabla. Así, la probabilidad que se desea calcular será:

$$P(X > 100) = P\left(Z > \frac{100-80}{10}\right) = P(Z > 2)$$

Como el área total bajo la curva es igual a 1, se puede deducir que:

$$P(Z > 2) = 1 - P(Z \leq 2)$$

Finalmente, la probabilidad buscada que una persona elegida al azar tenga un peso entre 60 y 100 Kg., es de 0.9772-0.0228=0.9544, es decir, aproximadamente de un 95%. Resulta interesante comprobar que se obtendría la misma conclusión recurriendo a la propiedad (III) de la distribución normal.

No obstante, es fácil observar que este tipo de situaciones no corresponde a lo que habitualmente nos encontramos en la práctica. Generalmente no se dispone de información acerca de la distribución teórica de la población, sino que más bien el problema se plantea a la inversa: a partir de una muestra extraída al azar de la población que se desea estudiar, se realizan una serie de mediciones y se desea extrapolar los resultados obtenidos a la población de origen. En un ejemplo similar al anterior, supongamos que se dispone del peso de n=100 individuos de esa misma población, obteniéndose una media muestral de X = 75 Kg, y una desviación estándar muestral S = 12Kg, se quiere extraer alguna conclusión acerca del valor medio real de ese peso en la población original. La solución a este tipo de cuestiones se basa en un resultado elemental de la teoría estadística, el llamado teorema central del límite. Dicho axioma viene a decirnos que las medias de muestras aleatorias de cualquier variable siguen ellas mismas una distribución normal con igual media que la de la población y desviación estándar la de la población dividida por √n. En nuestro caso, podremos entonces considerar la media muestral.

$$\overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Con lo cual, a partir de la propiedad (III) se conoce que aproximadamente un 95% de los posibles valores de X caerían dentro del intervalo.

$$\left(\mu-\frac{1.96\sigma}{\sqrt{n}};\mu+\frac{1.96\sigma}{\sqrt{n}}\right)$$

Puesto que los valores de μ y σ son desconocidos, podríamos pensar en aproximarlos por sus análogos muestrales, resultando (78 – 1.96 x 12/√100; 78 + 1.96 x 12/√100) = (75.6; 80.3). Estaremos, por lo tanto, un 95% seguros de que el peso medio real en la población de origen oscila entre 75.6 Kg y 80.3 Kg. Aunque la teoría estadística subyacente es mucho más compleja, en líneas generales éste es el modo de construir un intervalo de confianza para la media de una población. (Díaz, S. (2001). La distribución normal. Recuperado de https://www.fisterra.com/mbe/investiga/distr_normal/distr_normal.asp).

Contrastes de normalidad

La verificación de la hipótesis de normalidad resulta esencial para poder aplicar muchos de los procedimientos estadísticos que habitualmente se manejan. Tal y como ya se apuntaba antes, la simple exploración visual de los datos observados mediante, por ejemplo, un histograma o un diagrama de cajas, podrá ayudarnos a decidir si es razonable o no el considerar que proceden de una característica de distribución normal.

Resulta obvio que este tipo de estudio no puede llevarnos sino a obtener una opinión meramente subjetiva acerca de la posible distribución de nuestros datos, y que es necesario disponer de otros métodos más rigurosos para contrastar este tipo de hipótesis. En primer lugar, deberemos plantearnos el saber si los datos se distribuyen de una forma simétrica con respecto a su media o presentan algún grado de asimetría, pues es ésta una de las características fundamentales de la distribución de Gauss. Aunque la simetría de la distribución pueda valorarse, de modo simple, atendiendo a algunas medidas descriptivas de la variable en cuestión (comparando, por ejemplo, los valores de media, mediana y moda), resultará útil disponer de algún índice que nos permita cuantificar cualquier desviación. Si se dispone de una muestra de tamaño n, de una característica X, se define el coeficiente de asimetría de Fisher como.

$$\gamma_1=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{3/2}}$$

A partir del cual podemos considerar que una distribución es simétrica ($\gamma=0$), asimétrica hacia la izquierda ($\gamma<0$) o hacia la derecha ($\gamma>0$). En segundo lugar, podemos preguntarnos si la curva es más o menos "aplastada", en relación con el grado de apuntamiento de una distribución gaussiana. El coeficiente de aplastamiento o curtosis de Fisher, permite clasificar una distribución de frecuencias en mesocúrtica (tan aplanada como una normal, $\gamma_2 = 0$), leptocúrtica (más apuntada que una normal, $\gamma_2 > 0$) o platicúrtica (más aplanada que una normal, $\gamma_2 < 0$).

En cuanto a los niveles de curtosis, no hay apenas diferencias, siendo de –0.320 para el peso y de –0.366 para la edad. Los gráficos de probabilidad normal constituyen otra importante herramienta gráfica para comprobar si un conjunto de datos puede considerarse o no procedente de una distribución normal. La idea básica consiste en enfrentar, en un mismo gráfico, los datos que han sido observados frente a los datos teóricos que se obtendrían de una distribución gaussiana. Si la distribución de la variable coincide con la normal, los puntos se concentrarán en torno a una línea recta, aunque conviene tener en cuenta que siempre tenderá a observarse mayor variabilidad en los extremos. En los gráficos P-P se confrontan las proporciones acumuladas de una variable con las de una distribución normal. Los gráficos Q-Q se obtienen de modo análogo, esta vez representando los cuartiles respecto a los cuartiles de la distribución normal. Además de permitir valorar la desviación de la normalidad, los gráficos de probabilidad permiten conocer la causa de esa desviación. Una curva en forma de "U" o con alguna curvatura, significa que la distribución es asimétrica con respecto a la gaussiana, mientras que un gráfico en forma de "S" significará que la distribución tiene colas mayores o menores que la normal, esto es, que existen pocas o demasiadas observaciones en las colas de la distribución. Parece lógico que cada uno de estos métodos se complemente con procedimientos de análisis que cuantifiquen de un modo más exacto las desviaciones de la distribución normal. Existen distintos tests estadísticos que podemos utilizar para este propósito. El test de Kolmogorov-Smirnov es el más extendido en la práctica. Se basa en la idea de comparar la función de distribución acumulada de los datos observados con la de una distribución normal, midiendo la máxima distancia entre ambas curvas. Como en cualquier test de hipótesis, la hipótesis nula se rechaza cuando el valor del estadístico supera un cierto valor crítico que se obtiene de una tabla de probabilidad. Dado que en la mayoría de los paquetes estadísticos, como el SPSS, aparece programado dicho procedimiento, y proporciona tanto el valor del test como el p-valor correspondiente, no nos detendremos más en explicar su cálculo. Existen modificaciones de este test, como el de Anderson-Darling que también pueden ser utilizados. Otro procedimiento muy extendido es también el test chi-cuadrado de bondad de ajuste. No obstante, este tipo de procedimientos deben ser utilizados con precaución. Cuando se dispone de un número suficiente de datos, cualquier test será capaz de detectar diferencias pequeñas aún cuando estas no sean relevantes para la mayor parte de los propósitos. El test de Kolmogorov-Smirnov, en este sentido, otorga un peso menor a las observaciones extremas y por la tanto es menos sensible a las desviaciones que normalmente se producen en estos tramos. (Díaz, S. (2001). La distribución normal. Recuperado de https://www.fisterra.com/mbe/investiga/distr_normal/distr_normal.asp).

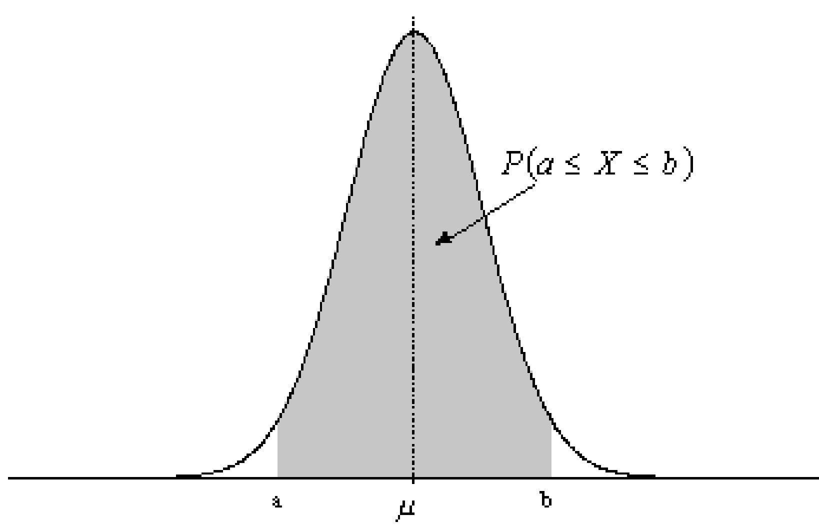


Imagen 1.Gráfica de la distribución normal.
Recuperado de https://www.fisterra.com/mbe/investiga/distr_normal/distr_normal.asp

La Distribución Normal es la piedra angular de la teoría estadística moderna. Conocida y estudiada desde hace mucho tiempo, es utilizada para describir el comportamiento aleatorio de muchos procesos que ocurren en la naturaleza y también realizados por los humanos.

Definición: Función de densidad de la distribución normal

Sea **X:** Variable aleatoria continua con media μ y varianza σ^2
X tiene distribución normal si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < +\infty$$

Se puede demostrar que **f** cumple las propiedades de una función de densidad:

1) $f(x) \geq 0, \quad -\infty<x<+\infty$;

2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

La gráfica de f es similar al perfil del corte vertical de una campana y tiene las siguientes características:

1) Es simétrica alrededor de μ
2) Su asíntota es el eje horizontal
3) Sus puntos de inflexión están ubicados en $\mu - \sigma$ y $\mu + \sigma$

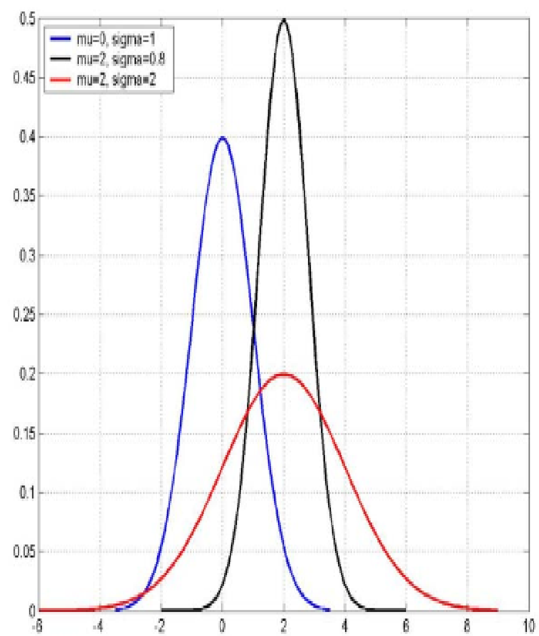


Gráfico de la distribución normal para varios valores de μ y σ

Para calcular probabilidad se tiene la definición

$$P(a \leq X \leq b) = \int_a^b f(x)dx, \qquad \text{siendo } \mathbf{a,b} \in \Re$$

También se puede usar la definición de distribución acumulada o función de distribución:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad \text{para } -\infty < x < +\infty$$

Esta definición es útil para calcular probabilidad con la propiedad: **$P(a \leq X \leq b) = F(b) - F(a)$**

Distribución normal estándar

El modelo estándar al cual pueden ser trasladados todos los casos particulares de distribuciones normales, modelo cuyos parámetros son $\mu = 0$ y $\sigma = 1$ (que son el origen y la unidad o escala de medida estándar); para el cual el símbolo X de la variable se cambia por el símbolo Z.

$Z \sim N(0, 1)$

La distribución normal estándar es el método general para calcular las probabilidades de eventos relativos a cualquier a cualquier distribución normal $X \sim N(\mu, \sigma)$ o, a la inversa, para encontrar eventos relativos a $X \sim N(\mu, \sigma)$ que satisfacen probabilidades conocidas de antemano.

La estandarización de una distribución normal significa que cualquier evento de una $X \sim N(\mu, \sigma)$ se traslada a un evento de la normal estándar $Z \sim N(0, 1)$ que mantenga la misma probabilidad (las probabilidades de ambos eventos deben corresponder a áreas iguales).

La distribución normal estándar, o tipificada o reducida, es aquella que tiene por media el valor cero, $\mu = 0$, y por desviación típica la unidad, $\sigma = 1$.

Su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Su gráfica es:

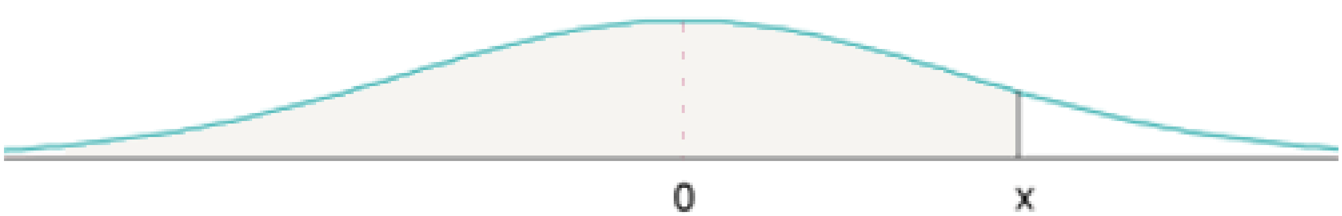


Imagen 2. Distribución normal estándar.
Recuperado de http://www.vitutor.com/pro/5/a_2.html

Para generalizar y facilitar el cálculo de probabilidad con la distribución normal, es conveniente definir la **Distribución Normal Estándar** que se obtiene haciendo $\mu = 0$, y $\sigma^2 = 1$ en la función de densidad de la distribución normal

Definición: Función de densidad de la distribución normal estándar

Sea **Z:** Variable aleatoria continua con media $\mu = 0$ y varianza $\sigma^2 = 1$
Z tiene distribución normal estándar si su función de densidad es:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < +\infty$$

Para calcular probabilidad con la distribución normal estándar se puede usar la definición de la **distribución acumulada** o **función de distribución:**

$$F(z) = P(Z \leq z) = \int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad -\infty < z < +\infty$$

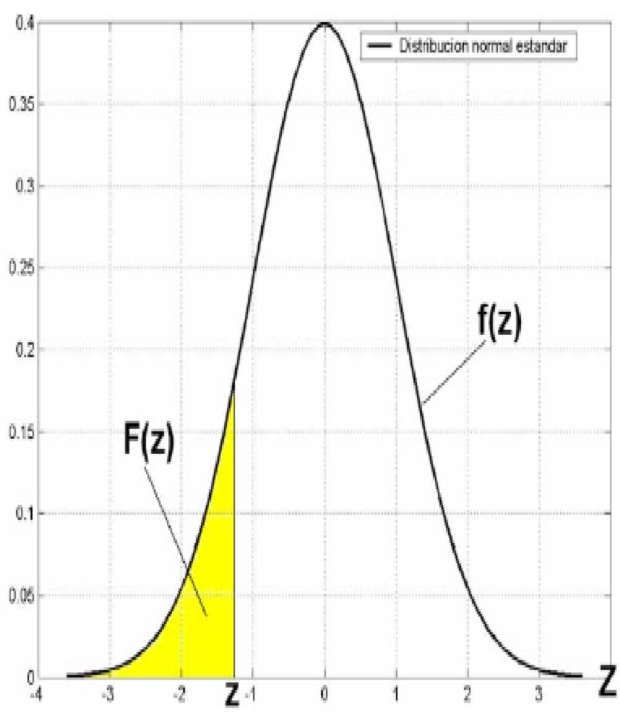


Gráfico de la distribución normal estándar

Esta distribución es frecuentemente utilizada en las aplicaciones estadísticas. Su propio nombre indica su extendida utilización, justificada por la frecuencia o normalidad con la que ciertos fenómenos tienden a parecerse en su comportamiento a esta distribución.

Muchas variables aleatorias continuas presentan una función de densidad cuya gráfica tiene forma de campana.

En otras ocasiones, al considerar distribuciones binomiales, tipo B(n,p), para un mismo valor de p y valores de n cada vez mayores, se ve que sus polígonos de frecuencias se aproximan a una curva en "forma de campana".

En resumen, la importancia de la distribución normal se debe principalmente a que hay muchas variables asociadas a fenómenos naturales que siguen el modelo de la normal

- *Caracteres morfológicos* de individuos (personas, animales, plantas,...) de una especie, p.ejm. tallas, pesos, envergaduras, diámetros, perímetros,...
- *Caracteres fisiológicos*, por ejemplo: efecto de una misma dosis de un fármaco, o de una misma cantidad de abono.
- *Caracteres sociológicos*, por ejemplo: consumo de cierto producto por un mismo grupo de individuos, puntuaciones de examen.
- *Caracteres psicológicos*, por ejemplo: cociente intelectual, grado de adaptación a un medio,...
- *Errores* cometidos al medir ciertas magnitudes.
- *Valores estadísticos* muestrales, por ejemplo : la media.
- *Otras distribuciones* como la binomial o la de Poisson son aproximaciones normales, ...

FUNCIÓN DE DENSIDAD

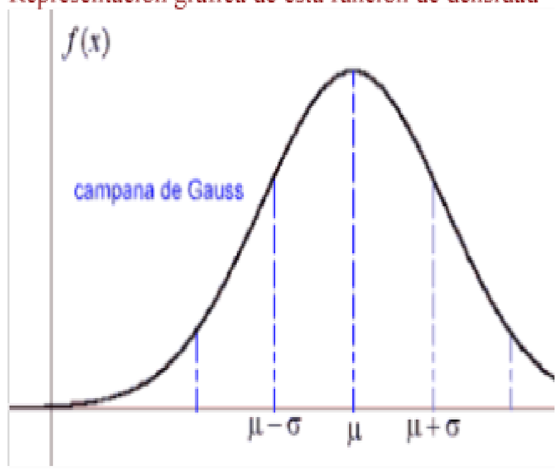
Empleando cálculos bastante laboriosos, puede demostrarse que el modelo de la **función de densidad** que corresponde a tales distribuciones viene dado por la fórmula

Función de Densidad

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ media	π = 3,1415...
σ desv. típica	e = 2,7182...
σ^2 varianza	x abscisa

Representación gráfica de esta función de densidad



FUNCIÓN DE DISTRIBUCIÓN

- Puede tomar cualquier valor (- ∞, + ∞)
- Son más probables los valores cercanos a uno central que llamamos media μ
- Conforme nos separamos de ese valor μ , la probabilidad va decreciendo de igual forma a derecha e izquierda (es simétrica).
- Conforme nos separamos de ese valor μ , la probabilidad va decreciendo de forma más o menos rápida dependiendo de un parámetro σ , que es la desviación típica.

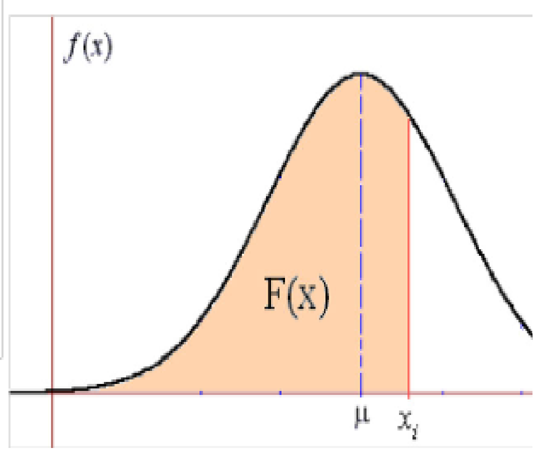
Función de Distribución

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$-\infty < x < +\infty$$

$$F(x) = P(X \leq x)$$

F(x) es el área sombreada de esta gráfica



[1]

Si la variable X es $N(\mu,\sigma)$ entonces la variable tipificada de X es $Z = \dfrac{X - \mu}{\sigma}$ y sigue también una distribución normal pero de $\mu = 0$ y $\sigma = 1$, es decir, $N(0,1)$

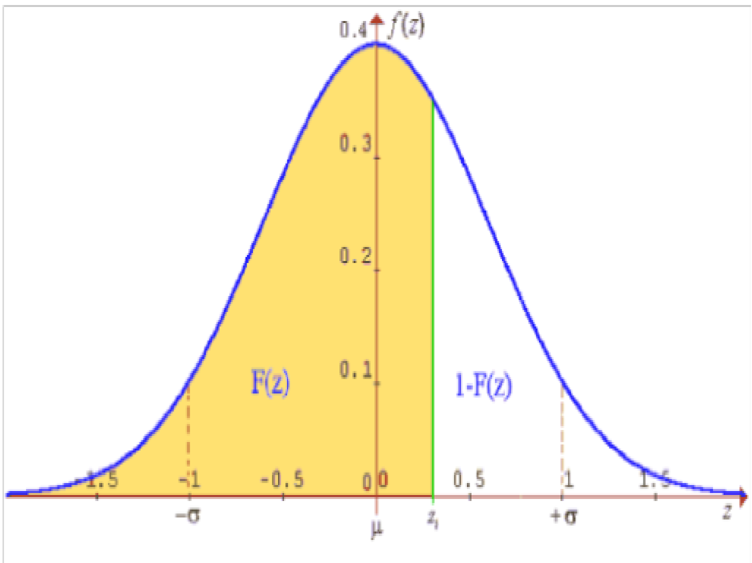
Por tanto su función de densidad es

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \qquad ; -\infty < z < +\infty$$

y su función de distribución es

$$F(z) = P(Z \leq z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

1Tomado de: http://electronica.ugr.es/~amroldan/asignaturas/sacpc/apuntes/va_normal.pdf siendo la representación gráfica de esta función



a la variable Z se la denomina *variable tipificada de X*, y a la curva de su función de densidad *curva normal tipificada*.

Característica de la distribución normal tipificada (reducida, estándar)

- No depende de ningún parámetro
- Su media es 0, su varianza es 1 y su desviación típica es 1.
- La curva *f(x)* es simétrica respecto del eje OY
- Tiene un máximo en este eje
- Tiene dos puntos de inflexión en $z = 1$ y $z = -1$

Ejemplos, ejercicios o casos de aplicación práctica.

1. En una ciudad se estima que la temperatura máxima en el mes de junio sigue una distribución normal, con media 23° y desviación típica 5°. Calcular el número de días del mes en los que se espera alcanzar máximas entre 21° y 27°.
2. Calcular la proporción de estudiantes que tienen puntuaciones que exceden por lo menos en cinco puntos de la puntuación que marca la frontera entre el Apto y el No-Apto (son declarados No-Aptos el 25% de los estudiantes que obtuvieron las puntuaciones más bajas.
3. Tras un test de cultura general se observa que las puntuaciones obtenidas siguen una distribución una distribución N(65, 18). Se desea clasificar a los examinados en tres grupos (de baja cultura general, de cultura general aceptable, de excelente cultura general) de modo que hay en el primero un 20% la población, un 65% el segundo y un 15% en el tercero. ¿Cuáles han de ser las puntuaciones que marcan el paso de un grupo al otro?
4. En una ciudad una de cada tres familias posee teléfono. Si se eligen al azar 90 familias, calcular la probabilidad de que entre ellas haya por lo menos 30 tengan teléfono.
5. En un examen tipo test de 200 preguntas de elección múltiple, cada pregunta tiene una respuesta correcta y una incorrecta. Se aprueba si se contesta a más de 110 respuestas correctas. Suponiendo que se contesta al azar, calcular la probabilidad de aprobar el examen.
6. Un estudio ha mostrado que, en un cierto barrio, el 60% de los hogares tienen al menos dos televisores Se elige al azar una muestra de 50 hogares en el citado barrio. Se pide¿Cuál es la probabilidad de que al menos 20 de los citados hogares tengan cuando menos dos televisores? ¿Cuál es la probabilidad de que entre 35 y 40 hogares tengan cuando menos dos televisores?
7. Un investigador científico reporta que unos ratones vivirán un promedio de 40 meses cuando sus dietas se restringen drásticamente y después se enriquecen con vitaminas y proteínas. Suponga que las vidas de tales ratones se distribuyen normalmente con una desviación estándar de 6.3 meses, encuentre la probabilidad de que un ratón dado viva.
8. Un abogado va todos los días de su casa en los suburbios a su oficina en el centro de la ciudad. El tiempo promedio para un viaje de ida es 24 minutos, con una desviación estándar de 3.8 minutos. Suponga que la distribución de los tiempos de viaje está distribuida normalmente.
9. La vida promedio de cierto tipo de motor pequeño es 10 años con una desviación estándar de dos años. El fabricante reemplaza gratis todos los motores que fallen dentro del tiempo de garantía. Si está dispuesto a reemplazar sólo 3% de los motores que fallan, ¿de qué duración debe ser la garantía que ofrezca? Suponga que la duración de un motor sigue una distribución normal.
10. Una compañía paga a sus empleados un salario promedio de \$15.90 por hora con una desviación estándar de \$1.50. Si los salarios se distribuyen aproximadamente de forma normal y se pagan al centavo más próximo.

Síntesis de cierre del tema

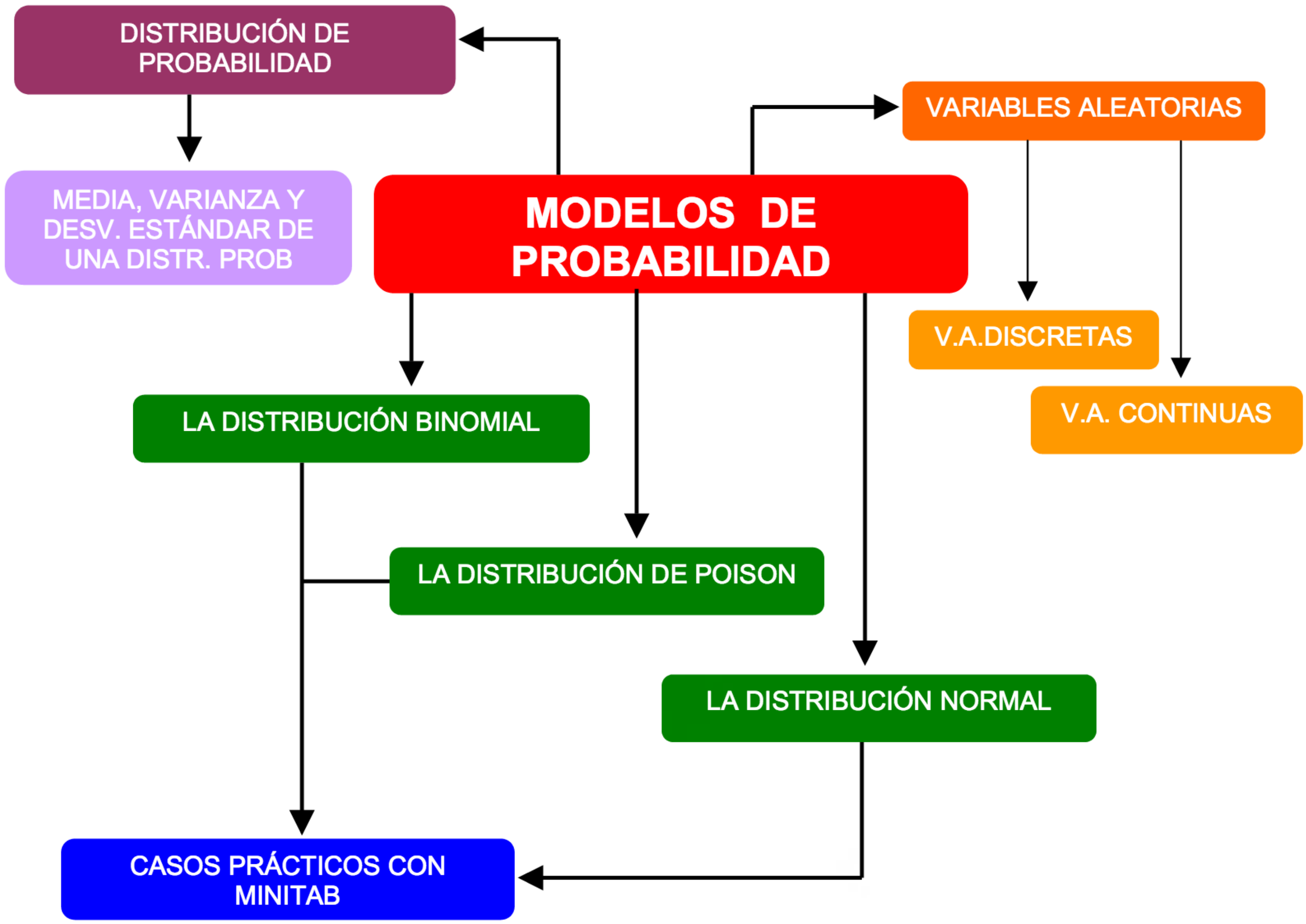


Imagen 3. Modelos de probabilidad.
Recuperado de http://www.uoc.edu/in3/emath/docs/Modelos_Probabilidad.pdf

Actividades auto-evaluativas propuestas al estudiante

1. ¿Cuál es el tamaño de la muestra que hay que extraer de una población normal para estimar la nota media de una asignatura con un error máximo de 0.5 puntos y con un riesgo del 5 %, sabiendo por estudios anteriores que la varianza poblacional es 8.52?
2. Una máquina debe introducir 375 gramos de cereales en cajas de envasado. La cantidad introducida es una variable aleatoria que se distribuye normalmente con media 375 gramos y desviación típica 20 gramos. Para comprobar que el peso medio de cada caja se mantiene en 375 gramos, se toman periódicamente muestras aleatorias de 25 cajas y se pesan sus contenidos. El encargado tiene orden de parar el proceso y ajustar la máquina cada vez que el promedio obtenido sea menor que 365 o mayor que 385 gramos. ¿Cuál es la probabilidad de tener que detener el proceso cada vez que se toma una muestra?
3. Se ha tomado una muestra de los precios de un mismo producto alimenticio en 16 comercios elegidos al azar en un barrio de una ciudad, y se han encontrado los siguientes precios: 95, 108, 97, 112, 99, 106, 105, 100, 99, 98, 104, 110, 107, 111, 103, 110. Suponiendo que los precios de este producto se distribuyen según una ley normal de varianza 25 y media desconocida: ¿Cuál es la distribución de la media muestral?
4. La variable altura de las alumnas que estudian en una escuela de idiomas sigue una distribución normal de media 1’62 m y desviación típica 0’12 m. ¿Cuál es la probabilidad de que la media de una muestra aleatoria de 100 alumnas sea mayor que 1’60 m?
5. Se supone que la estatura de los chicos de 18 años de cierta población sigue una distribución normal de media 162 cm y desviación típica 12cm. Se toma una muestra al azar de 100 de estos chicos encuestados y se calcula la media. ¿Cuál es la probabilidad de que esta media esté entre 159 y 165 cm?
6. Si un estudiante responde al azar a un examen de 8 preguntas de verdadero o falso ¿Cuál es la probabilidad de que acierte 4? ¿Cuál es la probabilidad de que acierte dos o menos? ¿Cuál es la probabilidad de que acierte cinco o más? ¿Cuánto valen la media y la varianza del número de preguntas acertadas?
7. En una población en la que hay un 40% de hombres y un 60% de mujeres seleccionamos 4 individuos ¿Cuál es la probabilidad de que haya 2 hombres y 2 mujeres? ¿Cuál es la probabilidad de que haya más mujeres que hombres?
8. Las calificaciones en un examen siguen una distribución Normal de media 5,6 y desviación típica 0,8. ¿Qué proporción de alumnos tendrá puntuaciones inferiores o iguales a 4? ¿Qué proporción de alumnos aprobará? ¿Qué proporción de alumnos obtendrá Notable o Sobresaliente?
9. Las puntuaciones en un test de ansiedad-rasgo siguen, en una población de mujeres, una distribución Normal de media 25 y desviación Típica 10. Si queremos clasificar la población en cuatro grupos de igual tamaño ¿Cuáles serán las puntuaciones que delimiten estos grupos?

10. Una prueba consta de 200 preguntas de verdadero o falso, para un sujeto que respondiese al azar ¿Cuál sería la probabilidad de que acertase? 50 preguntas o menos. Más de 50 y menos de 100. Más de 120 preguntas.

2

Unidad 2

Estimación de parámetros

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

Fundación Universitaria del Área Andina

MIEMBRO DE LA RED ILUMNO

Introducción

En esta cartilla se definen los referentes de la estimación de parámetros para el cálculo de intervalos de confianza, es importante leer cuidadosamente el contenido de la cartilla. Al interpretar cada una de las expresiones, le permite comprender los elementos expuestos para así aplicar los contenidos y asimilarlos.

Frecuentemente, para solucionar situaciones cotidianas, se requiere diferenciar los procedimientos de cálculo para resolver los diferentes contextos matemáticos.

U2

Metodología

El proceso de aprendizaje académico por medio de esta cartilla se desarrolla en tres etapas. La primera etapa es la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido por medio de la revisión de los ejercicios resueltos con el fin de ganar confianza en la solución de ejercicios propuestos. La segunda etapa es el análisis del video y el contenido del mismo. La tercera etapa es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

U2

Desarrollo temático

Componente motivacional

Este tema sirve para solucionar situaciones reales o ideales de manera eficiente, aporta herramientas teóricas para identificar apropiadamente los cálculos de la estimación de parámetros y los intervalos de confianza con el fin de entender los aspectos básicos de la estadística inferencial.

Por ejemplo, identificar las diferencias entre la relación de parámetros y los intervalos de confianza con el fin de establecer los conceptos prácticos de la solución de ejercicios.

Recomendaciones académicas

La presente cartilla contiene la fundamentación teórica de los aspectos teóricos y prácticos sobre la estimación de parámetros y los intervalos de confianza que permite evidenciar la solución de situaciones problémicas que conduzcan a la comprensión de ejercicios sobre los aspectos esenciales de la estadística inferencial.

La estadística inferencial es más sencilla de lo que nos parece. Es la utilización y aplicación de normas o leyes denominadas propiedades. Cuando se aprende apropiadamente estas propiedades, se puede afrontar de manera correcta cada una de las situaciones o ejercicios que se presentan, por esto la matemática no es difícil lo difícil es comprender y aplicar las propiedades que orientan el desarrollo de la misma asignatura.

Desarrollo de cada una de las unidades temáticas

Competencia general	Competencia específica
El estudiante aprende a identificar la estimación de parámetros	El estudiante diferencia las características de la estimación de parámetros
El estudiante relaciona las características de la estimación de parámetros y los intervalos de confianza	El estudiante aplica las características de la estimación de parámetros
El estudiante entiende las características de los intervalos de confianza	El estudiante entiende las características de la estimación de parámetros.

Estimación de parámetros

La estimación de parámetros hace referencia al estudio de una muestra de una población a partir de los diferentes parámetros que permite caracterizar los valores hallados en las poblaciones y las muestras.

Estimador puntual

Se define como un valor estadístico único que se usa para estimar un parámetro. El estadístico usado se denomina estimador.

Estimación por intervalo

Se define como un valor estadístico que proporciona el estadístico para estimar el parámetro dependiendo de los rangos de trabajo del valor estimado.

Un estimador es confiable cuando su probabilidad es un rango correcto dentro del valor del parámetro, el cual responde a la solución del teorema del límite central, para una distribución normal de muestras grandes.

Según Weimer (2007), las estimaciones se pueden calcular teniendo en cuenta:

Los valores poblacionales suelen estimarse por sus correspondientes valores muestrales, una media muestral $\bar{x}$ es una estimación de la media poblacional μ; una proporción muestral $\hat{p} = x/n$ es una estimación de la proporción poblacional p. La varianza poblacional σ2 se estima por s2, la varianza muestral; mientras la desviación estándar σ de una población se estima por la desviación estándar s de una muestra extraída de la población. (p. 415).

Estimado insesgado

Se define como un valor para el cual la media de la distribución muestral es el parámetro estimado. Si se usa la media muestral $\bar{x}$ para estimar la media poblacional μ, se afirma que es insesgado si μx = μ.

El valor absoluto de la diferencia entre una estimación y el parámetro estimado se conoce como error de estimación.

Ejemplo

Se realiza un estudio para determinar el promedio de gaseosa vendida por un dispensador automático, en un estudio anterior se calculó que la desviación estándar era s = 0.50 onzas. Si se usa una muestra de 50 para estimar μ, encuentre el nivel de confianza de que el error de estimación sea menor de 0.01 onzas.

$$\sigma_{g} = \frac{\sigma}{\sqrt{n}} = \frac{0.50}{\sqrt{50}} = 0.070$$

Entonces,

$$P(|\overline{X}-\mu|) < P\left(|Z| < \frac{0.01}{0.070}\right) = P(|Z| < 0.14) = 2P(0 < Z < 0.14) = 2(0.0736)$$
$$= 0.1472$$

El valor de confianza para esta estimación puntual es de 14.72 %, de que el error de estimación sea menor que 0.01.

Si se desea calcular el error máximo de estimación E para muestras grandes se utiliza la expresión matemática:

$$E = Z_{\alpha/2}\frac{\alpha}{\sqrt{n}}$$

Si se desea calcular el error máximo de estimación para muestras pequeñas se utiliza

$$E = t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Intervalos de confianza

Se define como el valor de un intervalo que se encuentra en un límite de valor designado que cumple la siguiente estructura equivalente entre si:

Límite para un intervalo de confianza

1. $\overline{x} \pm E$
2. $\overline{x} \pm z_{\alpha/2}{}^{\sigma}/_{\sqrt{n}}$
3. $\overline{x} \pm z_{\alpha/2}\sigma\overline{x}$

Intervalos de confianza para muestras pequeñas

En una muestra pequeña se tiene en cuenta la siguiente expresión matemática

$$\overline{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Ejemplo

Cinco bolsas de maíz pesan 14.5; 13.3; 16.9; 12; 8 y 16.5 libas respectivamente. Encuentre un intervalo del 95% de confianza para la media del peso de estas bolsas, suponiendo que la medida de los pesos es normal.

La media muestral y la desviación estándar ser $\overline{x}$ = 15.36 y s = 0.472, respectivamente. Para gl = 5 – 1 = 4, entonces t0.025 = 2.776.

$$E = t_{0.025}\frac{s}{\sqrt{n}} = (2.776)\left(\frac{0.472}{\sqrt{5}}\right) = 0.59$$

y los límites para el intervalo del 95% de confianza son:

$$\overline{x} \pm E$$

$$15.36 \pm 0.59$$

Intervalo de confianza para proporciones

Los límites del intervalo del (1 – α) 100% de confianza para aproximar la proporción poblacional p están dados por una de las tres expresiones equivalentes:

$$\widehat{p} \pm E$$

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

$$\widehat{p} \pm z_{\alpha/2}\sigma_{\widehat{p}}$$

Ejemplo

Un candidato político está planeando su campaña y quiere determinar su popularidad. En una muestra aleatoria de 5000 de los 30000 votantes registrados en el país, 1900 manifestaron conocer al candidato. Construya un intervalo del 95% de confianza para la proporción de votantes en país que conocen al candidato.

$Z_{0.025}$= 1.96. Como 30000 < (20)(5000), entonces p= 1900/5000 = 0.38

$$\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{(0.38)(0.62)}{5000}}\sqrt{\frac{30000-5000}{30000-1}} = 0.0051$$

Intervalo de confianza para las medias

Al calcular una media se debe tener en cuenta la siguiente expresión

$$n = \left(\frac{Z_{\alpha/2\sigma}}{E}\right)^2$$

Con esta expresión se busca calcular el tamaño de la muestra para que el intervalo de confianza sea lo más cercano a E.

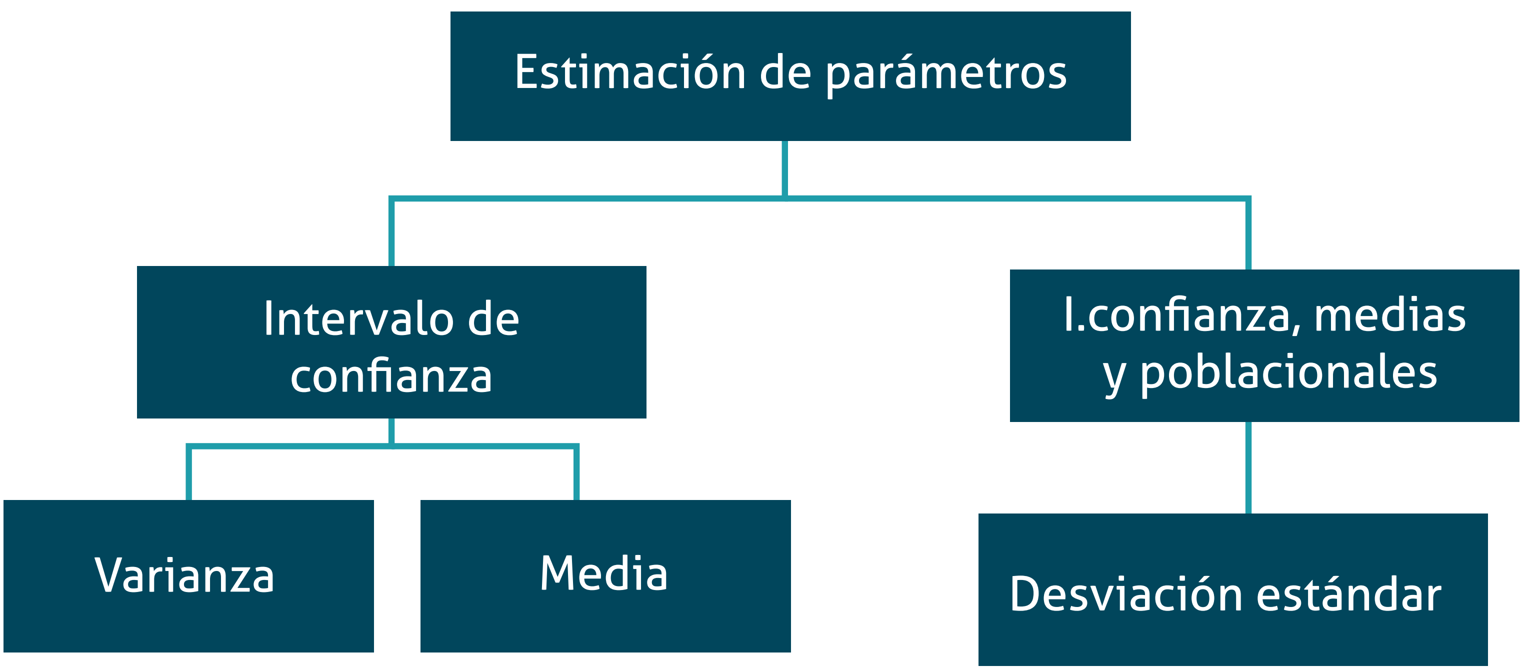
Ejemplo

Un científico desea estudiar el peso promedio de las águilas doradas. Un estudio anterior de 2 águilas mostró que la desviación estándar de sus pesos es de 14.6 libras. ¿Qué tan grande debe ser una muestra para que el científico tenga el 95 % de confianza de que el error de estimación es a lo más 5 libras?

$$n = \left(\frac{z_{\alpha/2\sigma}}{E}\right)^2 = \left(\frac{(1.96)(14.6)}{5}\right)^2 = 32.75$$

El tamaño de la muestra debe ser n = 33 águilas para que se pueda lograr un intervalo de confianza de 95%.

Síntesis de cierre del tema



Actividades auto-evaluativas propuestas al estudiante

1. En una muestra de 660 personas el nivel medio de glicemia es 15 mmol/l. y la desviación estándar es 0,5. ¿Cuál es el intervalo de referencia al 95% para los valores de glicemia en la población origen?
2. En una muestra sobre 108 pacientes se prueba la eficacia de un nuevo medicamento. Al final del tratamiento, 62 pacientes muestran una evolución positiva, mejorando apreciablemente. Estima la eficacia del medicamento con una confianza del 95%.
3. Una máquina produce piezas cuyas longitudes están distribuidas normalmente con s = 0,15 mm. La longitud media de una muestra de 1000 piezas es de 85,92 mm. Encuentre la estimación puntual y para un nivel de confianza del 95%.
4. En un hospital psiquiátrico se ha estudiado una muestra de 350 pacientes y se ha observado que 88 de ellos tienen una cierta tendencia al suicidio. Hallar un intervalo de confianza al nivel de 95% para el parámetro proporción de individuos con tendencia al suicidio de la población.
5. Se desea hacer un estudio de mercado sobre el nivel de aceptación de un tipo de detergente. Para ello, se ha tomado una muestra aleatoria formada por 160 personas, de las cuales 85 son compradoras frecuente del tipo de detergente. Hallar un intervalo de confianza al nivel de 99% para la población de usuarios del citado detergente en una comarca muy poblada.
6. Un psicólogo ha estudiado que el tiempo de juego en el descanso de los niños de 5° de primaria. Con una muestra de 200 alumnos, la media de tiempo de descanso fue de 45 segundos y la desviación típica de 0,04 segundos. Hallar un intervalo de confianza para la media de tiempos de reacción al nivel de confianza.
7. Se sospecha que el número de unidades que contiene cada dosis de un medicamento no llega a las 10.000 que se indican en el envase pero el laboratorio que lo fabrica afirma que el contenido medio de la dosis es de 10.000 unidades. Para comprobarlo, tomamos al azar 100 dosis y determinamos el número de unidades de cada una, obteniendo de media 9940 unidades y de desviación típica 120 unidades. Si suponemos que la distribución del número de unidades en la población es normal, ¿qué podemos decir acerca de la afirmación del laboratorio para un nivel de confianza del 99%?
8. Se afirma que la estatura media de las personas adultas de una determinada región es de 1,80m. Queremos tener una confianza del 99% en saber si la afirmación anterior es correcta o errónea. Para ello, tomamos una muestra al azar de 100 personas adultas, a las que medimos sus alturas, obteniendo de media 1,78 m. y de desviación típica 0,10 m. Suponemos que la variable objeto de estudio es normal.

Unidad 3

Distribuciones muestrales

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO

Introducción

En esta cartilla se aborda el conocimiento teórico y práctico sobre las bases de la estadística inferencial a partir de la conformación del muestreo y la distribución muestral.

Es importante leer cuidadosamente el contenido de la cartilla. Al interpretar cada una de las expresiones, le permite comprender los elementos expuestos para así aplicar los contenidos y asimilarlos.

Frecuentemente, para solucionar situaciones cotidianas, se requiere diferenciar los procedimientos de cálculo para resolver los diferentes contextos matemáticos.

U3

Metodología

El proceso de aprendizaje académico por medio de esta cartilla se desarrolla en tres etapas. La primera etapa es la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido por medio de la revisión de los ejercicios resueltos con el fin de ganar confianza en la solución de ejercicios propuestos. La segunda etapa es el análisis del video y el contenido del mismo. La tercera etapa es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

Cabe anotar que la segunda etapa puede ser retomada cuantas veces quiera, pero tiene un tiempo limitado para adquirir los conocimientos. Recuerde que debe avanzar con las siguientes cartillas y tener en cuenta las recomendaciones dadas a lo largo del avance del curso.

U3

Desarrollo temático

Componente motivacional

Este tema sirve para solucionar situaciones reales o ideales de manera eficiente, aporta herramientas teóricas para identificar apropiadamente los cálculos de interés simple y compuesto con el fin de entender los aspectos básicos de la estadística inferencial.

Por ejemplo, identificar las diferencias entre distribuciones muestrales con el fin de establecer los conceptos prácticos de la solución de ejercicios.

Recomendaciones académicas

La presente cartilla contiene la fundamentación teórica de los aspectos teóricos y prácticos sobre las distribuciones muestrales que permite evidenciar la solución de situaciones problémicas que conduzcan a la comprensión de ejercicios sobre los aspectos esenciales de la estadística inferencial.

La matemática financiera es más sencilla de lo que nos parece. Es la utilización y aplicación de normas o leyes denominadas propiedades. Cuando se aprende apropiadamente estas propiedades, se puede afrontar de manera correcta cada una de las situaciones o ejercicios que nos solicita, por esto la estadística no es difícil lo difícil es comprender y aplicar las propiedades que orientan el desarrollo de la misma asignatura. Todos, de una manera u otra podemos llegar a manejar, cada día mejor, la matemática.

Desarrollo de cada una de las unidades temáticas

Competencia general	Competencia específica
El estudiante constituyen el núcleo de la estadística inferencial, y su propósito es aportar técnicas, métodos y herramientas para el cálculo de operaciones de probabilidad.	El estudiante diferencia las características de las distribuciones muestrales.
El estudiante relaciona las características que contribuyen a la planeación, organización, dirección, coordinación, control y evaluación de cualquier proceso estadístico.	El estudiante calcula porcentajes que le permite reconocer las posibilidades teóricas y prácticas de la estadística inferencial.
El estudiante entiende que la estadística inferencial es de uso y aplicación diaria en transacciones comerciales, laborales, bancarias y bursátiles.	El estudiante entiende las características de las distribuciones muestrales y los diferentes errores.

Distribuciones muestrales

Para el estudio de un fenómeno, se requiere contar con información relacionada con el mismo. Esta información obtenida bien sea experimentalmente o, mediante la observación, está dada por datos. Estos datos son el resultado de medir en un conjunto de elementos o individuos, una o varias características a ser analizadas en una investigación. Ahora bien, el análisis puede llevarse a cabo en base a toda o, a una parte de la población. Si se hace uso de toda la información, decimos que se ha hecho una investigación exhaustiva o total (censo). No siempre es posible realizar un censo, por razones como; costos, tiempo, poco práctico, etc. Es necesario entonces, en estos casos, llevar a cabo una investigación parcial.

La misma consiste en realizar el análisis en base a la información correspondiente a un subconjunto de los elementos o individuos, una muestra, de forma tal que a un costo y esfuerzo razonable se logren obtener conclusiones tan validas como las que se obtendrían realizando una investigación exhaustiva o total, un censo.

Conceptos Básicos

Universo: es el conjunto de individuos o elementos (personas, fábricas, familias, etc.) que poseen características en común que se desean investigar.

Población: es el conjunto de todas las posibles mediciones que pueden hacerse de una o más características en estudio de los elementos del universo. Por lo tanto, la población está constituida por valores o datos bien sea numéricos o no.

Muestra: es una parte de una población, idealmente representativa de la misma.

Parámetro: es una función de los valores de la población que sirve para sintetizar alguna característica relevante de la misma. Es una medida resumen que se calcula para describir una característica de toda una población.

Ejemplos de parámetros son: la media poblacional, la proporción poblacional, la varianza poblacional, entre otros. (Distribuciones muestrales. Recuperado de http://webdelprofesor.ula.ve/economia/drivas/materias/metodosII/Distribuciones%20en%20el%20muestreo.pdf).

Estadístico: se denomina estadístico a toda función medible de los elementos de una muestra en la que no intervienen parámetros.

Muestreo: proceso de medición de la información en solo una parte de la población estadística. Se define como el proceso de seleccionar un número de observaciones.

Sujetos de un grupo en particular de la población (métodos para seleccionar muestras), que se utiliza cuando no es posible contar o medir todos los elementos de la población objeto de estudio.

Tipos de muestreo

Existen dos métodos para seleccionar muestras de poblaciones.

Muestreo no aleatorio o de juicio: es práctica común seleccionar una muestra en forma intencional, de acuerdo a opiniones o criterios personales, fundamentalmente con el objeto de obtener información sin mucho costo. A este tipo de muestreo se le denomina muestreo no probabilístico, no aleatorio o de juicio.

Este tipo de muestreo como puede observarse, no involucra ningún elemento aleatorio en el procedimiento de selección. Sin embargo, es importante resaltar que en condiciones apropiadas estos métodos pueden ofrecer resultados útiles, por ejemplo, cuando solo se necesitan estimaciones gruesas, las cuales no van a ser utilizadas para tomar decisiones importantes. Son ejemplos de muestreos no probabilísticos:

a. La muestra es restringida a la parte de la población que es fácilmente accesible.
b. La muestra consiste de los elementos que estén más a la mano.
c. Se selecciona un grupo de unidades tipo.
d. La muestra está compuesta por voluntarios.

Muestreo aleatorio o probabilístico: en el cual todos los elementos de la población tienen la oportunidad de ser escogidos para la muestra. Este procedimiento da a cada elemento de la población una probabilidad de ser seleccionado. Dentro de este tipo de muestreo se encuentran:

a. Muestreo aleatorio simple: el cual es un método de selección de muestras que permite que cada muestra posible pueda ser elegida con la misma probabilidad. Por su parte cada elemento de la población tiene la misma oportunidad de ser incluido en la muestra.
b. Muestreo sistemático: método en el cual los elementos que se seleccionan de la población en un intervalo uniforme que se mide con respecto al tiempo, al orden o al espacio.
c. Muestreo estratificado: método en el que la población se divide en grupos homogéneos, o estratos, y después se toma

una muestra aleatoria simple de cada estrato. Aquí la variabilidad dentro de cada grupo es pequeña y entre los grupos es grande.

d. Muestreo por conglomerados: método en el que la población se divide en grupos o racimos de elementos, y luego se selecciona una muestra aleatoria de estos racimos.

La variabilidad dentro de cada grupo es grande y entre los grupos es pequeña; es como si cada conglomerado fuese una pequeña representación de la población en sí misma.

Métodos para seleccionar una muestra aleatoria

Al seleccionar una muestra aleatoria se debe tomar en cuenta si la extracción se va realizar con reemplazo o sin reemplazo, en el primer caso, una vez extraída el elemento de la población este puede ser devuelto a la misma, en el segundo caso esto no es posible.

Existen varios métodos para seleccionar una muestra, entre los cuales se pueden mencionar: método del bingo, tabla de Números aleatorios y generación de números pseudoaletorios.

Método del bingo

Consiste en etiquetar N papeles, bolas o cualquier otro objeto del 1 al N e introducirlas en una urna o bolsa y agitarla hasta que queden bien mezcladas, luego extraer una a la vez hasta que se hayan seleccionado n artículos donde n es el tamaño deseado de la muestra. Los miembros de la población que correspondan a los números de los artículos extraídos son incluidos en la muestra, y las características de estas unidades se miden u observan. Si la población es bastante grande, este método mecánico de selección aleatoria puede ser difícil o prácticamente imposible de implementar. Esto nos lleva a la consideración de la tabla de números aleatorios.

Tabla de números aleatorios

Las tablas de números aleatorios contienen los dígitos 0; 1; 2; :::; 7; 8; 9. Tales dígitos se pueden leer individualmente o en grupos y en cualquier orden, en columnas hacia abajo, columnas hacia arriba, en fila, diagonalmente, etc., y es posible considerarlos como aleatorios. Las tablas se caracterizan por dos cosas que las hacen particularmente útiles para el muestreo al azar.

Una característica es que los dígitos están ordenados de tal manera que la probabilidad de que aparezca cualquiera en un punto dado de una secuencia es igual a la probabilidad de que ocurra cualquier otro. La otra es que las combinaciones de dígitos tienen la misma probabilidad de ocurrir que las otras combinaciones de un número igual de dígitos.

Estas dos condiciones satisfacen los requisitos necesarios para el muestreo aleatorio, establecidos anteriormente. La primera condición significa que en una secuencia de números, la probabilidad de que aparezca cualquier dígito en cualquier punto de la secuencia es 1/10. La segunda condición significa que todas las combinaciones de dos dígitos son igualmente probables, del mismo modo que todas las combinaciones de tres dígitos, y así sucesivamente. Para utilizar una tabla de números aleatorios:

a. Hacer una lista de los elementos de la población.

b. Numerar consecutivamente los elementos de la lista, empezando con el cero (0, 00, 000, etc.).
c. Tomar los números de una tabla de números aleatorios, de manera que la cantidad de dígitos de cada uno sea igual a la del último elemento numerado 1.3 muestreo 11 de su lista. De ese modo, si el último número fue 18, 56 ó 72, se debería tomar un número de dos dígitos.
d. Omitir cualquier dígito que no corresponda con los números de la lista o que repita cifras seleccionadas anteriormente de la tabla. Continuar hasta obtener el número de observaciones deseado.
e. Utilizar dichos números aleatorios para identificar los elementos de la lista que se habrían de incluir en la muestra.

Para ilustrar el uso de la tabla de números aleatorios se daría el siguiente ejemplo:

suponga que se tienen 40 latas de refrescos, y que se desea tomar una muestra de tamaño n = 4 para estudiar su condición. El primer paso es numerar las latas de 1 a 40 o apilarlas en algún orden de tal forma que puedan ser identificadas.

En la tabla de números aleatorios, los números deben escogerse de dos dígitos porque la población de tamaño N = 40 es un número de dos dígitos. Luego se selecciona arbitrariamente una fila y una columna de la tabla. Suponga que la selección es la fila 6 y la columna 4. Se leen los pares de dígitos a partir de la fila 6 y la columna 4 y moviéndonos hacia la derecha, ignorando los números mayores que 40 y también cualquier número repetido cuando aparezca una segunda vez. Se continúa leyendo pares de dígitos hasta que cuatro unidades diferentes hayan sido seleccionadas, es decir los números 05, 20, 08 y 17. Por lo tanto, las latas con la etiqueta correspondiente a dichos números constituyen la muestra.

Generación de números pseudoaletorios. Existen métodos más eficaces para generar números aleatorios, en muchos de los cuales se utilizan calculadoras o computadoras. La mayoría de los paquetes estadísticos generan números pseudoaleatorios y en Excel usando la función aleatoria se pueden generar dichos números.

Error de muestreo

Es el error que se comete debido al hecho de dar conclusiones sobre cierta realidad, a partir de la observación de sólo una parte de ella, es decir, es la diferencia entre el parámetro de la población y el estadístico de la muestra utilizado para estimar el parámetro.

Distribuciones muestrales

Se ha dicho que uno de los objetivos de la estadística es saber acerca del comportamiento de parámetros poblacionales tales como: la media, la varianza o la proporción. Para ello, se extrae una muestra aleatoria de la población y se calcula el valor de un estadístico correspondiente, por ejemplo, la media muestral, la varianza muestral o la proporción muestral.

Un estadístico es una variable aleatoria, informalmente esto es cierto, ya que su valor depende de los elementos elegidos en la muestra seleccionada. La veracidad formal de esta declaración se da en el siguiente teorema (sin demostración).

Sean X1;X2;:::;Xn n variables aleatorias. Definamos Y = f(X1;X2;::;Xn), entonces Y es también una variable aleatoria.

El teorema anterior establece que una función de una o más variables aleatorias es también una variable aleatoria,, y como un estadístico es una función de la muestra (las cuales son variables aleatorias), entonces un estadístico es una variable aleatoria, y en consecuencia tiene asociada una distribución de probabilidad la cual es llamada la distribución muestral del estadístico.

Empíricamente

Para hallar empíricamente la distribución muestral de un estadístico es necesario seleccionar todas las muestras de dicha población y a partir de dicha información construir la distribución de frecuencia relativa de los valores del estadístico, la cual es considerada como su distribución muestral. Veamos a continuación el cálculo de la distribución muestral de dos estadísticos muy importantes, la media muestral y la proporción.

Distribución muestral de la media

Para hallar la distribución muestral de la media se procede de la siguiente manera:

1. Se seleccionan desde la población todas las muestras posibles de tamaño n.
2. En cada muestra se calcula la media muestral.
3. A partir de dicha información se construye la distribución de frecuencias relativas de las medias.

DISTRIBUCIONES MUESTRALES

Como sabemos, cuando una medida tal como la media, la desviación estándar, la proporción, etc. describe cierto aspecto de una población, esta medida toma el nombre de ***parámetro*** de la población considerada. De igual manera, cuando una medida describe cierto aspecto de una muestra, se le da el nombre de ***estadístico*** de la muestra.

Un parámetro es una ***cantidad constante***. Una vez determinado su valor, este valor es permanente, ***mientras no cambie la población.***

De una muestra aleatoria de ***n*** valores, varios estadísticos podrían medirse. Usando estos ***n*** valores, se podría calcular una media, una mediana, una rango, una varianza, una desviación estándar, etc.
Ahora si en lugar de una muestra consideramos todas las muestras posibles del mismo tamaño, ***n,*** de una población y al calcular la media aritmética de cada muestra, algunas de las medias muestrales serían iguales; otras medias muestrales serían diferentes. Agrupando todas las medias muestrales que son semejantes y utilizando las reglas de probabilidad, podríamos determinar la probabilidad de obtener una muestra con una media particular.

Repitiendo este procedimiento para cada media posible de una muestra, se construye una distribución de probabilidad de la variable aleatoria "media muestral".

Una distribución de muestreo o muestral es una ***distribución de probabilidad*** *donde la variable aleatoria es una estadística.*

Es decir: un estadístico es una ***cantidad variable*** ya que un mismo tipo de estadístico obtenido en diferentes muestras de una misma población, puede variar su valor de muestra a muestra, por tanto podemos decir que una estadístico es una ***variable aleatoria*** y su valor depende de los elementos que conforman la muestra.

Dado que un estadístico es una variable aleatoria que depende únicamente de la muestra observada, debe tener una distribución de probabilidad.

Definición. *La distribución de probabilidad de un estadístico recibe el nombre de* ***distribución muestral.***

La distribución de probabilidad de X se llama "distribución muestral de la media".
La distribución de probabilidad de σ^2 se llama "distribución muestral de la varianza".

La distribución muestral de un estadístico depende del tamaño de la población, del tamaño de la muestra y del método de selección de las muestras.

Tenemos que cualquier distribución de probabilidad (y por lo tanto cualquier distribución muestral) puede ser descrita por su media, su desviación estándar y su forma.

Parámetros	Estadísticos
N: Tamaño de la población	n: tamaño de la muestra
μ: media de la población	$\bar{x}$: media de la muestra
σ: desviación estándar de la población	s: desviación estándar de la muestra
$\mu_{\bar{x}}$: media de la distribución muestral de la media (media de todas las muestras posibles de tamaño ***n***)	$\sigma_{\bar{x}}$: desviación estándar de la distribución muestral de la media (error estándar de la media)

¿Cómo podemos obtener la descripción de una distribución muestral de un estadístico?

Para explicar este concepto vamos a obtener la descripción de la distribución muestral de la media para una población con pocos elementos.

DESCRIPCIÓN DE UNA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Para describir la distribución muestral de la media, es necesario seguir los siguientes pasos:

1.- Obtener todas las muestras posibles de tamaño ***n*** de la población de tamaño N
2,. Para cada muestra se calcula la media.
3.- Se construye la distribución de probabilidad para los valores de las medias
4.- Se calcula la media y la desviación estándar de la distribución de probabilidad.
5.- Se construye la gráfica de la distribución de probabilidad

Ejemplo: Una población está formada por los valores:

12, 6, 7,10, 8, 11, 9, 13

Describir la distribución muestral de la media, al tomar todas las muestras de tamaño 3 ***sin reemplazo***,

1 y 2.- Para determinar todas las muestras necesitamos primero saber el número de muestras posible: Si el muestreo es sin reemplazo el numero de muestras está dado por:

$$ {}_N C_n = \frac{N!}{n!(N-n)!} $$

Para este ejemplo: N=8, n = 3, por lo que el número de muestras es:

$$ {}_8 C_3 = \frac{8!}{3!(8-3)!} = 56 $$

Ordenando los datos tenemos: 6, 7, 8, 9, 10, 11, 12, 13

Entonces todas las muestras de tamaño 3 así como sus respectivas medias son:

No.	Muestra			media	No	muestra			Media	No	muestra			Media	No.	muestra			media
1	6	7	8	21/3	15	6	9	13	28/3	29	7	9	12	28/3	43	8	10	13	31/3
2	6	7	9	22/3	16	6	10	11	27/3	30	7	9	13	29/3	44	8	11	23	31/3
3	6	7	10	23/3	17	6	10	12	28/3	31	7	10	11	28/3	45	8	11	13	32/3
4	6	7	11	24/3	18	6	10	13	29/3	32	7	10	12	29/3	46	8	12	13	33/3
5	6	7	12	25/3	19	6	11	12	29/3	33	7	10	13	30/3	47	9	10	11	30/3
6	6	7	13	26/3	20	6	11	13	30/3	34	7	11	12	30/3	48	9	10	12	31/3
7	6	8	9	23/3	21	6	12	13	31/3	35	7	11	13	31/3	49	9	10	13	32/3
8	6	8	10	24/3	22	7	8	9	24/3	36	7	12	13	32/3	50	9	11	12	32/3
9	6	8	11	25/3	23	7	8	10	25/3	37	8	9	10	27/3	51	9	11	13	33/3
10	6	8	12	26/3	24	7	8	11	26/3	38	8	9	11	28/3	52	9	12	13	34/3
11	6	8	13	27/3	25	7	8	12	27/3	39	8	9	12	29/3	53	10	11	12	33/3
12	6	9	10	25/3	26	7	8	13	28/3	40	8	9	13	30/3	54	10	11	13	34/3
13	6	9	11	26/3	27	7	9	10	26/3	41	8	10	11	29/3	55	10	12	13	35/3
14	6	9	12	27/3	28	7	9	11	27/3	42	8	10	12	30/3	56	11	12	13	36/3

3.- Tomando como base las medias de las muestras podemos ahora construir la siguiente distribución de probabilidad

x	$P(x)$
7	1/56
22/3	1/56
23/3	1/28
8	3/56
25/3	1/14
26/3	5/56
9	3/28
28/3	3/28
29/3	3/28
10	3/28
31/3	5/56
32/3	1/14
11	3/56
34/3	1/28
35/3	1/56
12	1/56
	1/1

4.- Calculemos ahora la media y la desviación estándar de la distribución muestral de la media

$\bar{x}$	$P(\bar{x})$	$\bar{x} \cdot P(\bar{x})$	$\bar{x}^2$	$\bar{x}^2 \cdot P(\bar{x})$
7	1/56	1/8	49	7/8
22/3	1/56	11/84	484/9	121/126
23/3	1/28	23/84	529/9	529/252
8	3/56	3/7	64	24/7
25/3	1/14	25/42	625/9	625/126
26/3	5/56	65/84	676/9	845/126
9	3/28	27/28	81	243/28
28/3	3/28	1	784/9	28/3
29/3	3/28	29/28	841/9	841/84
10	3/28	15/14	100	75/7
31/3	5/56	12/13	961/9	4805/504
32/3	1/14	16/21	1024/9	512/63
11	3/56	33/56	121	363/56
34/3	1/28	17/42	1156/9	289/63
35/3	1/56	5/24	1225/9	175/72
12	1/56	3/14	144	18/7
	1/1	19/2		183/2

Entonces tendremos que:

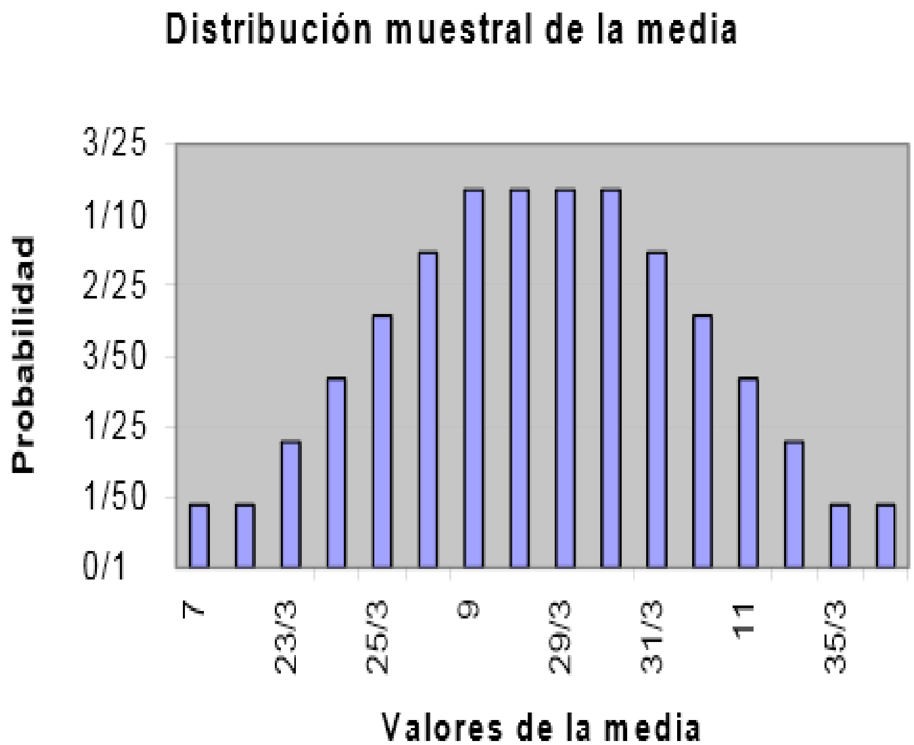
la media es igual a: $\mu_{\bar{x}} = \frac{19}{2}$

la varianza de la distribución es: $\sigma_{\bar{x}}^{2} = \frac{183}{2} - \left(\frac{19}{2}\right)^2 = \frac{5}{4}$

y la desviación estándar es: $\sigma_{\bar{x}} = \sqrt{\frac{5}{4}} = \frac{\sqrt{5}}{2}$

5.- Tomando como base la distribución de probabilidad para la media construimos la gráfica de la distribución muestral de la media.

La gráfica de la distribución muestral de la media es:



Podemos observar que la grafica tiene una forma semejante (aproximada) a una distribución normal. Como se establecerá más, adelante esta aproximación será cada vez mejor a medida que aumente el tamaño de la muestra.

Si este problema lo intentamos resolver cuando el ***muestreo es con reemplazo***, entonces tendremos que el número de muestras a extraer será de N^n, y por tanto para nuestro ejemplo, el número de muestras sería de: $8^3 = 512$. Realizar esta tarea requiere de mucho más de esfuerzo y tiempo.

Podemos observar que la grafica tiene una forma semejante (aproximada) a una distribución normal. Como se establecerá más, adelante esta aproximación será cada vez mejor a medida que aumente el tamaño de la muestra.

Si este problema lo intentamos resolver cuando el ***muestreo es con reemplazo***, entonces tendremos que el número de muestras a extraer será de N^n, y por tanto para nuestro ejemplo, el número de muestras sería de: $8^3 = 512$. Realizar esta tarea requiere de mucho más de esfuerzo y tiempo.

Podemos observar que cuando la población es finita y de tamaño moderado podemos construir una distribución muestral, obteniendo todas las muestras posibles de un tamaño dado, calculando para cada muestra el valor del estadístico que nos interesa y enumerando los diferentes valores calculados junto con sus probabilidades de ocurrencia.
Podemos hacer una aproximación de las verdaderas distribuciones muestrales basadas en poblaciones muy grandes o infinitas, sacando un gran número de muestras aleatorias es decir, de una población cualquiera se extraen k muestras; cada una permite calcular k estadísticos con los cuales se puede hacer un histograma como el de la derecha de la figura 1.

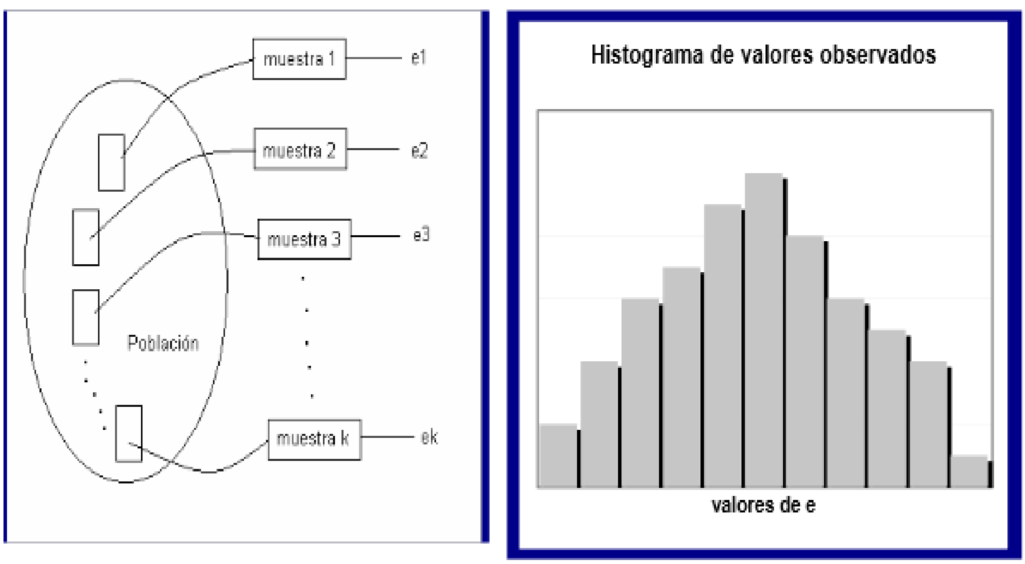


Figura 1

Se aprecia que el histograma adquiere forma de campana si se suavizan los escalones, al reducir los intervalos. La curva obtenida a partir de datos muestrales, observados a través del muestreo, tiende asintóticamente a otra curva teórica a medida que k aumenta, y los intervalos se hacen infinitesimales.

Ejemplos, ejercicios o casos de aplicación práctica

1. En un curso se va a elegir delegado y subdelegado entre 4 candidatos. ¿Cuántas hojas diferentes se pueden rellenar?
2. Si las matrículas de los coches constan de dos letras (elegibles entre 26) y cuatro dígitos. ¿Cuántas matrículas diferentes se pueden crear?
3. Se tiene una bolsa con 4 bolas numeradas del 1 al 4. Se realiza el experimento que consiste en la extracción de una bola de la bolsa, observar el número obtenido y reintegrar la bola a la bolsa.Se pide:

a. Hallar de cuántas formas diferentes se extraen las 4 bolas.

b. Hallar de cuántas formas diferentes se extraen las 4 bolas, si no se reintegran a la bolsa.

4. Un experimento consiste en la extracción de tres cartas de una baraja española.

a. Hallar de cuántas formas diferentes se extraen las 3 cartas si se reintegran a la baraja.

b. Hallar de cuántas formas diferentes se extraen las 3 cartas, si no se reintegran a la baraja.

5. Se tiene una bolsa con 8 bolas numeradas del 1 al 8. Se realiza el experimento que consiste en la extracción de una bola de la bolsa, observar el número obtenido y reintegrar la bola a la bolsa. Se pide:

a. El suceso "Obtener número primo".

b. El suceso "Obtener número par".

c. El suceso "Obtener número primo o par".

6. En el experimento aleatorio lanzar dos monedas al aire y observar el resultado obtenido, ¿cuándo decimos que se ha verificado el suceso "obtener al menos una cara"?
7. En el experimento aleatorio lanzar dos monedas al aire y observar el resultado obtenido enuncia dos sucesos compuestos.
8. Se aplica a una población de 100 individuos un muestreo aleatorio estratificado tomando una muestra de 12 elementos. Si los estratos los forman 10, 20, 30 y 40 individuos, ¿cuántos se toman en cada estrato con una afijación igual?, ¿y si la afijación es proporcional?

Síntesis de cierre del tema

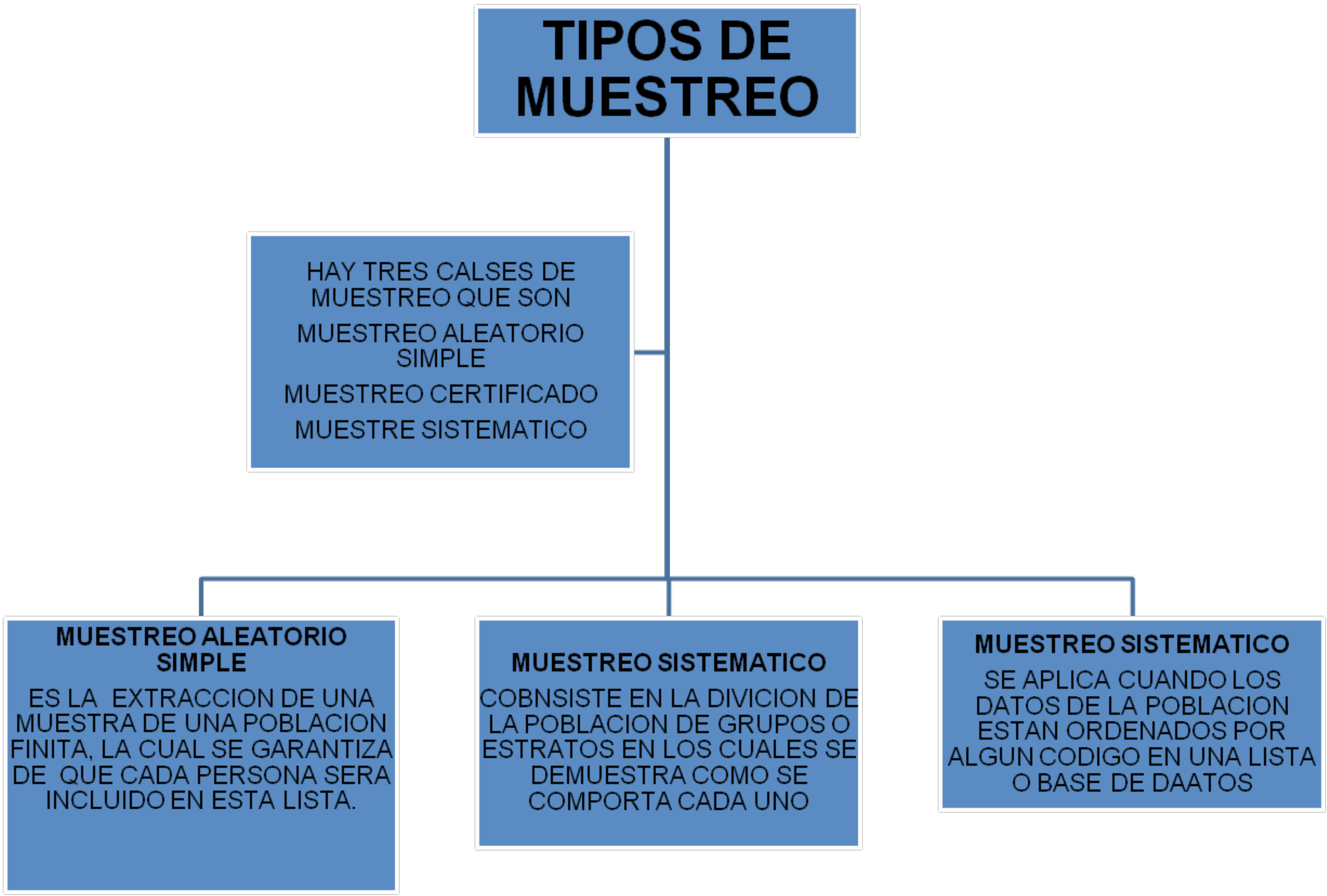


Imagen 1. Estadística – mapas conceptuales.
Recuperado de http://coflonorteltda.blogspot.com/2010/03/estadistica-mapas-conceptuales.html

Actividades auto-evaluativas propuestas al estudiante

1. Un investigador desea determinar el tamaño de muestra mínimo usando un diseño MAS para estimar la talla promedio de las ratas en la población con una confianza de 95% y un error de estimación no mayor a 2 centímetros si en la población hay 200 ratas. ¿Cuántas ratas se debe elegir en la muestra?
2. En una fábrica de artículos electrónicos generalmente el 10% de los artículos presenta algún defecto de fabricación. Para analizar la calidad del producto se desea estimar la proporción de artículos electrónicos defectuosos de un lote de 2000 artículos listo para ser embarcado. ¿Cuántos artículos deben ser elegidos del lote si se desea una confianza de 95% y un error de estimación no mayor a 0.05?
3. Un centro comercial acaba de recibir un pedido de sintonizadores TDT para ponerlos a la venta entre sus clientes. Dichos sintonizadores vienen numerados con códigos desde el 39456 al 48795. El gerente de dicho centro está preocupado por la calidad de dichos sintonizadores y decide obtener una muestra sistemática de 7 aparatos y someterlos a varias pruebas. Ayúdale a obtener la muestra.
4. Una gran empresa ha solicitado a su departamento de informática que realice una aplicación que permita gestionar on-line las ventas en todas sus tiendas. Para hacer las primeras comprobaciones deciden elegir 4 tiendas. Para ello disponen de un fichero con 1728 filas en el que en las 6 primeras filas aparece una cabecera y en la séptima los campos que definen cada uno de los datos tomados de cada tienda (Dirección postal, Nombre del gerente, etc.). A continuación están el resto de filas con los datos concretos de cada una de las tiendas. Simula de forma razonada un muestreo sistemático e indica en qué filas aparecerán los datos de las tiendas que van a formar parte de la muestra.

¿Qué tipo de muestreo se ha utilizado en los siguientes casos descritos?

5. Tenemos que seleccionar una muestra de 60 estudiantes que cursan 3º de ESO , por lo que acudimos a un Instituto que tiene 10 grupos de 30 alumnos del curso que nos interesa. Seleccionamos dos de los grupos y la muestra sería el total de los alumnos de esos dos grupos seleccionados.
6. Tenemos una población de 2.000 viviendas de las que debemos seleccionar 200 y el investigador ha actuado de la siguiente manera: asigna a cada vivienda un número, después los introduce en urna y extrae 200 números al azar, que serán los que compondrán la muestra.
7. Queremos hacer una encuesta por teléfono por lo que utilizaremos el listín telefónico y el nombre que ocupe el número 32 de cada página del listín será un elemento de la muestra.

3

Unidad 3

Distribución de la media

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO

Introducción

En esta cartilla se abordan las distribuciones muestrales para grupos de datos, teniendo en cuenta los cálculos de la distribución para la media, la proporción y las diferencias aplicando el método de la estadística inferencial.

Es importante leer cuidadosamente el contenido de la cartilla. Al interpretar las expresiones matemáticas, le permite comprender los elementos expuestos para así aplicar los contenidos y asimilarlos.

Frecuentemente, para solucionar situaciones cotidianas, se requiere diferenciar los procedimientos de cálculo para resolver las diferentes contextos matemáticos.

U3

Metodología

El proceso de aprendizaje académico por medio de esta cartilla se desarrolla en tres etapas. La primera etapa es la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido por medio de la revisión de los ejercicios resueltos con el fin de ganar confianza en la solución de ejercicios propuestos. La segunda etapa es el análisis del video y el contenido del mismo. La tercera etapa es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

Cabe anotar que la segunda etapa puede ser retomada cuantas veces quiera, pero tiene un tiempo limitado para adquirir los conocimientos. Recuerde que debe avanzar con las siguientes cartillas y tener en cuenta las recomendaciones dadas a lo largo del avance del curso.

U3 Desarrollo temático

Componente Motivacional

Este tema sirve para solucionar situaciones reales o ideales de manera eficiente, aporta herramientas teóricas para identificar apropiadamente los cálculos de la distribución muestral en la media, la proporción y las diferencias.

Por ejemplo, identificar las diferencias entre las distribuciones muestrales con el fin de establecer los conceptos prácticos de la solución de ejercicios.

Recomendaciones académicas

La presente cartilla contiene la fundamentación teórica de los aspectos teóricos y prácticos sobre las distribuciones muestrales que permite evidenciar la solución de situaciones problémicas que conduzcan a la comprensión de ejercicios sobre los aspectos esenciales de la matemática financiera.

La matemática financiera es más sencilla de lo que nos parece. Es la utilización y aplicación de normas o leyes denominadas propiedades. Cuando se aprende apropiadamente estas propiedades, se puede afrontar de manera correcta cada una de las situaciones o ejercicios que nos solicita, por esto la matemática no es difícil lo difícil es comprender y aplicar las propiedades que orientan el trabajo en la asignatura. Todos, de una manera u otra podemos llegar a manejar, cada día mejor, la matemática.

Desarrollo de cada una de las unidades temáticas

Competencia general	Competencia específica
El estudiante reconoce las características de las relaciones de las distribuciones muestrales en la estadística inferencial.	El estudiante diferencia las características de la distribución de la muestra en los promedios en diferentes contextos.

El estudiante relaciona las características que contribuyen a la planeación, organización, dirección, coordinación, control y evaluación de la estadística inferencial.	El estudiante calcula la distribución muestral a partir de la media, la proporción y las diferencias estadísticas.

Distribución de la media

Si se extraen muestras aleatorias de tamaño cualquiera de una población estadística infinita que tiene media poblacional y varianza, entonces la distribución en la población estudiada se calcula a partir de:

i) El promedio para una muestra es igual al promedio de la población. Es decir, $\mu_g = \mu$

ii) La varianza de una muestra es igual a la varianza de una población divida en el número de datos del estudio estadístico. En consecuencia la desviación estándar de las medias muestrales (llamada también el error estándar de la media muestral), es igual a la deviación estándar poblacional dividida por la raíz cuadrada de n. Es decir :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

De acuerdo con los procesos estadisticos mencionados en Acuña (sf) tenemos que

> Si la población fuera finita de tamaño, entonces se aplica el factor de corrección al error estándar de la media muestral. Pero en la práctica este factor es omitido a menos que la muestra sea lo suficientemente grande comparada con la población.
>
> Si además la población se distribuye normalmente, entonces la media muestral también tiene una distribución normal con la media y varianza anteriormente indicadas. (Acuña, E. (s.f). Distribuciones muestrales. Recuperado de http://academic.uprm.edu/eacuna/miniman6.pdf).

$$\sqrt{\frac{N-n}{N-1}}$$

Distribución de la proporción muestral

Al tener una población distribuida binomialmente con probabilidad de éxito, se utiliza una muestra aleatoria de tamaño, entonces se puede mostrar que el promedio de X, *es decir, el número de éxitos:* p número de éxitos en la muestra, es $\mu = np$ y que su varianza es $\sigma^2 = npq$. En consecuencia la proporción de la muestra p = X / n tiene medida p, y varianza pq / n. Entonces

$$z = \frac{X - np}{\sqrt{npq}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

La gráfica que se presenta puede ser una distribución de probabilidad normal, si los valores son mayores a 5, se considera una muestra confiable y si son cercanos a 0 y a 1 se debe tomar un tamaño de muestra más grande.

Es importante, destacar que la variable aleatoria solo puede tomar dos valores (p éxito o q fracaso), esto implica que en el desarrollo del ejercicio estadístico, se considere una distribución binomial B (n,p) que es muy cercana a la distribución normal N.

Si se tiene una muestra de tamaño menor que 30 datos, la expresión matemática para la distribución normal en un promedio es:

$$N = \left(p, \sqrt{\frac{pq}{n}} \right)$$

Donde p es la proporción de uno de los valores que presenta la variable estadística en la población y q = 1- p. (Teoria del muestreo. Recuperado de https://es.scribd.com/doc/24927306/TEORIA-DEL-MUESTREO-Muestras-Aleatorias-).

Distribución de la muestra de una diferencia

Al tener dos tipos de poblaciones diferentes, la primera que tiene un promedio μ_1 y desviación estándar σ_1 , y la segunda con promedio μ_2 y desviación estándar σ_2 Con una muestra aleatoria, elegida de la población 1 e independiente de la muestra aleatoria 2; se puede calcular las distribuciones para ambas promedios, a partir de la expresión que se conoce como distribución muestral de las diferencias entre medias o la distribución muestral del estadístico $\overline{x}_1 - \overline{x}_2$

La distribución es aproximadamente normal para $n_1 \geqq 30$ y $n_2 \geqq 30$. Si las poblaciones .

La expresión matemática que se utiliza para calcular esta distribución es:

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ejemplo:

En un estudio estadístico se decide comparar los pesos promedio de un grupo de cereales en un granero, usando una muestra aleatoria de 20 bultos de arroz y 25 bultos de cebada; se sabe que en el análisis practicado para ambas muestras se trabaja con una distribución normal; el peso promedio del arroz es de 100 libras y su σ es igual a 14.142; mientras que el peso promedio de la cebada es 85 libras y su σ es 12.247. Si $\overline{x}_1$ representa el promedio de los pesos de los 20 bultos de arroz y $\overline{x}_2$ es el promedio de los pesos de una muestra de 25 bultos de cebada, encuentre la probabilidad de que el promedio de los 20 bultos de arroz sea al menos 20 libras más grande que el de los 25 bultos de cebada.

Solución:

Datos:

μ_1 = 100 libras

μ_2 = 85 libras

σ_1 = 14.142 libras

σ_2 = 12.247 libras

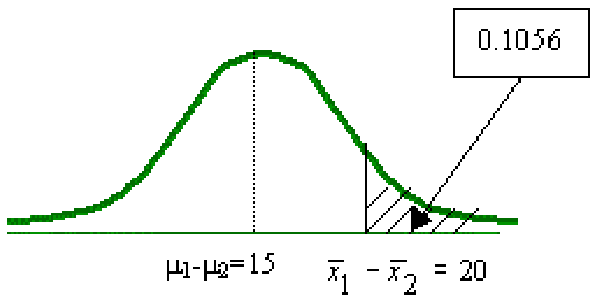
n_1 = 20 bultos de arroz

n_2 = 25 bultos de cebada

$p(\bar{x}_1 - \bar{x}_2 \rangle 20) = ?$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{20 - (100 - 85)}{\sqrt{\frac{(14.142)^2}{20} + \frac{(12.247)^2}{25}}} = 1.25$$

Por lo tanto, la probabilidad de que el promedio de los pesos de la muestra de arroz sea al menos 20 libras más grande que el de la muestra de cebada es 0.1056.



Ejemplo:

En un taller de mecánica se hace la medición de la vida media de los carburadores de los automóviles que se reparan, los automóviles A tiene una vida media de 7.2 años, con σ igual a 0.8 años; mientras que los automóviles B tiene una vida media de 6.7 años y una σ igual a 0.7. Halle la probabilidad de que una muestra aleatoria de 34 automoviles tenga una vida promedio de al menos 1 año más que la de una muestra aleatoria de 40 automoviles de la marca B.

Solución:

Datos:

μ_A = 7.2 años

μ_B = 6.7 años

σ_A = 0.8 años

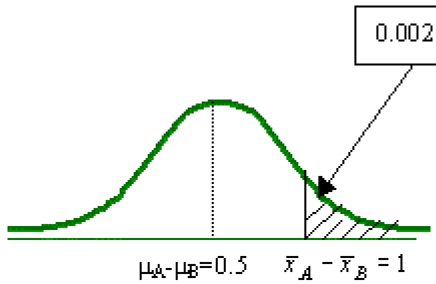
σ_B = 0.7 años

n_A = 34 tubos

n_B = 40 tubos

$p(\bar{x}_A - \bar{x}_B \rangle 1) = ?$

$$z = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{1 - (7.2 - 6.7)}{\sqrt{\frac{(0.8)^2}{34} + \frac{(0.7)^2}{40}}} = 2.84$$



Ejemplo:

El rendimiento en km/L de 2 tipos de combustible tienen una σ de 1.23 km/l para el combustible 1 y una σ de 1.37 km/L para el 2 combustible; el primer combustible se prueba eb 35 máquinas y el segundo en 42 máquinas.

a. ¿Cuál es la probabilidad de que la primer combustible de un rendimiento promedio mayor de 0.45km/L que la segundo combustible?

b. ¿Cuál es la probabilidad de que la diferencia en rendimientos promedio se encuentre entre 0.65 y 0.83km/L a favor del combustible 1.

Solución:

σ_1 = 1.23 Km/Lto

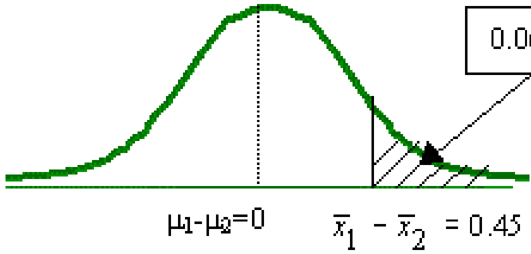
σ_2 = 1.37 Km/Lto

n_1 = 35 máquinas

n_2 = 42 máquinas

a. $p(\bar{x}_1 - \bar{x}_2 \rangle 0.45) = ?$

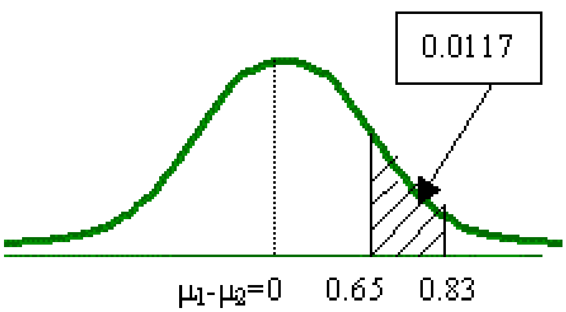
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{0.45 - 0}{\sqrt{\frac{(1.23)^2}{35} + \frac{(1.37)^2}{42}}} = 1.52$$



b. $p(0.65 \leq \bar{x}_1 - \bar{x}_2 \leq 0.83) = ?$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{0.65 - 0}{\sqrt{\frac{(1.23)^2}{35} + \frac{(1.37)^2}{42}}} = 2.19$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{0.83 - 0}{\sqrt{\frac{(1.23)^2}{35} + \frac{(1.37)^2}{42}}} = 2.80$$



La probabilidad de que la diferencia en rendimientos promedio en las muestras se encuentre entre 0.65 y 0.83 Km/Lto a favor del combustible 1 es de 0.0117.

Distribución de la muestra para la diferencia de dos proporciones

En el trabajo con muestras de dos poblaciones que sean binomiales y al mismo tiempo muestrales, la distribución de la muestra de la diferencia de dos proporciones se identifica con la relación ($n_1p_1 \geq 5$, $n_1q_1 \geq 5$, $n_2p_2 \geq 5$ y $n_2q_2 \geq 5$). Entonces p1 y p2 tienen distribuciones muestrales aproximadamente normales, así que su diferencia p_1-p_2 también tiene una distribución normal. (Restrepo, C. (s.f.). Compulibro estadística II. Recuperado de http://www.sites.upiicsa.ipn.mx/polilibros/portal/Polilibros/P_terminados/ESTADISTICA/Estadistica_old/Polilibro/CE/UNIDAD).

La expresión matemática para una distribución de la muestra para la diferencia de dos proporciones, así: para el cálculo de la distribución muestral de diferencia de proporciones es:

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

Ejemplo:

Un grupo de personas que viven en Bogotá realizan una encuesta sobre los índices de violencia en la ciudad, el 12% de encuestados son hombres y están de acuerdo con que la violencia ha aumentado, mientras que el 10% de mujeres encuestadas hacen la misma afirmación; si la encuesta se hace a 100 hombres y 100 mujeres; halle la probabilidad de que el porcentaje de hombres sea al menos 3% mayor que el de las mujeres.

Solución:

Datos:

$P_H = 0.12$

$P_M = 0.10$

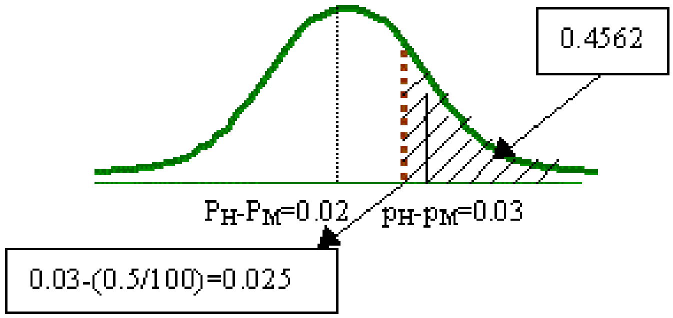
$n_H = 100$

$n_M = 100$

$p(p_H - p_M\ 0.03) \geq\ = ?$

Se recuerda que se está incluyendo el factor de corrección de 0.5 por ser una distribución binomial y se está utilizando la distribución normal. (Restrepo, C. (s.f.). Compulibro estadística II. Recuperado de http://www.sites.upiicsa.ipn.mx/polilibros/portal/Polilibros/P_terminados/ESTADISTICA/Estadistica_old/Polilibro/CE/UNIDAD).

$$z = \frac{(p_H - p_M) - (P_H - P_M)}{\sqrt{\frac{P_H q_H}{n_H} + \frac{P_M q_M}{n_M}}} = \frac{0.025 - (0.12 - 0.10)}{\sqrt{\frac{(0.12)(0.88)}{100} + \frac{(0.10)(0.90)}{100}}} = 0.11$$



La probabilidad de que el porcentaje de hombres, al menos 3% mayor que el de mujeres es de 0.4562.

Ejemplo:

Se sabe que 3 de cada 6 parea de zapatos fabricados por la máquina 1 son defectuosos y que 2 de cada 5 pares de zapatos fabricados por la máquina 2 son defectuosos; se toman muestras de 120 pares de zapatos de cada máquina:

a. ¿cuál es la probabilidad de que la proporción de zapatos defectuosos de la máquina 2 rebase a la máquina 1 en por lo menos 0.10?
b. ¿cuál es la probabilidad de que la proporción de zapatos defectuosos de la máquina 1 rebase a la máquina 2 en por lo menos 0.15?

Solución:

Datos:

$P_1 = 3/6 = 0.5$

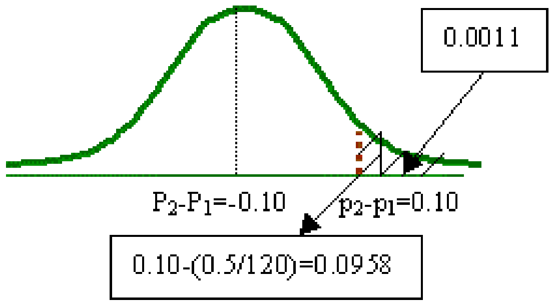
$P_2 = 2/5 = 0.4$

$n_1 = 120$ objetos

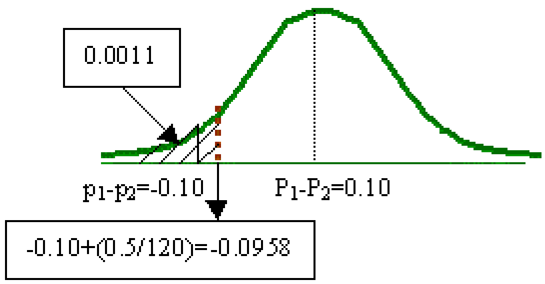
$n_2 = 120$ objetos

$p(p_2 - p_1 \geq 0.10) = ?$

$$z = \frac{(p_2 - p_1) - (P_2 - P_1)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{0.0958 - (-0.10)}{\sqrt{\frac{(0.50)(0.50)}{120} + \frac{(0.40)(0.60)}{120}}} = 3.06$$



Otra manera de hacer este ejercicio es poner P_1-P_2:



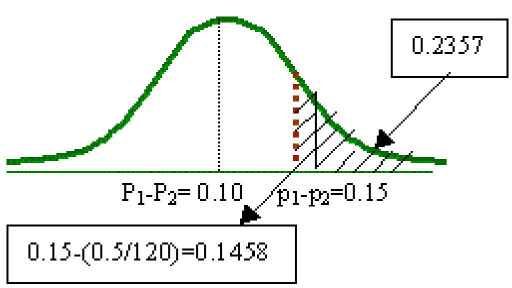
$$z=\frac{(p_1-p_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}}=\frac{-0.0958-0.10}{\sqrt{\frac{(0.50)(0.50)}{120}+\frac{(0.40)(0.60)}{120}}}=-3.06$$

La probabilidad de que exista una diferencia de proporciones de pares de zapatos defectuosos de por lo menos 10% a favor de la máquina 2 es de 0.0011. (Restrepo, C. (s.f.). Compulibro estadística II. Recuperado de http://www.sites.upiicsa.ipn.mx/polilibros/portal/Polilibros/P_terminados/ESTADISTICA/Estadistica_old/Polilibro/CE/UNIDAD).

p(p1-p2≥

0.15)=?

$$z=\frac{(p_1-p_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}}=\frac{0.1458-0.10}{\sqrt{\frac{(0.50)(0.50)}{120}+\frac{(0.40)(0.60)}{120}}}=0.72$$



La probabilidad de que exista una diferencia de proporciones de pares de zapatos defectuosos de por lo menos 15% a favor de la máquina 1 es de 0.2357.

Ejemplos, ejercicios o casos de aplicación práctica

1. Si se toma una muestra aleatoria simple (m.a.s) de tamaño 40 de una población considerada normal. Su varianza muestral es 4. Calcular la probabilidad.
2. Para conocer la proporción de niños a los que no les gusta la sopa. Realizamos una encuesta a una muestra (m.a.s) de tamaño 100. Si se conoce que dicha proporción es del 40%. Calcular la probabilidad de que la muestra llegue a una proporción superior al 46%.
3. El número de automóviles que tanquean en una estación de gasolina en una hora es un valor desconocido. Para estimar dicho valor uno de los empleados de la estación de gasolina plantea que dicho valor es de 8 vehículos a la hora otro, en cambio, da la estimación de 10 a la hora. Estimar el número de vehívulos que realizan su tanqueo por hora.
4. Una marca reconocida de cereal realiza una prueba en la que conoce que 2 de cada 10 consumidores reconocen su producto en una prueba a ciegas. ¿Cuántas pruebas ciegas de sabor deberían hacerse para que la proporción de que los que conocen la marca oscile entre el 16% y el 24% con una probabilidad mínima de 0,8?
5. Un cocinero realiza experimento acerca del peso de una cantidad de pollo de una marca determinada. Tiene 500 rebanadas de pollo, con un peso medio de 5.02 gramos y una desviación típica de 0.30 gramos cada una. Calcula la probabilidad de que, una muestra aleatoria de 100 rebanadas de pollo tenga un peso de 100 gramos.

Síntesis de cierre del tema

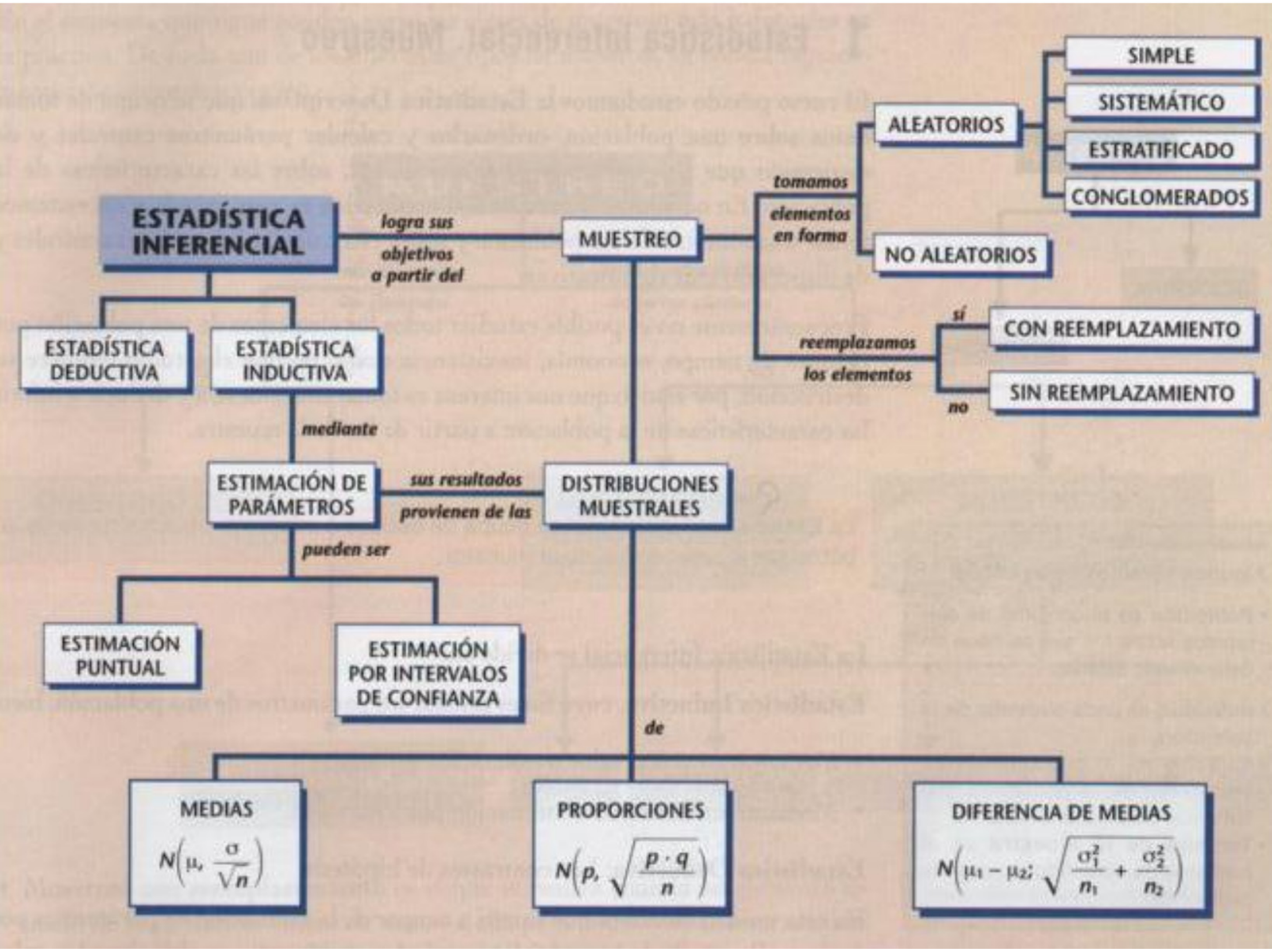


Imagen 1. Páez, P (s.f.). Teorema del límite central.
Recuperado de http://www.uam.es/personal_pdi/stmaria/jmurillo/Met_Inves_Avan/Presentaciones/Teorema_del_Limite_Central.pdf

Actividades auto-evaluativas propuestas al estudiante

1. Un banco ha realizado estudios de las cuentas y su distribución normal con un promedio de $ 2.000.000 y una σ de 600.000; si se toman al azar 100 cuentas. ¿Cuál es la probabilidad de el promedio de la muestra este entre $ 1.000.000 y $900.000?
2. En una peluquería, se ha observado que en los sábados asiste un promedio de 248 personas a realizarse maricure, con una desviación típica de 6 personas. Considerando que la distribución de muestreo es normal: ¿Qué porcentaje de promedios de la muestra estará entre 155 y 169 personas? ¿Qué porcentaje de los promedios de la muestra será mayor o igual de 170 personas?
3. Se cuenta con trescientas cajas que contienen libros de Probabilidad, cada caja tiene un peso medio de 5.02 kg. y una desviación típica de 0.3 kg. Encuentra la probabilidad de que en una muestra al azar de 100 cajas el peso promedio de cada caja esté comprendido entre 4.96 y 4.98 kg.
4. En una estación de policia ubicada enfrente de un parque, se han registrado en promedio 475 actas levantadas todas ellas en el transcurso de la noche, con una desviación típica de 12 actas. Considerando que la distribución de muestro es normal: Encuentra la probabilidad de que los valores medios de las muestras sean mayores o iguales a 450 actas. Encuentra la probabilidad de que los valores medios de las muestras sean menores o iguales a 482 actas.
5. En la zona centro de la Ciudad, se observó que en un 20% de las familias, uno de sus integrantes o más, presentan problemas de salud debido a la contaminación ambiental. Una muestra aleatoria de 150 familias dió una proporción del 27% de familias con la característica señalada. Si suponemos que el valor del 20% es correcto, cuál será la probabilidad de obtener una proporción mayor o igual a la observada (es decir respecto al 27%)?
6. En el evento del concurso de belleza, una de las participantes obtuvo el 46% de los votos. Hallar la probabilidad de que en un muestreo al azar de 200 jueces calificadores, se obtenga mayoría a su favor.
7. Un estudio llevado a cabo en el Municipio de Z, dió como resultado que, el 70% de una población de 100 niños de primaria asistió a consulta con el médico general, por lo menos una vez en su vida. ¿Cuál es la probabilidad de que la proporción de los niños que asistió a consulta con el médico general, por lo menos una vez en su vida, esté entre 0.65 y 0.75? ¿Cuál es la probabilidad de que la proporción sea mayor que 0.75?
8. En una investigación realizada entre los habitantes del municipio X, se ha encontrado que el l8% de ellos, ha tenido problemas de tránsito. Se seleccionó una muestra aleatoria de 100 personas ¿Cuál es la probabilidad de que, entre el l5% y el 25% haya tenido contacto con policías de tránsito?
9. Se ha encontrado que el 60% del personal ejecutivo de la empresa “El Farolito”, realiza con regularidad “la verificación”

de su automóvil. Extrayendo una muestra aleatoria de 150 personas de dicho nivel. ¿Cuál es la probabilidad de que la proporción muestral esté comprendida entre 0.5 y 0.7?

10. En el proceso de envasado de los afamados refrescos GUSTO, se presenta una producción promedio en la que el 10 % de las botellas no están completamente llenas. Se selecciona una muestra al azar de 225 botellas de un lote de 625 envases llenos. ¿Cuál es la probabilidad de que la proporción muestral de botellas parcialmente llenas, se encuentre en el intervalo que va de 9 a 11 % ?

4

Unidad 4

Pruebas de hipótesis

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO

Introducción

En esta cartilla se realiza el estudio de los modelos de pruebas de hipótesis, además de la aplicación de estos modelos a situaciones cotidianas.

Estos modelos matemáticos pueden aportar herramientas para interpretar situaciones cotidianas, para aprender estos elementos teóricos se deben leer los contenidos del módulo, comprender sus procedimientos y manejar estos contenidos en la aplicación.

U4

Metodología

El proceso de aprendizaje académico por medio de la cartilla se desarrolla en tres etapas. La primera hace referencia a la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido. La segunda etapa es el análisis del video y el contenido del mismo. Y la tercera etapa es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

La segunda etapa puede ser retomada cuantas veces quiera, pero tiene un tiempo limitado para adquirir los conocimientos. Recuerde que debe avanzar con las restantes cartillas y con ello los módulos.

U4 Desarrollo temático

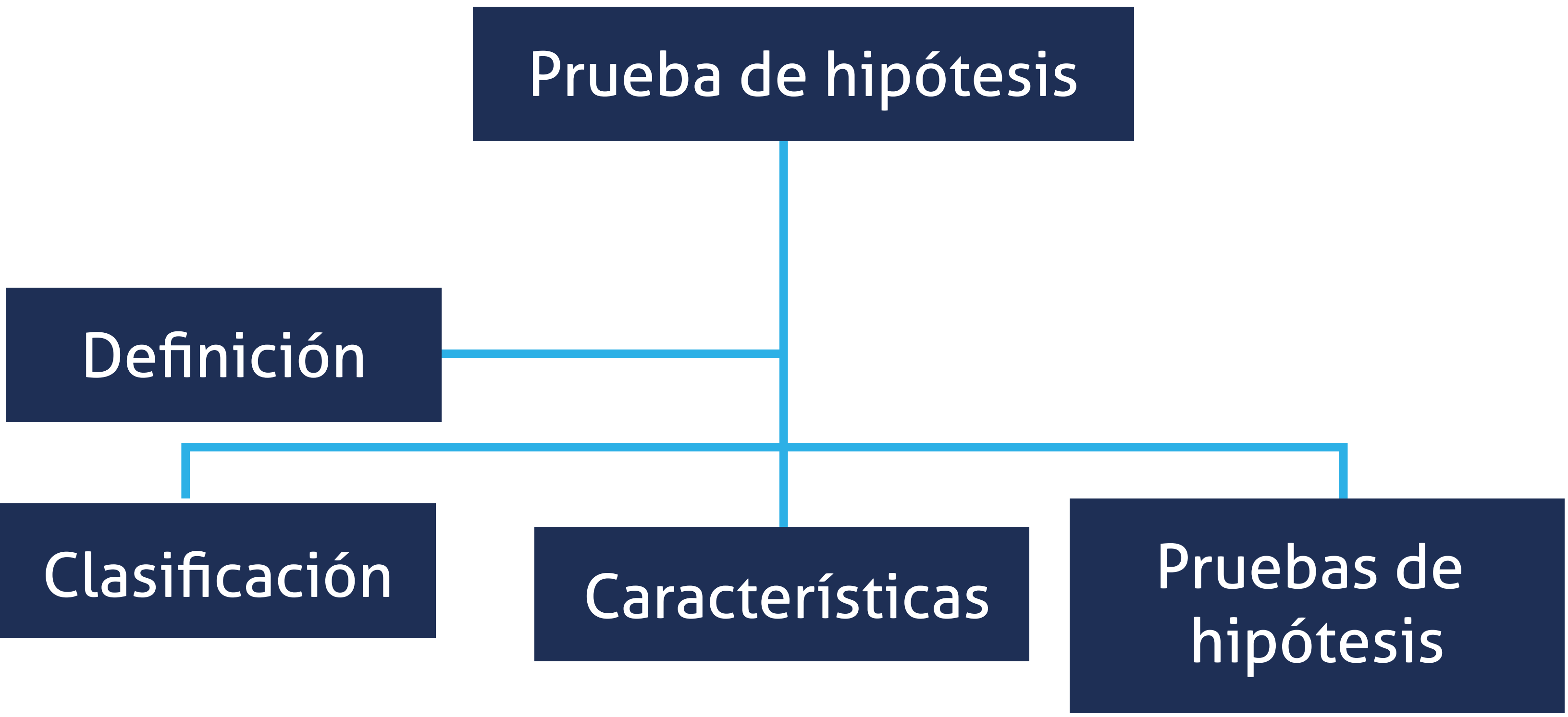


Imagen 1
Fuente: propia.

Componente Motivacional

La modelación de una prueba de hipótesis con su respectiva aplicación es fundamental en el desarrollo del conocimiento matemático en cualquier profesión, en especial cuando se trata de áreas del conocimiento como en la que usted está preparándose.

Es de relevante importancia la construcción de modelos lineales para la interpretación y solución de situaciones de la cotidianidad.

Recomendaciones académicas

La presente cartilla contiene elementos fundamentales sobre las pruebas de hipótesis a partir de las características y definiciones de sus elementos básicos.

Es fundamental el aprender a modelar apropiadamente sobre las pruebas de hipótesis para permitir la interpretación y comprensión de situaciones de la cotidianidad. Es fundamental la lectura comprensiva de los contenidos, interpretar los procedimientos y realizar las actividades durante la semana para que usted no se atrase en el proceso.

Desarrollo de cada una de las unidades temáticas.

Pruebas de hipótesis

La prueba de hipótesis se define como uno de los métodos más importantes para determinar la validez de un proceso estadístico, según García (2005).[1]

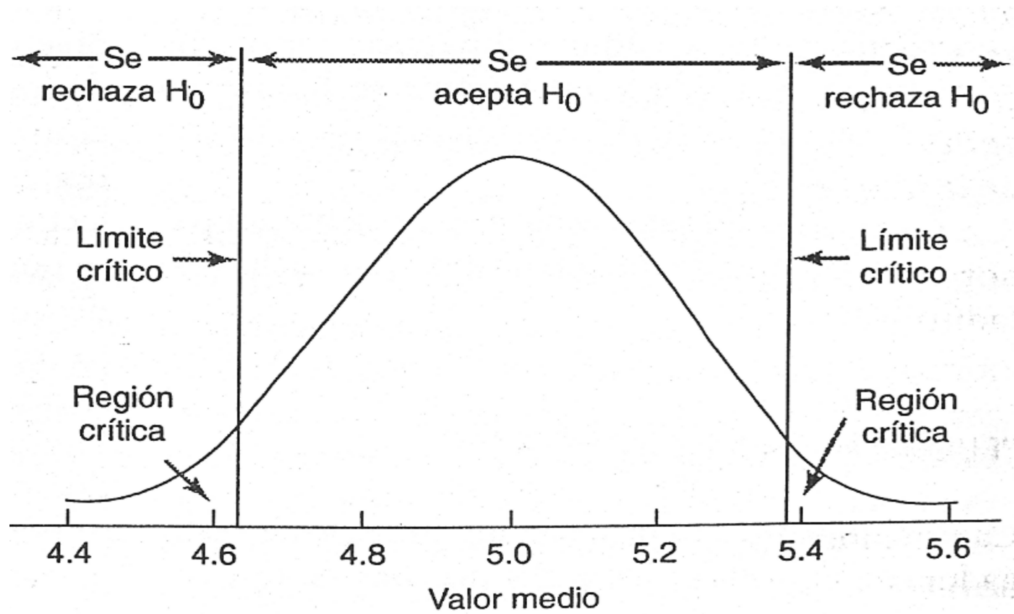
Este método consiste, en definitiva, en un proceso de toma de decisiones. Normalmente antes de iniciar una investigación se parte de una hipótesis lo que implica siempre la exclusión de otras. **La hipótesis nula** (Ho) , también llamada la hipótesis de no diferencia, indica que no existen diferencias significativas entre los resultados obtenidos en la práctica y los resultados teóricos, es decir, que no hay relación real entre las variables y que cualquier relación observada es producto del azar o la casualidad, o debida a las fluctuaciones del muestreo.

La necesidad de contar con una hipótesis nula radica en que la comprobación estadística de la hipótesis constituye generalmente un proceso de rechazo de ésta. Si bien resulta imposible demostrar en forma directa y concluyente que la primera explicación (hipótesis científica) es correcta, sí es posible demostrar que es muy probable que la hipótesis nula sea incorrecta y que dicho indicio apoya la hipótesis científica, de modo que el investigador pretende mediante la aplicación de pruebas estadísticas rechazar la hipótesis nula.

Si se tiene un intervalo de confianza y se desea determinar un procedimiento para probar la hipótesis se tiene que:

- Establecer la hipótesis nula
- Fijar el nivel de confianza
- Determinar el valor critico z (o t), correspondiente al nivel de confianza fijado y el número de grados de libertad, que será diferente en función de la prueba utilizada
- Calcular el error típico de la media
- Calcular el error muestral
- Calcular el intervalo de confianza
- Interpretar en términos de rechazo o no rechazo la Ho. Si el parámetro hipotéticamente establecido se encuentra entre los valores del intervalo de confianza, no podemos rechazar la Ho. Si el parámetro hipotético no se encuentra dentro del intervalo de confianza rechazamos la Ho y aceptamos la H1.

[1] Tomado de: Análisis de datos en los estudios epidemiológicos IV Estadística Inferencial. En: http://www.nureinvestigacion.es/FICHEROS_ADMINISTRADOR/F_METODOLOGICA/formacion_19.pdf recuperada el 06 - 06- 2015



Gráfica de prueba de hipótesis

Hipótesis nula y alternativa

La prueba de hipótesis involucra una pareja de hipótesis estadísticas, ya la aceptación de una implica el rechazo de la otra. Un investigador define como HIPOTESIS ALTERNATIVA H_1 a los experimentos que se pueden encontrar en un proceso de probar una muestra dada; mientras que la HIPOTESIS NULA H_0 que se describe como una afirmación sobre la relación de igualdad cuando un parámetro de población se compara con un valor numérico.

Pruebas de hipotesis

Las pruebas de hipótesis se clasifican como direccionales o no direccionales, dependiendo de cuando la hipótesis alternativa H_1 involucra el signo (ǂ). Si la afirmación de H_1 contiene ǂ, entonces la prueba se llama no direccional, mientras que si la afirmación involucra < ó >, entonces la prueba se llama direccional.

Prueba de hipotesis para la media

Si se tiene una muestra aleatoria que presenta varianza σx2 y media µx se puede calcular la hipótesis a partir de la expresión

$$Z_X = \frac{\bar{x} - \mu_x}{s_x / \sqrt{n}}$$

Siendo cada valor la desviación media, la desviación estándar, la población y el estadístico de prueba.

Prueba de hipotesis para la población

$$Z = \frac{\bar{X} - \mu}{\sigma^2/\sqrt{n}}$$

Ejemplos[2]

Se afirma que una brilladora gasta un promedio mínimo de 46 kilowatt-hora al año. Si una muestra aleatoria de 12 hogares que se incluye en un estudio planeado indica que las brilladoras gastan un promedio de 42 kilowatt-hora al año con una desviación estándar de 11.9 kilowatt-hora, ¿esto sugiere con un nivel de significancia de 0.05 que las aspiradoras gastan, en promedio, menos de 46 kilowatt-hora anualmente?

Solución:

1. Sea μ: gasto promedio en kilowatt-hora al año de una aspiradora.

   μ= 46 kilowatt-hora, s= 11.9 kilowatt-hora

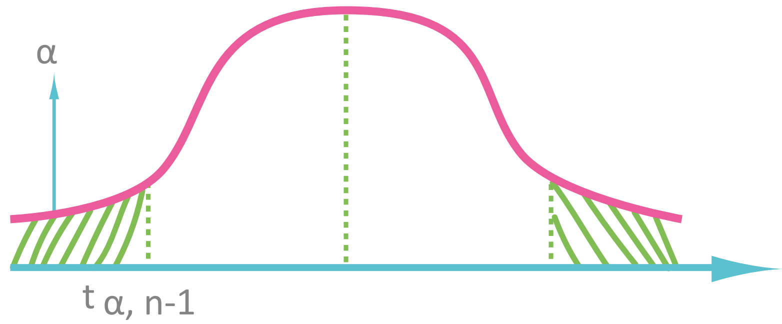
   $\bar{x}$ = 42 kilowatt-hora

   $n = 12$

2. H0 : μ ≥ 46
3. H1 : μ < 46
4. Nivel de significancia: 0.05
5. Estadístico de prueba: Como la varianza de la población es desconocida y el tamaño de muestra es menor de 30, y se supuso que la población es norma, utilizaremos la distribución t de student en el cálculo del estadístico:

$$t_0 = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

6. Región de Rechazo:



$t_{0,05,11} = 1.796$

Si $t_0 < -1,796$ Rechazar la H_0

Si $t_0 \geq -1,796$ Aceptar la H_0

Se evalua t_0 para los datos:

$$t_0 = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16$$

El peso en libras de una muestra aleatoria de niños de 2 años sigue una distribución normal con una desviación de 1.21 libras. Según se ha establecido, en promedio un niño de esta edad debe pesar alrededor de 25 libras. Un pediatra sin embargo considera que ahora los bebés han variado su peso y para ello ha considerado el peso de 100 bebés de esta edad obteniendo un peso medio de 25.3 libras. Con un nivel de confianza del 5%, pruebe si el pediatra tiene razón en lo planteado.

Solución:

1. Sea μ: peso promedio en libras de una población de niños de 2 años

   μ_0 = 25 Libras, σ = 1.21 libras

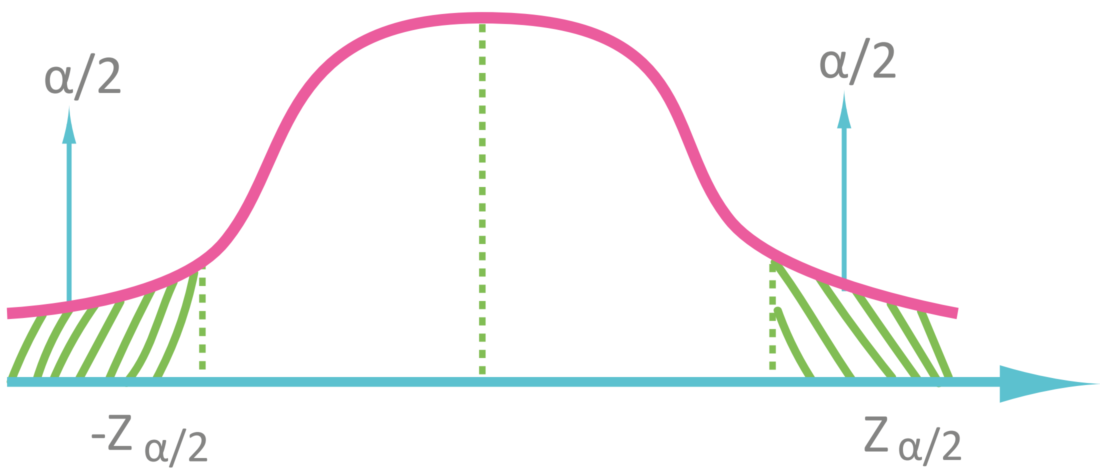
[2] Tomado de: https://jaimesotou.files.wordpress.com/2011/05/prueba-hipotesis-1.pdf

$\bar{x} = 25.3$ libras

$n = 100$

2. $H_0 : \mu = 25$
3. $H_1 : \mu \neq 25.3$
4. Nivel de significancia: 0.05
5. Estadístico de prueba: Como la varianza de la población es conocida, n>30 y la distribución es normal, utilizamos el estadístico z:

$$Z = \frac{\bar{X} - \mu}{\sigma^2/\sqrt{n}}$$



Puesto que es un intervalo bilateral se utiliza $z_{\alpha/2} \rightarrow$ Por la tabla II se obtiene el estadístico $z_{0,975} = 1.96$

Si $z_0 < -1{,}96$ ó $z_0 \geq 1{,}96$ Rechazar la H_0

Si $1.96 \leq z0 \leq 1{,}96$ Aceptar la H_0

Se evalua t_0 para los datos:

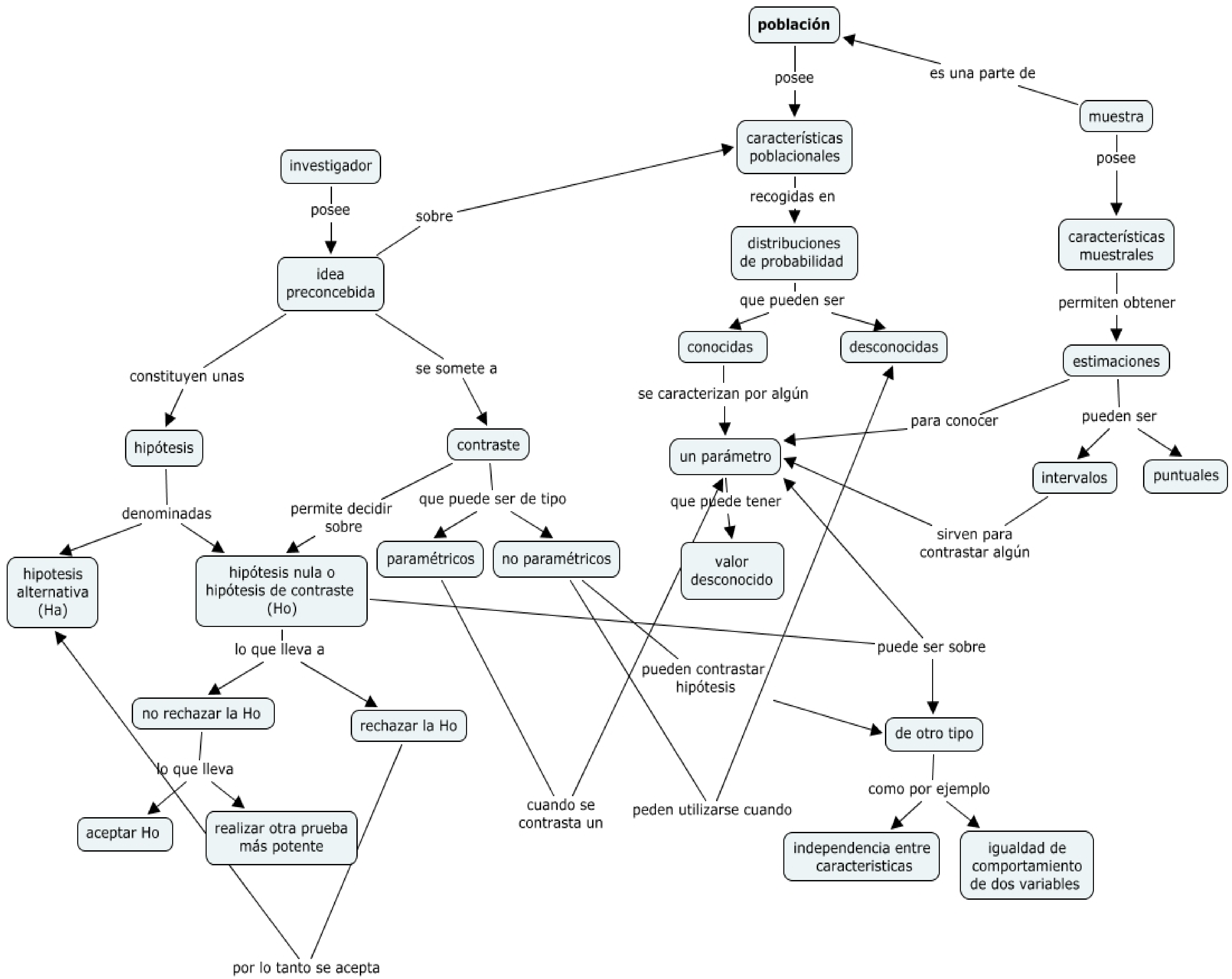
$$z_0 = \frac{14.3 - 14}{1.21/\sqrt{100}} = 2.48$$

Como $z_0 = 2.48 >$1.96, se rechaza Ho y se concluye con un nivel de significancia del 0.05 que el peso promedio de todos los bebés de seis meses ha variado según las pruebas disponibles.

Ejemplos, ejercicios o casos de aplicación práctica

1. Un manufacturero de juguetes Tailandés reclama que solo un 10% de los osos de juguete hechos para hablar están defectuosos. Cuatrocientos de éstos juguetes se sometieron a prueba de forma aleatoria y se encontró que 50 estaban defectuosos. Pruebe el reclamo del manufacturero con un nivel de significancia de 5%.
2. Una agencia de empleos afirma que el 80% de todas las solicitudes hechas por mujeres con hijos prefieren trabajos a tiempo parcial. En una muestra aleatoria de 200 solicitantes mujeres con niños, se encontró que 110 prefirieron trabajos a tiempo parcial. Pruebe la hipótesis de la agencia con un nivel de significancia de 5%.
3. Nacionalmente, un 16 % de los hogares tiene una computadora personal. En una muestra aleatoria de 80 hogares en Baltimore, solo 13 poseían una computadora personal. Con un nivel de significancia de 5%, pruebe si el porciento de hogares en Baltimore que tienen computadoras personales es menor que el porcentaje nacional.
4. El registrador de cierta universidad ha dicho que está dispuesto a permitir una sección del curso ESTAD 121 una vez a la semana si más del 65% de los estudiantes matriculados en el curso expresan que prefieren el curso una vez a la semana, en vez de dos veces a la semana. En una muestra aleatoria de 40 estudiantes, 26 indicaron su preferencia de una vez a la semana. Usando un nivel de significancia de 0.01, debe el registrador autorizar el ofrecimiento del curso ESTAD 121 una vez a la semana?
5. Los sistemas de escape de emergencia para aviones son impulsados por un combustible sólido. Una de las características importantes de este producto es la rapidez de combustión. Las especificaciones requieren que la rapidez promedio de combustión sea de 50 cm/s. Se sabe que la desviación estándar de esa rapidez es de El experimentador decide especificar un nivel de significancia, de 0.05. Selecciona una muestra aleatoria de n = 25 y obtiene una rapidez promedio muestral de combustión de ¿A qué conclusión debe legar?

Síntesis de cierre del tema[3]



[3] Tomado de: http://cmapserver.unavarra.es/rid=1195816861953_1078711613_2075/Inferencia%203.cmap

Actividades auto-evaluativas propuestas al estudiante.

1. La duración media de una muestra de 100 pilas producidos por una compañía resulta ser 157 horas, con una desviación estándar de 20 horas. Si μ es la duración media de todos los tubos producidos por la compañía, compruebe la hipótesis horas contra la hipótesis horas con un nivel de significancia de a) 0.01, b) 0.05. 1600.
2. En un banco consideran que los clientes tienen problemas de pago si se han retrasado en sus pagos por más de 120 días. Se revisaron los datos de los pagos realizados por 450 personas físicas y 210 empresas, seleccionados aleatoriamente. Se encontró que en las cuentas de las 150 personas físicas, ocho tenían vencimientos por más de 120 días; mientras que en los reportes de las cuentas de las empresas, doce tenía vencimientos por más de 120 días. Pruebe la hipótesis nula de que las proporciones son iguales. Utilice $\alpha = 0.05$.
3. En un hotel se desea saber si hay diferencias en los niveles de productividad entre los dos turnos laborales. Con ese objetivo analizó los niveles de atención durante 36 meses. La media del turno de la mañana es de 127 clientes con una desviación estándar de 20 clientes. También durante 36 meses, el turno de la tarde tuvo una media de 168 unidades con una desviación estándar de 44 clientes. ¿Afirmaría que el turno de la tarde tiene mayor productividad? Utilice un nivel de significancia de 0.05.
4. Una muestra aleatoria de 400 hombres y otro de 500 mujeres de una determinada población reveló que 150 hombres y 170 mujeres estaban a favor de cierto candidato. ¿Se puede concluir a un nivel de significación del 15% que la proporción de hombres a favor del candidato es mayor que la proporción de mujeres?
5. Los pesos en libras de una muestra aleatoria de bebés de seis meses son: 14.6, 12.5, 15.3, 16.1, 14.4, 12.9, 13.7 y 14.9. Haga una prueba con nivel de 5% de significancia para determinar si el peso promedio de todos los bebés de seis meses es distinto a 14 libras, suponga que sus pesos se distribuyen normalmente.

4

Unidad 4

Análisis de varianza

Estadística inferencial

Autor: Nelly Yolanda Cespedes Guevara

AREANDINA
Fundación Universitaria del Área Andina
MIEMBRO DE LA RED
ILUMNO

Introducción

En esta cartilla se presentan los contenidos necesarios para conocer los aspectos necesarios sobre el análisis de varianza y regresión lineal, teniendo en cuenta el análisis de la realización de un proyecto y la definición de los diferentes valores comerciales.

Estos modelos matemáticos pueden aportar herramientas para interpretar situaciones cotidianas, para aprender estos elementos teóricos se deben leer los contenidos del módulo, comprender sus procedimientos y manejar estos contenidos en la aplicación.

U4

Metodología

El proceso de aprendizaje académico por medio de la cartilla se desarrolla en tres etapas. La primera hace referencia a la lectura completa del módulo, realizando la respectiva interpretación, análisis y comprensión del contenido. La segunda etapa es el análisis del video y el contenido del mismo. Y la tercera etapa es la aplicación de los contenidos expuestos a situaciones generales, particulares y específicas.

Cabe anotar que la segunda etapa puede ser retomada cuantas veces quiera, pero tiene un tiempo limitado para adquirir los conocimientos. Recuerde que debe avanzar con las restantes cartillas y con ello los módulos.

U4

Desarrollo temático

Componente motivacional

La regresión lineal y el análisis de varianza con su respectiva aplicación es fundamental en el desarrollo del conocimiento de la estadística inferencial en cualquier profesión, en especial cuando se trata de áreas del conocimiento como en la que usted está preparándose.

Es de relevante importancia la construcción de evaluación de proyectos para la interpretación y solución de situaciones de la cotidianidad.

Recomendaciones académicas

La presente cartilla contiene elementos fundamentales sobre la regresión lineal y el análisis de varianza.

Es fundamental el aprender a modelar apropiadamente la evaluación de proyectos para permitir la interpretación y comprensión de situaciones de la cotidianidad. Es fundamental la lectura comprensiva de los contenidos, interpretar los procedimientos y realizar las actividades durante la semana para que usted no se atrase en el proceso.

Desarrollo de cada una de las unidades temáticas

Análisis de varianza

Cuando se analiza la varianza se puede verificar las diferencias estadísticas de dos muestras que presentan el mismo planteamiento matemático de solución.

Según Payares y Navarro (s.f.) las características de la varianza son:

> *Una varianza grande indica que hay mucha variación entre los sujetos, que hay mayores diferencias individuales con respecto a la media; una varianza pequeña nos indica poca variabilidad entre los sujetos, diferencias menores entre los sujetos. La varianza cuantifica todo lo que hay de diferente entre los sujetos u observaciones.*
>
> *El análisis de varianza no constituye un método o procedimiento único; según los diseños y datos disponibles existen diversos modelos de análisis de varianza.*

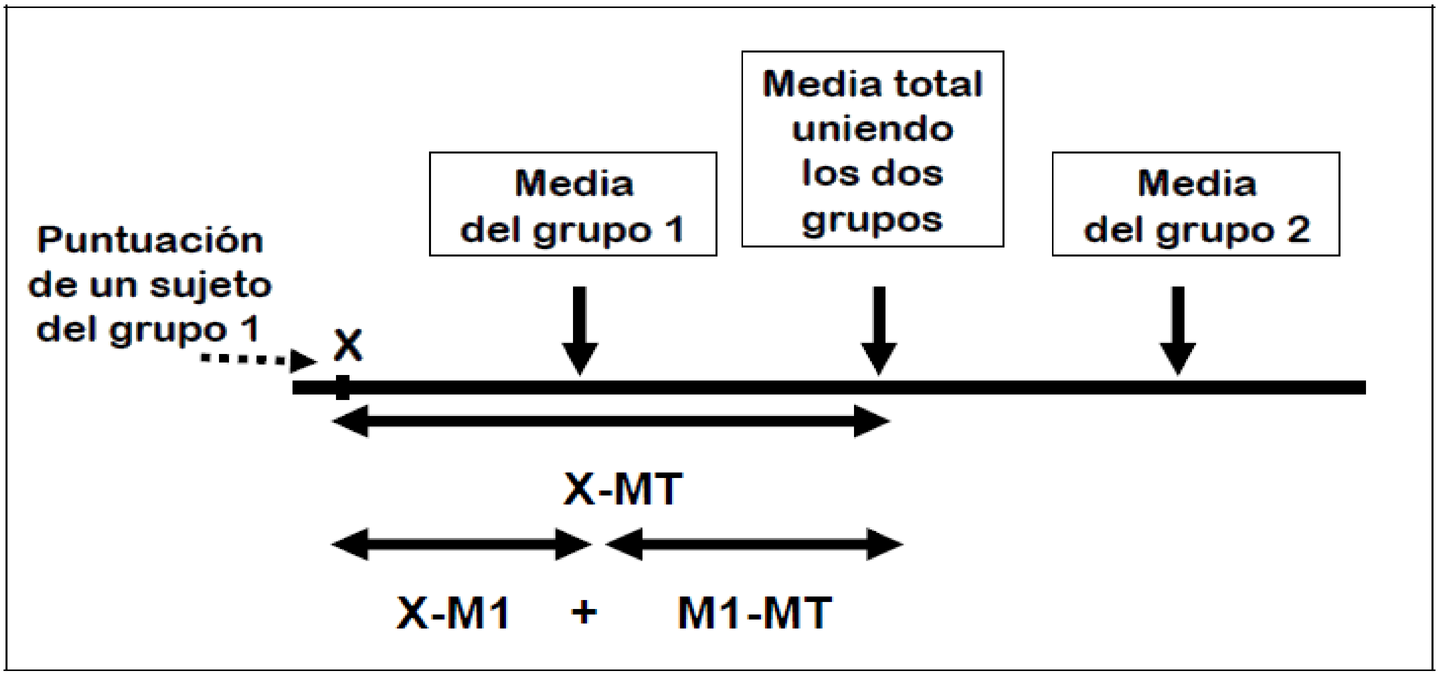
La expresión matemática es:

$$\sigma^2 = \frac{\sum(X-M)^2}{N-1}$$

En la figura se presenta el esquema de finalidad de la varianza, presentado por Payares y Navarro (s.f.).

1. Dos grupos o muestras, cada uno con su media (M1 y M2).

2. El grupo formado por las dos muestras con la media del total de ambos grupos (MT).
3. La puntuación (X) de un sujeto del primer grupo. Los puntos indicados en la figura representan las dos medias, la *media total y la puntuación* X de un sujeto concreto del grupo 1 (y podría hacerse la misma representación con todos los sujetos de todos los grupos).



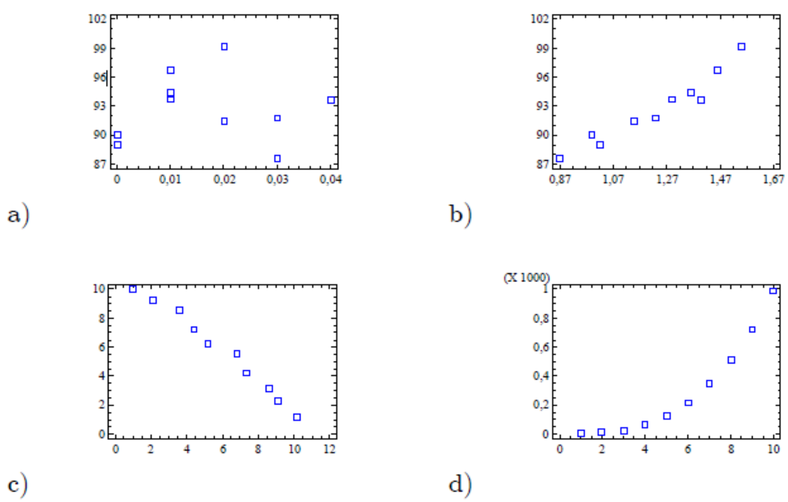
Regresión lineal

La regresión lineal permite establecer conexiones entre las variables a partir de modelos que permiten explicar cómo incide una variable en otra y viceversa.

En este sentido, se desarrolla el método de regresión lineal que permite:

Los métodos de regresión estudian la construcción de modelos para explicar o representar la dependencia entre una variable respuesta o dependiente (Y) y la(s) variable(s) explicativa(s) o dependiente(s), X[2].

La representación gráfica de una regresión lineal se encuentra a partir de las estructuras de los diagramas de dispersión o de puntos, así que



En a) hay ausencia de relación (independencia).

En b) existe asociación lineal positiva (varían en general en el mismo sentido).

En c) existe asociación lineal negativa (varían en sentido contrario).

En d) existe fuerte asociación, pero no lineal.

La expresión matemática de la regresión lineal responde a la ecuación. (Regresión lineal simple. Recuperado de http://clubensayos.com/Biografías/Weew/818491.html)

$Y = \beta 0 + \beta 1X + \varepsilon$

Este modelo de regresión permite calcular los errores aleatorios, es decir, $E[\varepsilon/X = x] = E[\varepsilon] = 0$, y por lo tanto:

$E[Y/X = x] = \beta 0 + \beta 1x + E[\varepsilon/X = x] = \beta 0 + \beta 1x$

En dicha expresión se observa que:

- La media de Y, para un valor fijo x, varía linealmente con x.
- $y = E[Y / X = x] = \beta 0 + \beta 1x$, ó $Y = \beta 0 + \beta 1X$.
- El parámetro β0 son los valores que puede tomar la variable en x y β1 hace referencia al valor de la pendiente calculada en y para la variable aleatoria.

i) La varianza de ε es constante para cualquier valor de x, es decir,

$Var(\varepsilon/X = x) = \sigma^2$

ii) La distribución de ε es normal, de media 0 y desviación σ.

iii) Los errores asociados a los valores de Y son independientes unos de otros.

Ejemplos

Cinco grupos de semillas clasificadas en 2, 3, 5, 7 y 8 tipos de semillas pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.

Hallar la ecuación de la recta de regresión de las semillas sobre el peso. ¿Cuál sería el peso aproximado de un grupo de seis semillas?

x_i	y_i	$x_i - y_i$	x_i^2	y_i^2
2	14	4	196	28
3	20	9	400	60
5	32	25	1024	160
7	42	49	1764	294
8	44	64	1936	352
25	152	151	5320	894

Luego se procede a calcular la ecuación de la recta de regresión lineal.

$$\bar{x} = \frac{25}{5} = 5 \qquad \bar{y} = \frac{152}{5} = 30.4$$

$$\sigma_x^2 = \frac{151}{5} - 5^2 = 5.2 \qquad \sigma_y^2 = \frac{5320}{5} - 30.4^2 = 139.84$$

$$\sigma_{xy} = \frac{894}{5} - 5 \cdot 30.\ 4 = 26.8$$

$$x - 5 = 0.192\ (y - 30) \qquad x = 0.192y - 0.76$$

$$y - 30.4 = 5.15\ (x - 5) \qquad y = 5.15x + 4.65$$

$$y = 5.15 \cdot 6 + 4.65 = 35.55\ \text{Kg}$$

Ejemplos, ejercicios o casos de aplicación práctica

1. Un fabricante de tijeras tiene que elegir entre tres colores para las cajas de tijeras: verde, rojo y azul. Para averiguar si el color influye en las ventas, se eligen 26 tiendas de tamaño parecido. Se envían cajas verdes a 6 de estas tiendas, cajas rojas a 5 y cajas azules a las 5 restantes. Después de unos días, se comprueba el númerode cajas vendidas en cada tienda.

Rojo	Amarillo	Azul
43	52	61
52	37	29
59	38	38
76	64	53
61	74	79
81		

Calcule la suma de los cuadrados dentro de los grupos, entre los grupos y total.

Complete la tabla del análisis de la varianza y contraste la hipótesis nula de que las medias poblacionales de los niveles de ventas de las cajas de los tres colores son iguales.

2. Un profesor tiene una clase de 23 estudiantes. Al comienzo de cada bimestre asigna a cada estudiante aleatoriamente a uno de los cuatro monitores: M1, M2, M3, M4. La tabla adjunta muestra las calificaciones obtenidas por los estudiantes que trabajan con los monitores.

M1	M2	M3	M4
72	78	80	79
69	93	68	70
84	79	59	61
76	97	75	74
64	88	82	85
81	68	63	

Calcule la suma de los cuadrados dentro de los grupos, entre los grupos y total.

Complete la tabla del análisis de la varianza y contraste la hipótesis nula de la igualdad de las medias poblacionales de las calificaciones de estos monitores.

3. Una serviteca desea corresponder a sus clientes por su confianza, en especial con los neumáticos para camiones que más acreditan el negocio. Ofrece dos marcas que tienen la misma duración, pero no está muy seguro de su variabilidad. Selecciona muestras de tamaño 16 y 21 neumáticos de cada marca, con varianzas de 36.000 y 42.000 km2. ¿Hay evidencia de una diferencia entre las varianzas poblacionales?

4. Supongamos la selección de dos muestras, de tamaños n1 = 21 y n2 = 20, cuyas varianzas son s2 = 57.12 y s2 = 25.68, respectivamente. ¿Dan estos resultados suficiente evidencia, que nos haga pensar que las varianzas poblacionales sean distintas?

5. Tres proveedores suministran piezas en envíos de 800 unidades. Se han comprobado minuciosamente muestras aleatorias de 10 envíos de cada uno de los tres proveedores y se ha anotado el número de piezas que no se ajustan a las normas. La tabla muestra este número.

Proveedor A	Proveedor B	Proveedor C
28	22	33
37	27	29

34	29	39
29	20	33
31	18	37
33	30	38

6. Cinco niños de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos. Hallar la ecuación de la recta de regresión de la edad sobre el peso. ¿Cuál sería el peso aproximado de un niño de seis años.

Síntesis de cierre del tema

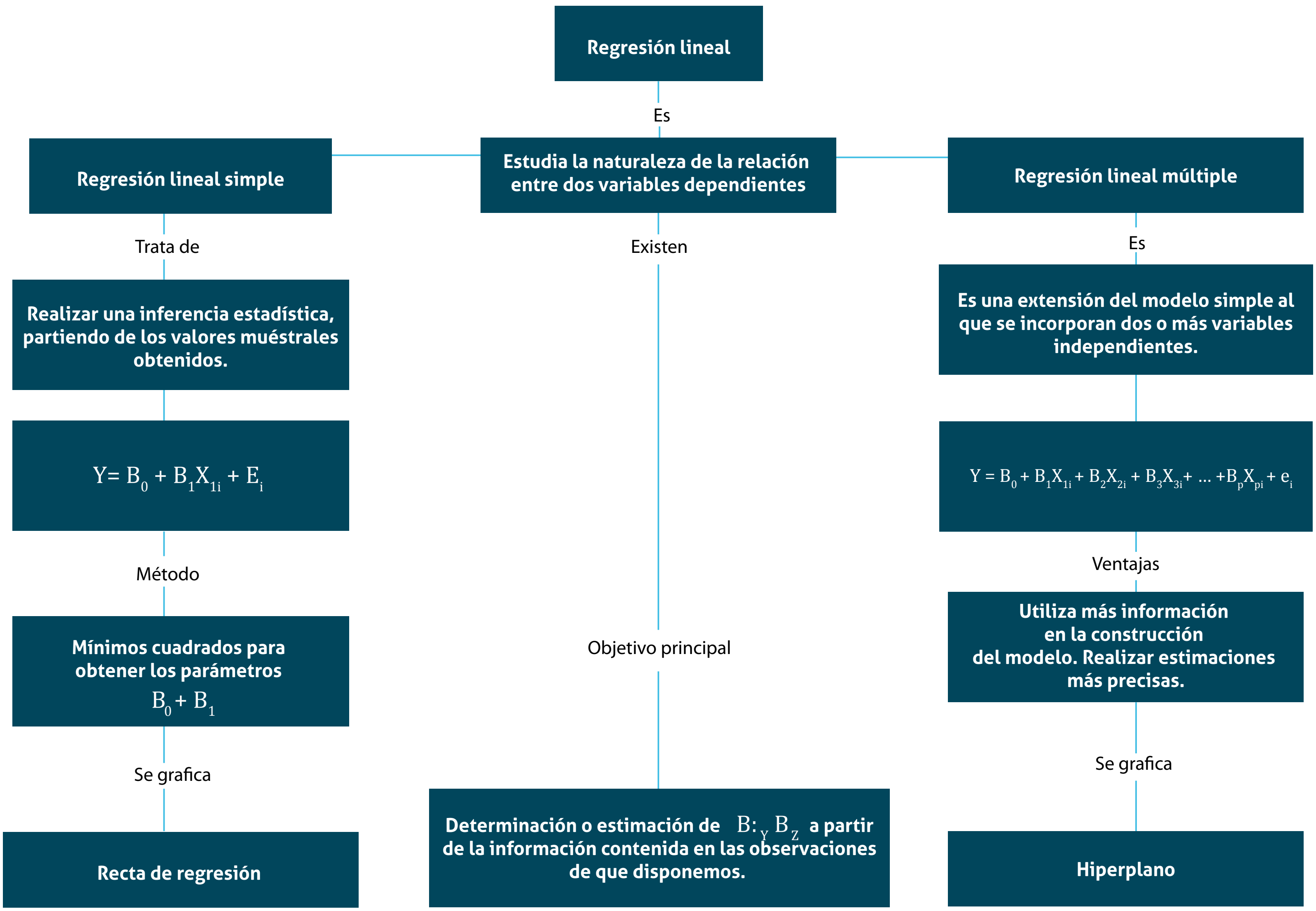


Imagen 1. Regresión lineal simple y múltiple. Recuperado de http://deaingenieriaambientalerikaflores3a.blogspot.com/2012/09/regresion-lineal-simple-y-multiple.html

Actividades auto-evaluativas propuestas al estudiante

7. Un equipo de ciclistas se dividen al azar en 4 grupos que entrenan con métodos diferentes. El primer grupo realiza largos recorridos a ritmo pausado, el segundo grupo realiza series cortas de alta intensidad y el tercero trabaja en el gimnasio con pesas y se ejercita en el pedaleo de alta frecuencia y el cuarto grupo entrena con aeróbicos. Después de un mes de entrenamiento se realiza un test de rendimiento consistente en un recorrido cronometrado de 15 Km. Los tiempos empleados fueron los siguientes, un nivel de confianza del 65% ¿Puede considerarse que los tres métodos producen resultados equivalentes?

Método I	Método II	Método III
15	14	13
16	13	12
14	15	11
15	16	14
17	14	11

8. Se realiza una prueba de memoria sobre distintas palabras, obteniéndose los siguientes resultados, ¿Qué conclusiones pueden sacarse acerca de las cuatro formas de presentación, con un nivel de significación del 5%?

Procdmt.I	Procdmt.II	Procdmt.III	Procdmt.IV
5	9	8	1
7	11	6	3
6	8	9	4
3	7	5	5
9	7	7	1
7		4	4
4		4	
2			

3. Un test de aptitud (X) de seis estudiantes que aplican una prueba de ventas en el primer mes (Y) en cientos de pesos.

x	25	42	33	54	29	36
y	42	72	50	90	45	48

Hallar el coeficiente de correlación e interpretar el resultado obtenido.

Calcular la recta de regresión de Y sobre X. Predecir las ventas de un vendedor que obtenga 50 en el test.

4. Una serviteca desea corresponder a sus clientes por su confianza, en especial con los neumáticos para camiones que más acreditan el negocio. Ofrece dos marcas que tienen la misma duración, pero no está muy seguro de su variabilidad. Selecciona muestras de tamaño 16 y 21 neumáticos de cada marca, con varianzas de 36.000 y 42.000 km^2. ¿Hay evidencia de una diferencia entre las varianzas poblacionales?

5. Se ha solicitado a un grupo de 80 personas información sobre el número de horas que dedican diariamente a caminar y leer. La clasificación de las respuestas ha permitido elaborar la siente tabla:

No. De horas caminadas (X)	6 7 8 9 10
No. De horas de lectura(Y)	4 3 3 2 1
Frecuencias (fi)	3 15 23 11 3

Calcular el coeficiente de correlación.

Determinar la ecuación de la recta de regresión de Y sobre X.

Si una persona camina 5 horas y media, ¿cuánto cabe esperar que vea lea un libro? (Análisis de varianza. Recuperado de http://www.ugr.es/~jsalinas/weproble/T14res.PDF).

Bibliografia

- Freund, J. Miller, I. & Miller, M. (2000). Estadística Matemática con aplicaciones. Madrid: editorial Prentice Hall.
- Kennett,Ron. y Zacks, S. (2000). Estadística Industrial Moderna. Barcelona: editorial Thomson.
- Newbold, P. (1988). Estadística para los Negocios y la Economía. Madrid: editorial Prentice Hall.
- Gutierrez, H. & De La Vara, R. (2005). Control estadístico de Calidad y Seis Sigma (6s). México: editorial Mc Graw Hill.
- Montgomery, D. & Runger, G.(2002). Probabilidad y Estadística aplicadas a la Ingeniería. México: editorial Limusa.
- Walpole, R. Myers, R. & Myers, S. (1998). Probabilidad y Estadística para Ingenieros. Madrid: editorial Prentice Hall.

Esta obra se terminó de editar en el mes de noviembre
Tipografá Myriad Pro 12 puntos
Bogotá D.C,-Colombia.

AREANDINA
Fundación Universitaria del Área Andina

MIEMBRO DE LA RED
ILUMNO