ESTADISTICA Y PROBABILIDADES

Editado Por
G. Aaron Estuardo Morales

ADMIN2766@ GMAIL.COM

Chile 2012

Indice

Contenido **Página**

Unidad Nº1: Estadística Descriptiva

Unidad Nº2: Probabilidades

Unidad N°3: Intervalos de Confianza

Unidad N°4: Pruebas de Hipótesis

Unidad Nº5: Regresión Lineal

Unidad N°1: Estadística Descriptiva

Introducción

La Estadística, nace de las necesidades reales del hombre. La variada y cuantiosa información relacionada con éste y que es necesaria para la toma de decisiones, hace que la estadística sea hoy, una importante herramienta de trabajo.

Entre las tareas principales de la Estadística, está el de reunir la información integrada por un conjunto de datos, con el propósito de obtener conclusiones válidas del comportamiento de éstos, como también hacer una inferencia sobre comportamientos futuros.

En cuanto al uso y la aplicación, puede decirse que abarca todo el ámbito humano encontrándose en las relaciones comerciales, financieras, políticas, sociales, etc. siendo fundamental en el campo de la investigación y en la toma de decisiones.

Es así también como en el área de las empresas de servicio y manufactura es posible realizar un análisis profundo del proceso estadístico al control de la productividad y de la calidad.

Estadística

Es el conjunto de métodos y procedimientos que implican recopilación, presentación, ordenación y análisis de datos, con el fin que a partir de ellos puedan inferirse conclusiones.

Pueden distinguirse dos ramas diferentes en Estadística:

– ***Estadística Descriptiva***, la cual es la que se utiliza en la descripción y análisis de conjuntos de datos o población.

– ***Inferencia Estadística***, la cual hace posible la estimación de una característica de una población, o la toma de una decisión con respecto a una población, con base únicamente en resultados muestrales.

Conceptos de elementos utilizados en el análisis estadístico

1) Población o Universo: Conjunto completo de individuos, objetos, o medidas los cuales poseen una característica común observable y que serán considerados en un estudio.

2) Muestra: Es un subconjunto o una porción de la población.

3) Variable: Característica o fenómeno de una población o muestra que será estudiada, la cual puede tomar diferentes valores.

4) Datos: Números o medidas que han sido recopiladas como resultado de la observación.

5) Estadístico: Es una medida, un valor que se calcula para describir una característica a partir de una sola muestra.

6) Parámetro: Es una característica cuantificable de una población.

Recopilación de Información

La **Estadística Descriptiva** tiene como función el manejo de los datos recopilados en cuanto se refiere a su ordenación y presentación, para poner en evidencia ciertas características en la forma que sea más objetiva y útil.

Una **población o universo** objeto de una investigación estadística puede ser **finita** si sus elementos se pueden contar. Por ejemplo, número de alumnos de un curso.

Una **población o universo** es **infinita** cuando no es finita. En Estadística, el sentido del término población infinita se refiere a una población con un número tan grande de elementos que no le es posible al investigador someter a medida cada uno de ellos.

Cuando se miden cualitativamente las características de una población, resultan categorías que deben ser **exhaustivas**, es decir, que se pueda clasificar a toda la población, y también deben ser mutuamente **excluyentes**, es decir, un mismo elemento no puede pertenecer simultáneamente a dos o más categorías. Por ejemplo, sexo de una persona: masculino o femenino.

Una **muestra** debe cumplir ciertas condiciones, de aquí surge el concepto de **muestra aleatoria** que es aquella obtenida de modo que cada elemento de la población tiene una oportunidad igual e independiente de ser elegido.

La **investigación estadística** es toda operación orientada a la recopilación de información sobre una población.

La investigación puede ser tan simple como la recopilación de datos estadísticos obtenidos de informaciones provenientes de fuentes oficiales a nivel institucional o de publicaciones de organismos altamente especializados en estas materias, o tan complejas que requiera de la colaboración de especialistas en diferentes materias, como ocurre en los censos de población de un país.

Se denomina **variable** a fenómenos o características que son medidas en algún tipo de investigación estadística.

Variables

Es muy probable que un especialista en Estadística que realiza una encuesta desee desarrollar un instrumento que le permita hacer varias preguntas y manejar diversos fenómenos o características. A estos fenómenos o características se les denomina ***variables aleatorias***.

Según la forma en que se expresen las variables, se dividen en:

1) Variables Cualitativas: son aquellas que pueden expresarse sólo en forma de atributo.

Ejemplo:

1) Estado civil :
- soltero
- casado
- viudo
- separado

2) Satisfacción con un producto:
- muy insatisfecho
- regularmente insatisfecho
- neutral
- satisfecho
- muy satisfecho

3) Tamaño de un tablero :
- grande
- mediano
- pequeño

2) Variables Cuantitativas, son aquellas variables que pueden expresarse en forma numérica. Se dividen en discretas y continuas.

2.1) Variables Cuantitativas Discretas, son respuestas numéricas que surgen de un proceso de conteo, siendo siempre un número entero.

Ejemplos :

1) Número de asignaturas inscritas en el primer semeste.
2) Número de integrantes del grupo familiar.
3) Número de salas de clases del IPVG.

2.2) Variables Cuantitativas Continuas, son respuestas numéricas que surgen de un proceso de medición, las cuales pueden tomar valores entre dos números enteros.

Ejemplo :

1) Estatura
2) Temperatura
3) Peso

Tabulación de los datos

En los experimentos estadísticos los datos recolectados pueden corresponder a una **población** o **muestra**. En ambos casos los procedimientos de resumen de datos son análogos y designaremos por:

$N =$ Tamaño de la población estudiada
$n =$ Tamaño de la muestra (parte de la población)

Con el objeto de realizar un mejor estudio de los datos es necesario organizar éstos, mediante el uso de distribuciones de frecuencia.

Una **distribución de frecuencia** es una tabla resumen en la que se disponen los datos divididos en grupos ordenados numéricamente y que se denominan **clases o categorías**.

A) Tabulación de datos cualitativos

La construcción de una distribución de frecuencia de atributos o distribución de frecuencia de variable cualitativa es simple, basta enumerar los diversos atributos con su respectiva frecuencia de ocurrencia.

***Frecuencia absoluta* :** (f_i) indica el número de veces que se repite un atributo.

Ejemplo:

Considérese una muestra 400 trabajadores de una cierta empresa de la región los cuales han sido encuestados sobre su actual estado civil. La información es tabulada de la siguiente manera:

Estado Civil	f_i
Soltero	75
Casado	200
Viudo	50
Separado	75
Total	400

$n = 400$ (tamaño de la muestra)
$m = 4$ (número de clases)

B) Tabulación de variable cuantitativa

Distinguiremos dos casos:

B.1) Tabulación de variable discreta (que toma un conjunto pequeño de datos distintos)

Las tablas de frecuencia de variable discreta llevan cinco columnas donde los elementos que participan son los siguientes:

a) *Frecuencia absoluta* : (f_i) indica el número de veces que se repite una variable.

b) *Tamaño de la muestra* : (n) indica la cantidad de elementos que conforman la muestra, se obtiene sumando todas las frecuencias absolutas.

$$n = \sum_{i=1}^{m} f_i \qquad\qquad m = \text{número de clases distintas}$$

c) *Frecuencia relativa* : (h_i) es la proporción de datos que se encuentra en una clase, se obtiene dividiendo la frecuencia absoluta de la clase por el tamaño de la muestra.

$$h_i = \frac{f_i}{n}$$

<u>**Obs:**</u> a) $\sum h_i = 1$

b) $0 \leq h_i \leq 1$

d) *Frecuencia absoluta acumulada* : (F_i) indica la cantidad de datos que se encuentran hasta cierta clase.

$$F_i = \sum_{j=1}^{i} f_j$$

e) *Frecuencia relativa acumulada* : (H_i) es la proporción de datos acumulados que se encuentran hasta cierta clase.

$$H_i = \sum_{j=1}^{i} h_j$$

Obs: a) $H_m = 1$

b) $0 \leq H_i \leq 1$

Ejercicio

Una empresa que tiene 50 trabajadores se propone reestructurar las remuneraciones, se estudia los años de servicio de los trabajadores determinándose los siguientes resultados:

4	5	4	6	7	9	7	7	5	8
8	7	6	7	7	4	6	8	8	9
6	8	9	5	6	5	4	7	9	6
7	6	5	4	4	4	6	8	8	7
8	9	5	5	4	6	7	9	5	4

$N = 50$ (tamaño de la población)

Se pide:

$1. -$ Tabular la información.

$2. -$ ¿ Qué cantidad de trabajadores tiene 8 años de servicio ?.

$3. -$ ¿ Qué porcentaje de trabajadores tiene 6 años de servicio ?.

$4. -$ Si aquellos trabajadores que tengan a lo menos siete años de servicio reciben un aumento del 8% .¿ Qué porcentaje de los trabajadores recibió dicho aumento?.

$5. -$ Si todos los trabajadores que tengan a lo más cinco años de servicio reciben una bonificación de $\$20.000$.¿ Qué cantidad de trabajadores recibió dicha bonificación?.

$6. -$ Si la empresa decide otorgar una bonificación especial de $\$13.200$ por cada año de servicio.¿ Cuánto será el dinero necesario para cumplir dicha bonificación?.

Solución

$1. -$

Años de servicio	f_i	h_i	F_i	H_i
4	9	$0,18$	9	$0,18$
5	8	$0,16$	17	$0,34$
6	9	$0,18$	26	$0,52$
7	10	$0,20$	36	$0,72$
8	8	$0,16$	44	$0,88$
9	6	$0,12$	50	$1,00$
Total	50	$1,00$		

$2. -$ Ocho trabajadores tienen 8 años de servicio

$3. -$ El 18% de los trabajadores tiene 6 años de servicio.

$4. -$ El 48% de los trabajadores recibió el aumento de sueldo.

$5. -$ 17 trabajadores recibieron la bonificación.

$6. -$ $\$\ 4.197.600$ se necesitan para la bonificación por año de servicio.

B.2) Tabulación de variable continua o discreta

Para tabular una variable continua o discreta (que tome un gran número de datos distintos) se necesitan los siguientes elementos:

a) ***Rango o recorrido*** **:** Es la diferencia entre el valor máximo y valor mínimo que toma la variable.

$R = x_{\text{máx}} - x_{\text{mín}}$

b) ***Número de intervalos o clases*** **(m) :** Es el número de grupos en que es posible dividir los valores de la variable.

El número de clases no debe ser ni muy grande ni muy pequeño, un número pequeño de clases puede ocultar la naturaleza general de los datos y un número muy grande puede ser demasiado detallado como para revelar alguna información útil. Como regla general se recomienda que el número de clases esté entre cinco y veinte. Hay una regla llamada ***Regla de Sturges*** que puede dar una aproximación razonable para el número de clases, ella es:

$m = 1 + 3,3\,log(n)$ donde n es el número de datos de la muestra.

c) ***Amplitud del intervalo o amplitud de la clase*** **(a) :**

$$a = \frac{\text{Recorrido}}{\text{N° de clases}} = \frac{R}{m}$$

d) ***Límites de un intervalo*** **:** Son los valores extremos de una clase. El menor valor es considerado como el **límite inferior** y el valor que se obtiene sumando al límite inferior la amplitud del intervalo es el **límite inferior** de la segunda clase.

e) ***Límites reales de un intervalo*** **:** Se obtienen calculando el promedio entre el límite superior de una clase y el límite inferior de la clase siguiente.

f) ***Marca de clase*** **:** (x_i) Es el punto medio de un intervalo.

g) ***Frecuencia absoluta*** **:** (f_i) indica el número de observaciones que pertenece a un intervalo dado.

Observación: $n = \sum_{i=1}^{m} f_i$ $n =$ tamaño de la muestra

h) *Frecuencia relativa* : (h_i) es la proporción de datos que se encuentra en un intervalo, se determina dividiendo la frecuencia absoluta del intervalo por el tamaño de la muestra.

$$h_i = \frac{f_i}{n}$$

i) *Frecuencia absoluta acumulada* : (F_i) indica el número de datos de la muestra menores o iguales al límite real superior del intervalo i.

$$F_i = \sum_{j=1}^{i} f_j$$ **Obs:** $F_m = n$

j) *Frecuencia relativa acumulada* : (H_i) indica la proporción de datos de la muestra menores o iguales al límite real superior del intervalo i.

$$H_i = \sum_{j=1}^{i} h_j$$

Observación: Existe más de un método para construir una tabla de distribución de frecuencias, a continuación se presentan dos formas de construirla:

Ejemplo

Los siguientes datos corresponden a las notas obtenidas por 100 alumnos en un curso de Estadística :

100	87	54	82	93	47	40	53	88	58
84	65	57	66	25	70	85	36	61	34
33	33	100	69	77	88	63	17	42	55
98	70	68	70	65	70	84	52	60	54
57	47	57	86	25	66	40	100	32	39
90	83	64	95	85	100	67	60	42	65
82	85	62	72	65	76	23	96	30	45
77	55	100	80	55	52	85	68	53	82
55	51	47	47	64	75	65	60	45	75
62	93	98	58	95	83	33	70	51	60

$1. -$ Construya la correspondiente distribución de frecuencia.

$2. -$ ¿En qué clase se concentra el mayor número de notas?

$3. -$ ¿Cuál es la frecuencia absoluta del cuarto intervalo?. Interprete el resultado .

$4. -$ ¿Qué porcentaje de los alumnos tienen una nota inferior a 57?

5. $-$ ¿Cuántos alumnos tienen una nota superior a 46?

6. $-$ Interprete la frecuencia acumulada del sexto intervalo.

7. $-$ Interprete la frecuencia relativa acumulada del quinto intervalo.

Solución:

$R = 100 - 17 = 83$

$n = 100$

$m = 1 + 3,3\,log(100) = 7,6 \simeq 8$

$a = \dfrac{83}{8} = 10,36 \simeq 10$

1. $-$

Notas	Límites reales	x_i	f_i	h_i	F_i	H_i
$17-26$	$16,5-26,5$	$21,5$	4	$0,04$	4	$0,04$
$27-36$	$26,5-36,5$	$31,5$	7	$0,07$	11	$0,11$
$37-46$	$36,5-46,5$	$41,5$	7	$0,07$	18	$0,18$
$47-56$	$46,5-56,5$	$51,5$	16	$0,16$	34	$0,34$
$57-66$	$56,5-66,5$	$61,5$	22	$0,22$	56	$0,56$
$67-76$	$66,5-76,5$	$71,5$	13	$0,13$	69	$0,69$
$77-86$	$76,5-86,5$	$81,5$	15	$0,15$	84	$0,84$
$87-96$	$86,5-96,5$	$91,5$	9	$0,09$	93	$0,93$
$97-106$	$96,5-106,5$	$101,5$	7	$0,07$	100	$1,00$
Total			100	$1,00$		

2. $-$ El mayor número de notas se concentra en el quinto intervalo, que coresponde al intervalo entre $57-66$.

3. $-$ La frecuencia absoluta del cuarto intervalo es 16. Esto nos indica que son 16 los alumnos que tienen una nota entre $47-56$.

4. $-$ El 34 % de los alumnos tiene una nota inferior a 57.

5. $-$ El 82 % de los alumnos tiene una nota superior a 46.

6. $-$ Existen 69 alumnos con nota inferior a 77.

7. $-$ El 56 % de los alumnos tiene una nota inferior a 67.

Ejercicios

1) Los siguientes datos corresponden al sueldo (en miles de pesos) de 40 trabajadores de una empresa :

119	135	138	144	146	150	156	164
125	135	140	144	147	150	157	165
126	135	140	145	147	152	158	168
128	136	142	145	148	153	161	173
132	138	142	146	149	154	163	176

a) Construya la tabla de frecuencia con todos sus elementos.
b) ¿En qué clase se encuentra el mayor número de trabajadores?.
c) ¿Qué porcentaje de trabajadores gana entre $ 139.000 y $ 168.000 ?.
d) ¿Cuántos trabajadores ganan a lo menos $ 159.000 ?.
e) ¿Cuántos trabajadores ganan a lo más $ 148.000 ?.

2) En una industria es necesario realizar un estudio respecto al peso de engranajes de gran tamaño. Los siguientes datos corresponden al peso, en kilógramos, de 30 de estas piezas, que poseen las mismas dimensiones, pero distinta aleación.

58	52	50	52	40	50	38	52	50	45
36	45	55	42	42	52	50	45	42	38
42	38	40	46	45	45	55	42	45	40

a) Construir una tabla de frecuencias de amplitud 5 comenzando desde 36.
b) ¿Cuántos engranajes pesan entre 46 y 55 Kg.?.
c) ¿Qué porcentaje representa a aquellos engranajes cuyo peso es inferior a 51 Kg.?.
d) ¿Cuál es la frecuencia relativa para aquel intervalo cuya marca de clase es 48 ?.
e) ¿Qué porcentaje representa a aquellas piezas que pesan más de 50 Kg. ? .

3) En una industria automotriz es necesario realizar un estudio debido a una partida defectuosa de discos de embrague. Para ello se ha recopilado la siguiente información referente a la duración en horas de 50 de ellos.

285	300	286	302	313	314	289	292	321	327
293	289	292	289	308	326	303	287	293	322
304	329	295	307	297	302	294	301	285	313
308	307	304	291	288	297	316	322	317	308
321	324	323	316	292	286	299	294	328	296

a) Construir una tabla de frecuencia de amplitud cinco comenzando desde 285.
b) ¿Cuántos discos duraron entre 290 y 299 horas?.
c) ¿Cuántos discos no alcanzaron a durar 300 horas?.
d) ¿Qué porcentaje representan los discos que duraron entre 310 y 314 horas?.
e) ¿Qué porcentaje representan los discos que duraron menos de 305 horas?.
f) ¿Cuántos discos duraron más de 309 horas?.
g) ¿Cuántos discos duraron menos de 305 horas?.
h) ¿Qué porcentaje representan los discos que duraron entre 285 y 294 horas?.
i) ¿Cuál es el intervalo de mayor frecuencia absoluta?.

4) En un conjunto habitacional se pretende hacer un estudio del número de personas que consumen productos enlatados. Los datos que han sido obtenidos de 50 bloques del conjunto habitacional son :

63	69	83	85	93	73	81	94	104	125
64	132	115	120	127	130	105	114	123	121
128	90	75	137	131	73	62	100	109	117
124	103	133	138	133	110	60	91	87	136
137	134	129	96	99	72	104	97	84	98

a) Construir una tabla de fecuencia de amplitud 10 partiendo desde 60.
b) ¿Cuántas personas consumen entre 100 y 129 productos enlatados ?.
c) ¿Qué porcentaje representa a las personas que consumen menos de 90 productos enlatados?.
d) ¿Qué cantidad de personas consumen más de 80 productos enlatados?.

5) Las ganancias por acción de 40 compañías de la industria de la construcción son:

$4,6$	$0,3$	$1,1$	$5,7$	$0,1$	$1,3$	$2,5$	$1,6$
$1,3$	$2,1$	$2,1$	$1,4$	$7,3$	$5,4$	$3,5$	$1,9$
$6,0$	$0,8$	$1,9$	$2,1$	$3,2$	$0,2$	$7,1$	$2,8$
$9,6$	$3,7$	$5,1$	$3,6$	$4,9$	$2,3$	$1,8$	$0,4$
$4,2$	$2,1$	$0,9$	$3,2$	$3,7$	$1,1$	$0,5$	$1,9$

a) Construya una distribución de frecuencias que comience en $0,1$ y tenga una amplitud de $2,0$
b) ¿Cuál es la frecuencia absoluta del tercer intervalo?. Interprete el resultado .
c) ¿Qué porcentaje de las compañías tienen a lo más una ganancia de $6,0$?
d) ¿Cuántas compañías tienen una ganancia a lo menos de $4,1$?
e) Interprete la frecuencia acumulada del segundo intervalo.
f) Interprete la frecuencia relativa acumulada del cuarto intervalo.

Solución

1) a) $R = 176 - 119 = 57$

$N = 40$

$m = 1 + 3,3log(40) = 6,28 \simeq 6$

$a = \dfrac{57}{6} = 9,5 \simeq 10$

Sueldo	Límites reales	x_i	f_i	h_i	F_i	H_i
$119 - 128$	$118,5 - 128,5$	$123,5$	4	$0,100$	$0,1$	4
$129 - 138$	$128,5 - 138,5$	$133,5$	7	$0,175$	$0,275$	11
$139 - 148$	$138,8 - 148,5$	$143,5$	13	$0,325$	$0,600$	24
$149 - 158$	$148,5 - 158,5$	$153,5$	9	$0,225$	$0,825$	33
$159 - 168$	$158,5 - 168,5$	$163,5$	5	$0,125$	$0,950$	38
$169 - 178$	$168,5 - 178,5$	$173,5$	2	$0,050$	1	40
Total			40			

b) En la tercera clase se encuentra el mayor número de trabajadores.
c) 67,5 % de los trabajadores gana entre $139.000 y $ 168.000
d) 7 trabajadores ganan a lo menos $ 159.000
e) 24 trabajadores ganan a lo más $ 148.000

2) a)

Peso	Límites reales	x_i	f_i	h_i	F_i	H_i
$36 - 40$	$35,5 - 40,5$	38	7	$0,23$	7	$0,23$
$41 - 45$	$40,5 - 45,5$	43	11	$0,37$	18	$0,60$
$46 - 50$	$45,5 - 50,5$	48	5	$0,17$	23	$0,77$
$51 - 55$	$50,5 - 55,5$	53	6	$0,20$	29	$0,97$
$56 - 60$	$55,5 - 60,5$	58	1	$0,30$	30	1
Total			30			

b) 11 engranajes pesan entre 46 y 55 kilos.
c) El 77 % de las piezas pesan menos de 51 kilos.
d) La frecuencia relativa es 0,17
e) El 23 % de las piezas pesa más de 50 kilos.

3) a)

Duración	Límites reales	x_i	f_i	h_i	F_i	H_i
$285-289$	$284,5-289,5$	287	9	$0,18$	9	$0,18$
$290-294$	$289,5-294,5$	292	8	$0,16$	17	$0,34$
$295-299$	$294,5-299,5$	297	5	$0,10$	22	$0,44$
$300-304$	$299,5-304,5$	302	7	$0,14$	29	$0,58$
$305-309$	$304,5-309,5$	307	5	$0,10$	34	$0,68$
$310-314$	$309,5-314,5$	312	3	$0,06$	37	$0,74$
$315-319$	$314,5-319,5$	317	3	$0,06$	40	$0,80$
$320-324$	$319,5-324,5$	322	6	$0,12$	46	$0,92$
$3215-329$	$324,5-329,5$	327	4	$0,08$	50	1
Total			50			

b) 13 discos duraron entre 290 y 299 horas.
c) 22 discos no alcanzaron a durar 300 horas.
d) El 6 % de los engranajes duraron entre 300 y 314 horas.
e) El 58 % de los engranajes duraron menos de 305 horas.
f) 16 engranajes duraron más de 309 horas.
g) 29 engranajes duraron menos de 305 horas.
h) El 16 % de los engranajes duraron entre 285 y 294 horas.
i) El primer intervalo.

4) a)

Nº de personas	f_i	h_i	F_i	H_i
$60-69$	5	$0,10$	5	$0,10$
$70-79$	4	$0,08$	9	$0,18$
$80-89$	5	$0,10$	14	$0,28$
$90-99$	8	$0,16$	22	$0,44$
$100-109$	6	$0,12$	28	$0,56$
$110-119$	4	$0,08$	32	$0,64$
$120-129$	8	$0,16$	40	$0,80$
$130-139$	10	$0,20$	50	1
Total	50			

b) 18 personas consumen entre 100 y 129 productos enlatados.
c) El 28 % de las personas consume menos de 90 productos enlatados.
d) 41 personas consume más de 79 productos enlatados.

5) a)

Ganancias	Límites Reales	x_i	f_i	h_i	F_i	H_i
$0,1-2,0$	$0,05-2,05$	$1,05$	17	$0,425$	17	$0,425$
$2,1-4,0$	$2,05-4,05$	$3,05$	13	$0,325$	30	$0,75$
$4,1-6,0$	$4,05-6,05$	$5,05$	7	$0,175$	37	$0,925$
$6,1-8,0$	$6,05-8,05$	$7,05$	2	$0,05$	39	$0,975$
$8,1-10,0$	$8,05-10,05$	$9,05$	1	$0,025$	40	$1,000$
Total			40	$1,000$		

b) La frecuencia absoluta del tercer intervalo es 7, es decir, existen 7 compañías cuyas ganancias están entre $4,1$ y $6,0$ por acción.
c) El $92,5$ % de las compañías tienen a lo más una ganancia de $6,0$ por acción.
d) 10 compañías tienen a lo menos una ganancia de $4,1$ por acción.
e) 30 compañías tienen una ganancia igual o menor a $4,0$ por acción.
f) El $97,5$ % de las compañías tienen una ganancia por acción de a lo más $8,0$.

Representación Gráfica

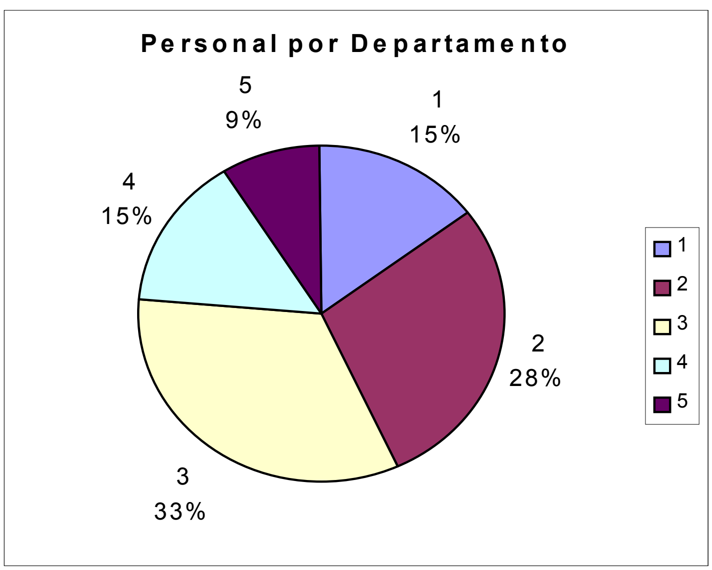
Su objetivo es captar la información obtenida en los datos en forma rápida por cualquier persona, así cada representación debe llevar un título adecuado.

Las normas en la construcción de un gráfico estadístico son similares a los de gráficos de funciones, las variables independientes, se ubican en las abscisas y las dependientes en las ordenadas.

Tipos de gráficos

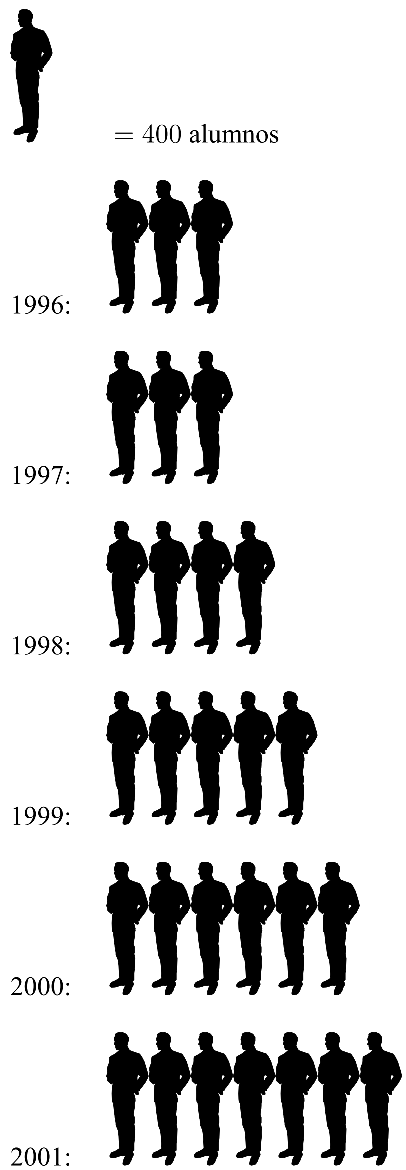
a) *Gráfico circular*: se usan para mostrar el comportamiento de las frecuencias relativas, absolutas o porcentuales de las variables. Dichas frecuencias son representadas por medio de sectores circulares, proporcionales a las frecuencias.

Departamento	f_i	%
A (1)	54	15
B (2)	101	28
C (3)	119	33
D (4)	54	15
E (5)	32	9
Total	360	100



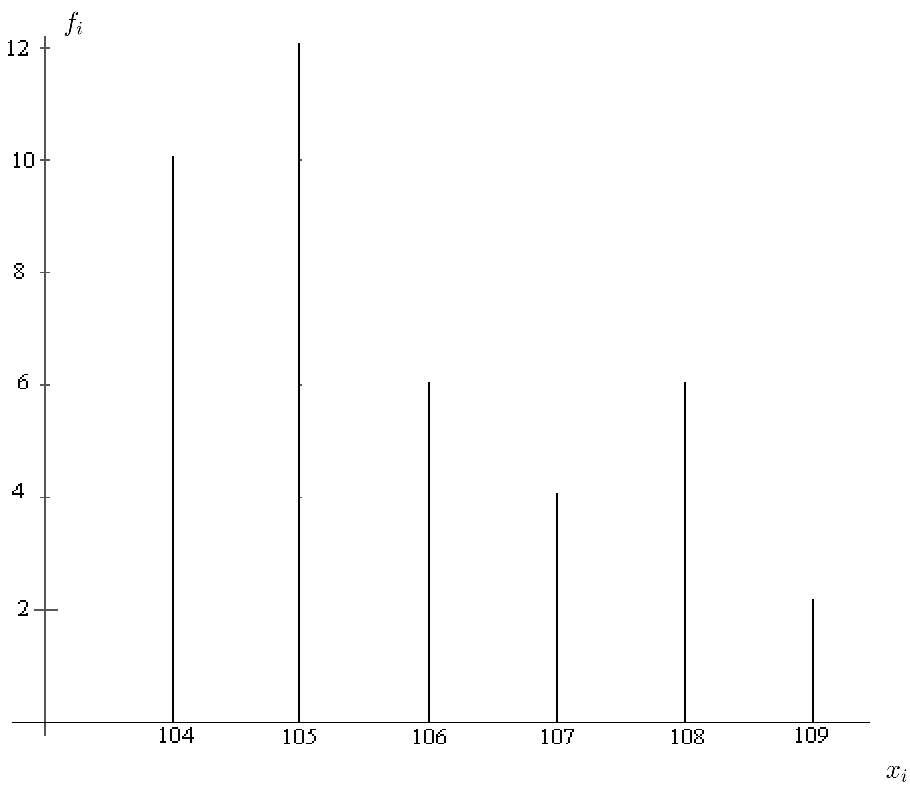
b) ***Pictograma*****:** es un gráfico cuyo uso es similar al de sector circular, pero la frecuencia es representada por medio de una figura o dibujo que identifique a la variable en estudio. Este gráfico se utiliza para mostrar producciones en una serie cronológica.

Por ejemplo, Alumnos del Instituto Profesional Dr. Virginio Gómez:



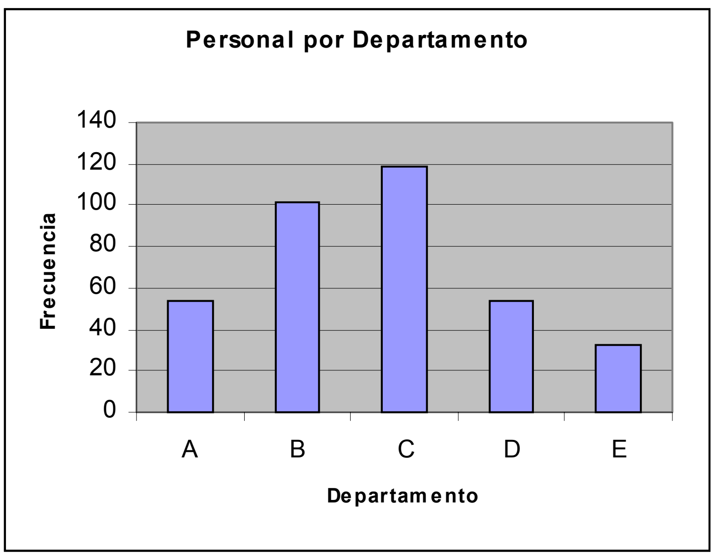
c) ***Gráfico lineal*****:** se utiliza para mostrar las frecuencias absolutas o relativas de una variable discreta, son representadas mediante líneas verticales proporcionales a dichas frecuencias.

x_i	f_i
104	10
105	12
106	6
107	4
108	6
109	2
Total	40



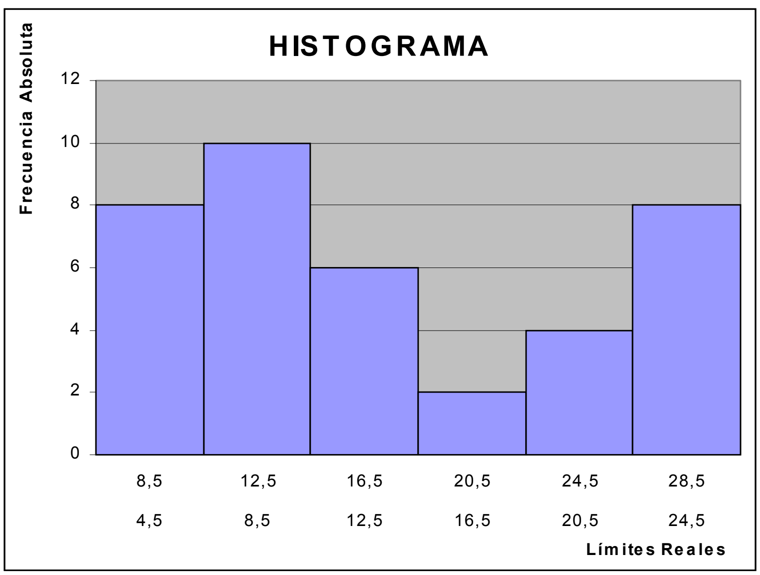
d) ***Gráfico de barra*****:** Se utiliza para representar tablas de frecuencia con atributos o con variables discretas y pocos valores. Sobre un eje horizontal se construyen bases de rectángulo del mismo ancho cada uno correspondiente a una modalidad del atributo, sobre estas bases se levantan rectángulos cuya altura es proporcional a la frecuencia absoluta de la modalidad. El espacio entre ellas debe ser uniforme.

Departamento	f_i
A	54
B	101
C	119
D	54
E	32
Total	360



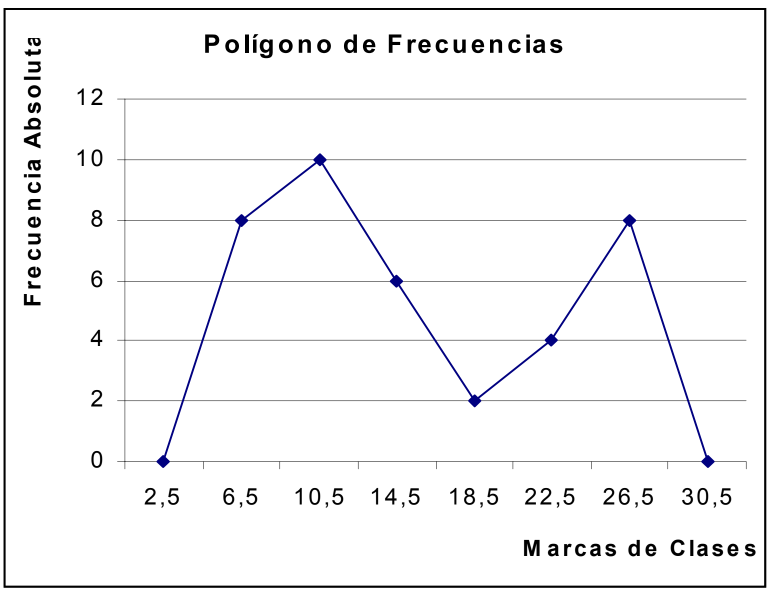
e) *Histograma*: es el gráfico adecuado cuando los datos están ordenados en tablas con intervalos, es decir, para datos de variables continuas. También el histograma es una conformación de rectángulos, pero uno al lado de otro cuya área es proporcional a la frecuencia de cada intervalo. Los extremos de la base de cada rectángulo son los límites reales del intervalo.

Límites Reales	f_i
$4,5-8,5$	8
$8,5-12,5$	10
$12,5-16,5$	6
$16,5-20,5$	2
$20,5-24,5$	4
$24,5-28,5$	8
Total	38



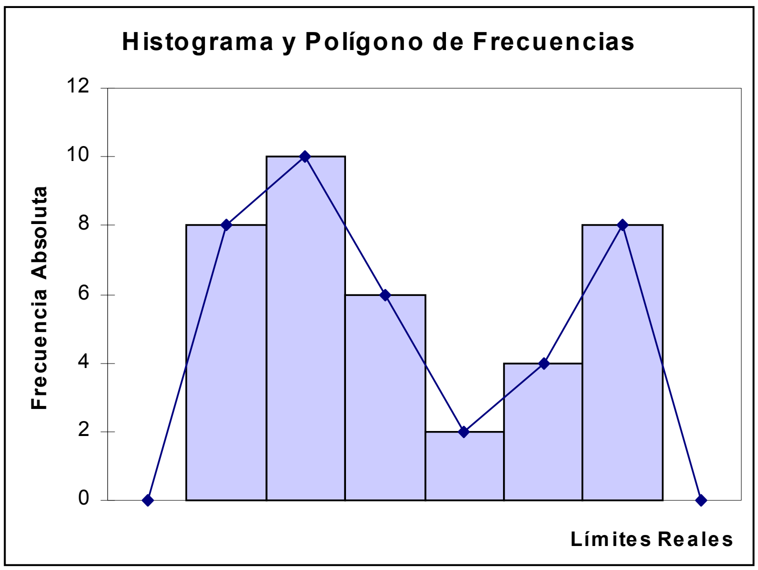
f) *Polígono de frecuencia*: este gráfico sirve para mostrar la tendencia de la variable, se puede determinar a partir de un histograma uniendo los puntos medios superiores de cada rectángulo del histograma. También, se determina el polígono uniendo los puntos formado por la marca de clase con la frecuencia absoluta del intervalo respectivo.

Límites reales	x_i	f_i
$4,5-8,5$	$6,5$	8
$8,5-12,5$	$10,5$	10
$12,5-16,5$	$14,5$	6
$16,5-20,5$	$18,5$	2
$20,5-24,5$	$22,5$	4
$24,5-28,5$	$26,5$	8
Total		38



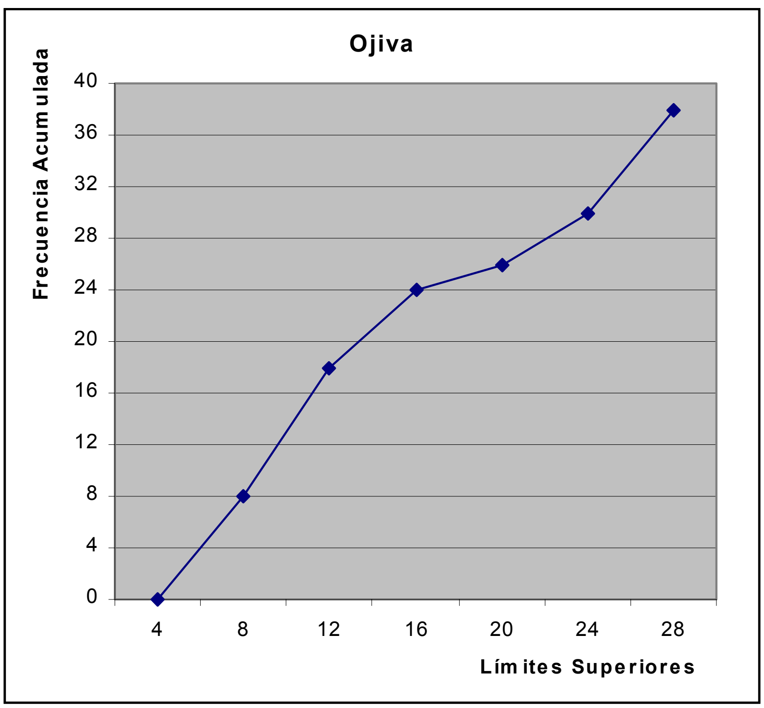
Observación: El polígono de frecuencias se convierte en polígono de frecuencias relativas, cambiando la frecuencia absoluta por la frecuencia relativa, en este caso, el área bajo el polígono de frecuencias relativas es igual a 1.

Histograma y Polígono de Frecuencias



e) ***Ojiva*:** es un gráfico que se usa para mostrar como se acumulan las frecuencias absolutas, relativas o porcentuales. Se obtiene al unir los puntos formados por los límites superiores de cada intervalo con la frecuencia absoluta o relativas acumuladas del intervalo respectivo. Si se consideran las frecuencias porcentuales acumuladas se llama ***ojiva porcentual.***

Límites reales	x_i	f_i	F_i
$4-8$	6	8	8
$8-12$	10	10	18
$12-16$	14	6	24
$16-20$	18	2	26
$20-24$	22	4	30
$24-28$	26	8	38
Total		38	



Ejercicios

1) Dada la información referente a la ubicación de personas dentro de cuatro departamentos de una empresa, se pide :

a) Tabular la información.
b) Realizar gráfico circular.
c) Indique frecuencias relativas porcentuales en cada grupo.

M	A	P	CC	A	CC	M	P	P	M
P	CC	M	A	M	CC	P	P	M	P
A	P	A	M	M	A	M	A	P	M
M	A	CC	A	A	M	P	M	M	P

donde : A = abastecimiento ; CC = control de calidad ; M = mantención ; P = producción.

2) Se realizó un número determinado de compras de materia prima. El volumen de la materia prima viene dado en m^3.Parte de la información se registra en la siguiente tabla :

Volumen	x_i	f_i	h_i	F_i	H_i	Límites reales
$6-10$		1				
$11-15$						
$16-20$		6		9		
$21-25$				18		
$26-30$				27		
Total		27				

a) Complete la tabla dada.
b) En un sólo gráfico, dibuje un histograma y un polígono de frecuencia.
c) ¿Cuántas compras se realizaron entre 11 y 30 m^3?.
d) ¿Cuántas compras se realizaron entre 16 y 25 m^3?.
e) ¿Qué porcentaje de compras se realizaron entre 16 y 20 m^3?.
f) ¿Cuántas compras se realizaron en total?.

3) Los siguientes datos corresponden a la duración, en horas, de 50 válvulas que fueron sometidas a un cierto control.

Tiempo	x_i	f_i	h_i	F_i	H_i	Límites reales
$450-499$		4				
$500-549$		5				
$550-599$		12				
$600-649$		10				
$650-699$		15				
$700-749$		3				
$750-799$		1				
Total		50				

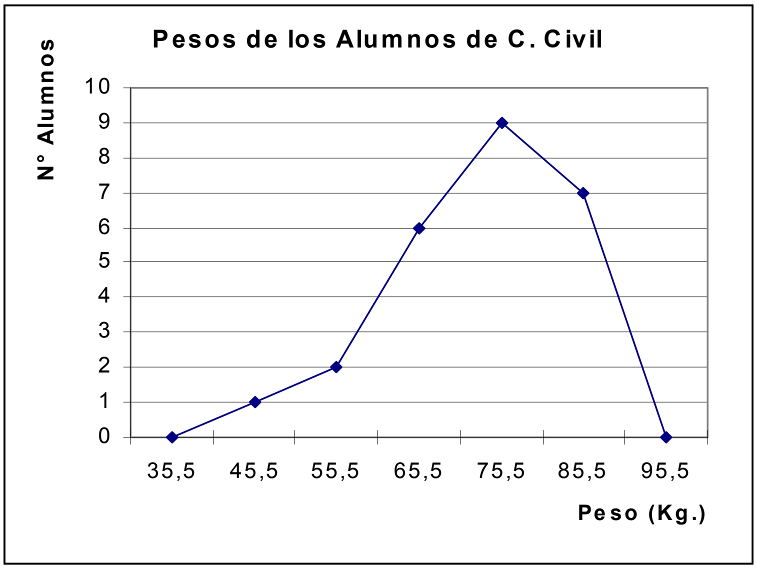
a) Complete la tabla dada.
b) Grafique la ojiva
c) ¿Qué porcentaje de las válvulas duraron, en promedio $674,5$ horas?.
d) ¿Qué porcentaje de las válvulas duraron entre 650 y 749 horas?.
e) ¿Cuántas válvulas duraron menos de 550 horas?.
f) ¿Qué porcentaje de las válvulas duraron más de 649 horas?

4) Se realizaron dos experimentos referente al peso, en Kg., aplicado sobre una cierta cantidad de tableros.

Peso (Kg.)	A	B
$15-19$	7	3
$20-24$	3	6
$25-29$	2	8
$30-34$	11	8
$35-39$	10	12
$40-44$	7	3
Total	40	40

a) Grafique el histograma del experimento A.
b) Grafique la ojiva porcentual del experimento B.
c) Realice, en un mismo gráfico, los polígonos de frecuencia.
d) Realice, en un mismo gráfico, las ojivas.

5) Dado el siguiente Polígono de Frecuencias:



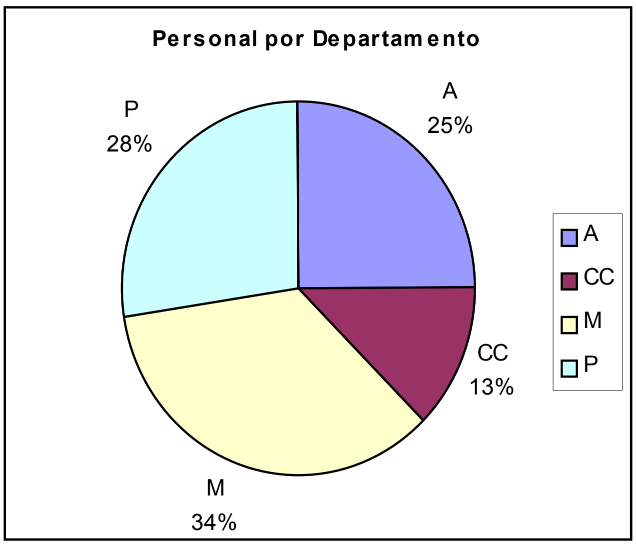
a)¿Cuáles son los límites reales del cuarto intervalo?.
b) Interprete la frecuencia del cuarto intervalo.
c) Interprete el porcentaje de datos que hay en el quinto intervalo.
d) ¿Qué porcentaje de pesos es igual o menor que $60,5$ Kg.?.
e) ¿Cuántos pesos son iguales o mayores que $50,5$ Kg.?.

Solución

1) a)

Departamento	f_i
A	10
CC	5
M	14
P	11
Total	40

b) Gráfico Circular



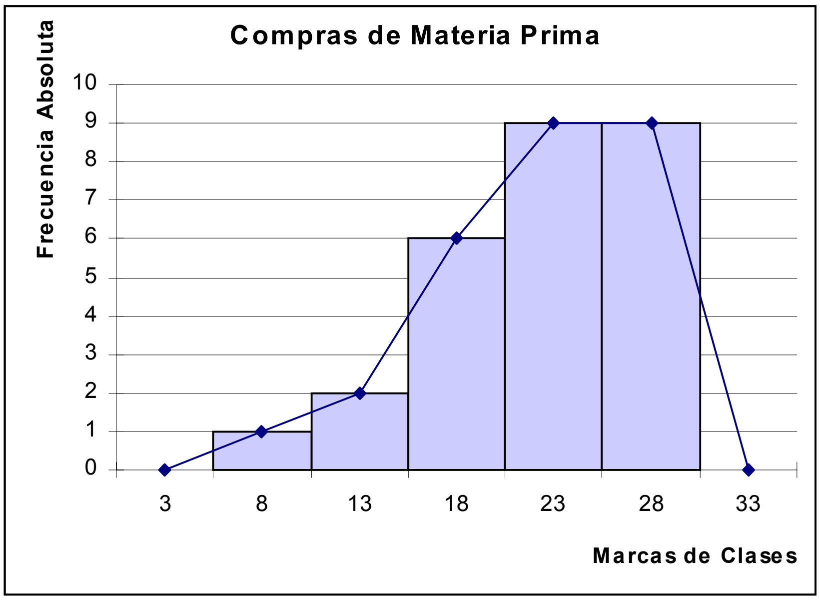
c)

Departamento	f_i	h_i	%
A	10	0.25	25
CC	5	0.125	13
M	14	0.35	35
P	11	0.275	28
Total	40	1	100

2) a)

Volumen	x_i	f_i	h_i	F_i	H_i	Límites reales
$6 - 10$	8	1	0.037	1	0.037	$5.5 - 10.5$
$11 - 15$	13	2	0.074	3	0.111	$10.5 - 15.5$
$16 - 20$	18	6	0.222	9	0.333	$15.5 - 20.5$
$21 - 25$	23	9	0.333	18	0.666	$20.5 - 25.5$
$26 - 30$	28	9	0.333	27	0.999	$25.5 - 30.5$
Total		27	0.999			

b) Histograma y Polígono de Frecuencia



c) Entre 11 y 30 m^3 se realizaron 26 compras

d) Entre 16 y 25 m^3 se realizaron 15 compras

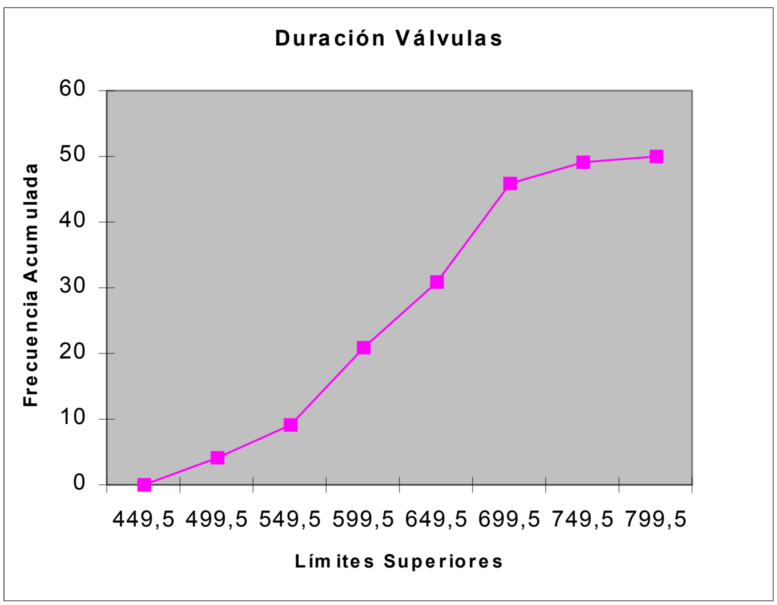
e) Entre 16 y 20 m^3 se realizaron un porcentaje de 22,2 % de compras

f) En total se realizaron 27 compras

3) a)

Tiempo	x_i	f_i	h_i	F_i	H_i	Límites reales
$450-499$	$474,5$	4	$0,08$	4	$0,08$	$449,5-499,5$
$500-549$	$524,5$	5	$0,10$	9	$0,18$	$499,5-549,5$
$550-599$	$574,5$	12	$0,24$	21	$0,42$	$549,5-599,5$
$600-649$	$624,5$	10	$0,20$	31	$0,62$	$599,5-649,5$
$650-699$	$674,5$	15	$0,30$	46	$0,92$	$649,5-699,5$
$700-749$	$724,5$	3	$0,06$	49	$0,98$	$699,5-749,5$
$750-799$	$774,5$	1	$0,02$	50	$1,00$	$749,5-799,5$
Total		50	$1,00$			

b) Ojiva



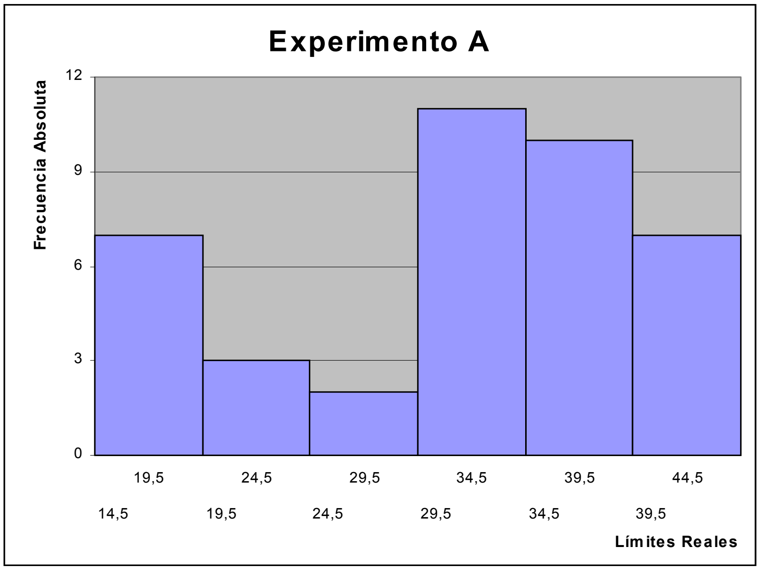
c) 30 % de las válvulas duraron en promedio 674,5 horas

d) 36 % de las válvulas duraron entre 650 y 749 horas

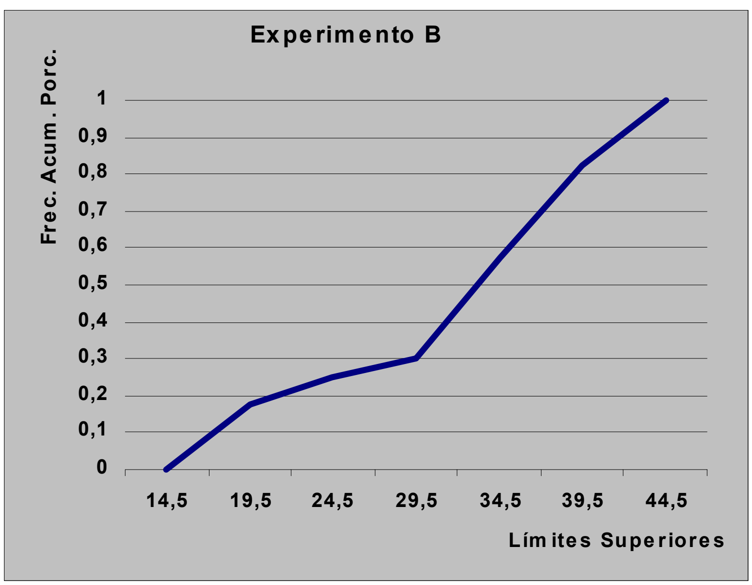
e) 9 válvulas duraron menos de 550 horas

f) 38 % de las válvulas duraron más de 649 horas

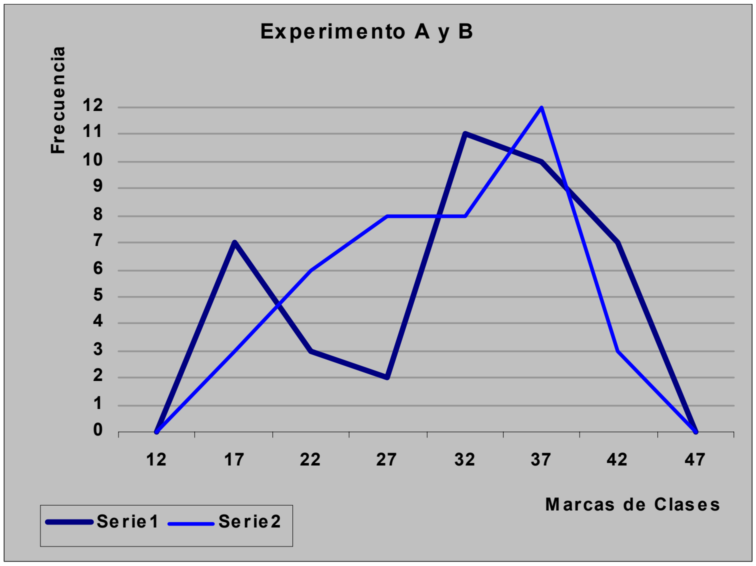
4) a) Histograma



b) Ojiva Porcentual

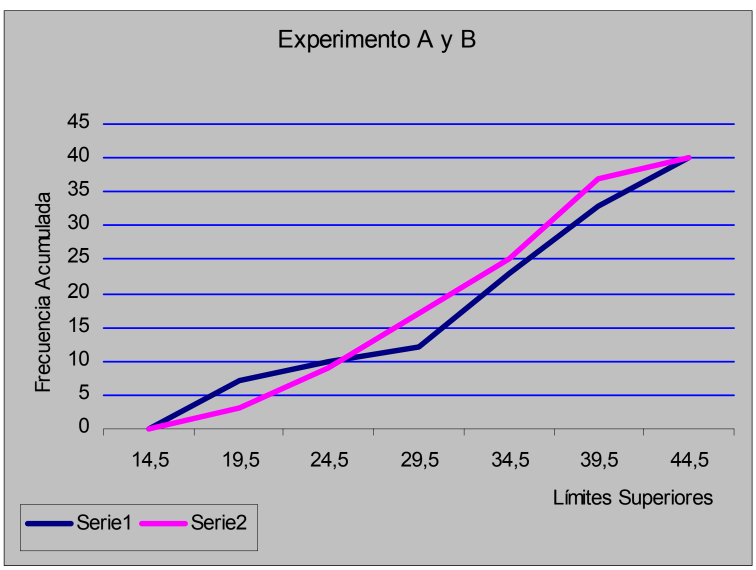


c) Polígonos de Frecuencia



Serie 1 = Experimento A Serie 2 = Experimento B

d) Ojivas



Serie 1 = Experimento A Serie 2 = Experimento B

5) a) Los límites reales del cuarto intervalo son $70,5 - 80,5$

b) 9 alumnos de C. Civil tienen pesos que van desde 71 kilos hasta 80 kilos

c) 28 % de los alumnos pesan más de $80,5$ kilos y menos de $90,5$ kilos

d) El 12 % de los pesos de los alumnos es igual o menor que $60,5$ kilos

e) 24 alumnos pesan a lo menos $50,5\,$Kg.

Medidas de tendencia central y de dispersión

En todo análisis y/o interpretación se pueden utilizar diversas medidas descriptivas que representan las propiedades de tendencia central, dispersión y forma para extraer y resumir las principales características de los datos. Si se calculan a partir de una muestra de datos, se les denomina **estadísticos**; si se les calcula a partir de una población se les denomina **parámetros**.

Medidas de tendencia central

La mayor parte de los conjuntos de datos muestran una tendencia a agruparse alrededor de un punto "central" y por lo general es posible elegir algún valor que describa todo un conjunto de datos. Un valor típico descriptivo como ese es una medida de tendencia central o "posición". Las medidas de tendencia central a estudiar son: media aritmética, mediana y moda.

Media aritmética

La ***media aritmética*** (también denominada media) es la medida de tendencia central que se utiliza con mayor frecuencia. Se calcula sumando todas las observaciones de un conjunto de datos, dividiendo después ese total entre el número total de elementos involucrados.

La media aritmética de un conjunto de valores $x_1, x_2, ..., x_n$ se define como el cuociente entre la suma de los valores y el número de ellos. Su símbolo es $\overline{x}$ si la media aritmética es de una muestra y μ si la media aritmética es de una población.

a) Para datos no agrupados:

Media muestral: $\overline{x} = \dfrac{x_1 + x_2 + ... + x_n}{n} = \displaystyle\sum_{i=1}^{n} \frac{x_i}{n}$; $n =$ tamaño de la muestra

Media poblacional: $\mu = \dfrac{x_1 + x_2 + ... + x_N}{N} = \displaystyle\sum_{i=1}^{N} \frac{x_i}{N}$; $N =$ tamaño de la población

Ejemplo : Calcular la media aritmética de los siguientes datos relacionados con las notas de test en Estadística obtenidas por un cierto alumno:

$45, 80, 56, 35, 25, 90$

$$\overline{x} = \frac{45 + 80 + 56 + 35 + 25 + 90}{6} = 55,17 \approx 55$$

El promedio de test es 55 puntos.

b) Para datos agrupados:

Si los datos están ordenados en tablas de frecuencia la media aritmética se obtiene como sigue :

Muestra **Población**

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + ... + x_m f_m}{f_1 + f_2 + ... + f_m} = \sum_{i=1}^{m} \frac{x_i f_i}{n} \qquad \mu = \sum_{i=1}^{m} \frac{x_i f_i}{N}$$

donde: x_i es la marca de clase del intervalo i-ésimo
f_i es la frecuencia del intervalo i-ésimo
n es el número de datos de la muestra y N es el número de datos de la población
m es el número de intervalos

Ejemplo : Calcular la media aritmética para el peso de 40 trabajadores, según tabla adjunta:

Peso (Kg.)	x_i	f_i	$x_i f_i$
$55 - 62$	$58,5$	5	$292,5$
$63 - 70$	$66,5$	15	$997,5$
$71 - 78$	$74,5$	12	894
$79 - 86$	$82,5$	5	$412,5$
$87 - 94$	$90,5$	3	$271,5$
Total		40	2868

$$\overline{x} = \sum_{i=1}^{5} \frac{x_i f_i}{n} = \frac{2868}{40} = 71,7 \approx 72$$

El peso promedio de los 40 trabajadores es de 72 kilos

Propiedades de la media aritmética

Propiedad 1 : La media aritmética de una constante es igual a la constante.

$\overline{x}$: x_1 x_2 x_3 ... x_n

valores : a a a ... a

$$\overline{x} = \frac{a + a + a + ... + a}{n} = \frac{na}{n} = a$$

Por lo tanto, $\overline{x} = a$

Propiedad 2 : La media aritmética de una variable más una constante es igual a la media aritmética de la variable más la constante.

$$\overline{x} \quad : \quad x_1 \qquad x_2 \qquad x_3 \quad \ldots \quad x_n$$

$$\overline{y} \quad : \quad x_1 + c \qquad x_2 + c \qquad x_3 + c \quad \ldots \quad x_n + c$$

$$\overline{y} = \sum_{i=1}^{n} \frac{yi}{n}$$

$$\overline{y} = \frac{\sum_{i=1}^{n} (x_i + c)}{n} = \frac{(x_1 + c) + (x_2 + c) + ... + (x_n + c)}{n}$$

$$= \frac{x_1 + x_2 + x_3 + ... + x_n + nc}{n}$$

$$= \sum_{i=1}^{n} \frac{x_i}{n} + \frac{nc}{n}$$

$$= \overline{x} + c$$

Propiedad 3 : La media aritmética de una variable por una constante es igual al producto de la constante por la media de la variable.

$$\overline{x} \; : \; x_1 \qquad x_2 \qquad \ldots \qquad x_n$$

$$\overline{z} \; : \; x_1c \; x_2c \qquad \ldots \qquad x_nc$$

$$\overline{z} = \frac{x_1c + x_2c + ... + x_nc}{n}$$

$$= \frac{c(x_1 + x_2 + ... + x_n)}{n}$$

$$= c\overline{x}$$

Propiedad 4 : Media Ponderada

$$\overline{x} = \frac{\overline{x}_1 \cdot n_1 + \overline{x}_2 \cdot n_2 + \ldots + \overline{x}_p \cdot n_p}{n_1 + n_2 + \ldots + n_p}$$

Ventajas y desventajas del uso de la media aritmética:

Ventajas	Desventajas
- Estable muestra a muestra	- No aplicable a atributos
- Fácil cálculo e interpretación	- Influyen en su valor los valores extremos

Ejemplos:

1) De un grupo de contribuyentes se determinó que el promedio de impuestos es de $32.200. Determinar en cada uno de los siguientes casos, la nueva media aritmética:

a) Los impuestos aumentan en un 2 %

b) A los impuestos se les disminuye la cantidad de $2.300

c) A cada contribuyente, se le disminuye un 3 % y además se le condona $2.550

Solución:

1) a) $\overline{x} = 32.200 \cdot 1,02 = 32.844$ La nueva media aritmética es $\$\,32.844$

b) $\overline{x} = 32.200 - 2.300 = 29.900$ La nueva media aritmética es $\$\,29.900$

c) $\overline{x} = 32.200 \cdot 0,97 - 2.550 = 28.684$ La nueva media aritmética es $\$\,28.684$

2) En tres cursos de un mismo nivel los promedios de las calificaciones fueron $5,6$; $6,1$ y $4,9$; si los cursos tenían respectivamente 34 ; 30 y 36 alumnos, determine la calificación promedio de los tres cursos.

Solución:

$$\overline{x} = \frac{5,6 \cdot 34 + 6,1 \cdot 30 + 4,9 \cdot 36}{34 + 30 + 36} = \frac{549,8}{100} = 5,498 \approx 5,5$$

El promedio de las calificaciones de los tres cursos es $5,5$

Mediana

La mediana es el valor que se encuentra en el centro de una secuencia ordenada de datos. La mediana no se ve afectada por observaciones extremas en un conjunto de datos. Por ello, cuando se presenta alguna información extrema, resulta apropiado utilizar la mediana, y no la media, para describir el conjunto de datos.

Su símbolo es Me.

a) Mediana para datos no agrupados

Se deben ordenar los datos de forma creciente o decreciente. Para muestras con un número par de observaciones, la mediana es el dato que queda en el centro de dicha ordenación y para muestras con número impar de observaciones la mediana es el promedio de los dos datos centrales.

Ejemplos :

1) Para muestra con número impar de datos: $Me = X_{\frac{n+1}{2}}$

datos : $4,\ 7,\ 5,\ 6,\ 3,\ 2,\ 7$

datos ordenados : $2,\ 3,\ 4,\ \underline{\mathbf{5}},\ 6,\ 7,\ 7$ $\quad \Rightarrow Me = X_{\frac{7+1}{2}} = X_4 = 5$

2) Para muestra con número par de datos: $Me = \dfrac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

datos : $12,\ 15,\ 14,\ 16,\ 11,\ 10,\ 10,\ 13$

datos ordenados : $16,\ 15,\ 14,\ \underline{\mathbf{13,\ 12}},\ 11,\ 10,\ 10$

$$Me = \frac{X_{\frac{8}{2}} + X_{\frac{8}{2}+1}}{2} = \frac{X_4 + X_5}{2} = \frac{13+12}{2} = 12,5$$

b) Mediana para datos agrupados

$$Me = L_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right) \cdot a$$

donde: i es el primer intervalo cuya frecuencia acumulada supera a $\dfrac{n}{2}$

L_i es el límite real inferior del intervalo de la mediana.
n es el número de datos.
F_{i-1} es la frecuencia acumulada anterior al intervalo de la mediana.
f_i es la frecuencia absoluta del intervalo de la mediana.
a es la amplitud del intervalo.

Ejemplo : Distribución de frecuencias de la duración, en horas, de uso continuo de 212 dispositivos electrónicos iguales, sometidos a un cierto control.

Duración	f_i	F_i
$350 - 399$	4	4
$400 - 449$	6	10
$450 - 499$	9	19
$500 - 549$	20	39
$550 - 599$	31	$\mathbf{70}$
$\mathbf{600 - 649}$	$\mathbf{80}$	150
$650 - 699$	42	192
$700 - 749$	10	202
$750 - 799$	8	210
$800 - 849$	2	212
Total	212	

El intervalo donde se encuentra la Mediana es el primer intervalo en el cual:

$$\frac{n}{2} \leq F_i$$

En este caso, $\frac{n}{2} = \frac{212}{2} = 106 \leq F_i$ $\Rightarrow 106 \leq 150$ $\Rightarrow 6^{to}$ intervalo : $i = 6$

$a = 50$ $Me = 599,5 + \left(\frac{106 - 70}{80}\right) \cdot 50$

$F_5 = 70$ $Me = 622$ horas

$f_6 = 80$

$L_6 = 599,5$

Moda

La moda es el valor de un conjunto de datos que aparece con mayor frecuencia. Se le obtiene fácilmente a partir de un arreglo ordenado. A diferencia de la media aritmética, la moda no se afecta ante la ocurrencia de valores extremos. Sin embargo, sólo se utiliza la moda para propósitos descriptivos porque es más variable, para distintas muestras, que las demás medidas de tendencia central. Un conjunto de datos puede tener más de una moda o ninguna.

Su símbolo es Mo.

a) Moda para datos no agrupados

Ejemplos

1) datos : $2,\ 4,\ 5,\ 6,\ 7,\ 7,\ 8,\ 7,\ 6 \quad \Rightarrow Mo = 7$

2) datos : $1,\ 1,\ 3,\ 1,\ 1,\ 2,\ 2,\ 4,\ 2,\ 3,\ 2,\ 5,\ 6 \Rightarrow Mo = 1$ y 2

3) datos : $0,\ 0,\ 2,\ 3,\ 4,\ 5 \quad \Rightarrow Mo = 0$

4) datos : $0,\ 1,\ 2,\ 3,\ 4,\ 5 \quad \Rightarrow Mo =$ no existe

b) Moda para datos agrupados

Existe más de una forma de calcular la moda:

Caso a) $Mo = L_i + \left(\dfrac{d_1}{d_1 + d_2} \right) \cdot a$

donde : i es el intervalo de mayor frecuencia absoluta.
L_i es el límite real inferior del intervalo que contiene a la moda.
d_1 es la diferencia entre la frecuencia absoluta del intervalo de la moda y el intervalo anterior : $d_1 = f_i - f_{i-1}$
d_2 es la diferencia entre la frecuencia absoluta del intervalo de la moda y el intervalo posterior : $d_2 = f_i - f_{i+1}$
a es la amplitud del intervalo.

Caso b) $Mo = L_i + \left(\dfrac{f_{i+1}}{f_{i-1} + f_{i+1}} \right) \cdot a$

donde : i es el intervalo de mayor frecuencia absoluta.

Ejemplo : Sea la tabla:

Duración	f_i	F_i
$350 - 399$	4	4
$400 - 449$	6	10
$450 - 499$	9	19
$\mathbf{500 - 549}$	**80**	**99**
$550 - 599$	31	130
$600 - 649$	20	150
$650 - 699$	42	192
$700 - 749$	10	202
$750 - 799$	8	210
$800 - 849$	2	212
Total	212	

Caso a): En este caso, el intervalo de mayor frecuencia absoluta es el $4^{to} \Rightarrow i = 4$

$f_4 = 80$ $\qquad Mo = 499,5 + \left(\frac{71}{71+49}\right) \cdot 50$

$d_1 = 80 - 9 = 71$ $\qquad Mo = 529,08$ horas

$d_2 = 80 - 31 = 49$

$L_4 = 499,5$

$a = 50$

Caso b): $i = 4$

$f_{4+1} = f_5 = 31$

$f_{4-1} = f_3 = 9$

$L_4 = 499,5$

$a = 50$

$$Mo = 499,5 + \left(\frac{31}{9+31}\right) \cdot 50$$

$Mo = 538,25$ horas

Ejercicios

1) En una industria dos operarios en siete días de trabajo, son capaces de producir, por día, y en forma individual la siguiente cantidad de árboles para fresa de 250 mm de longitud por 300 mm de diámetro.

Operario A	105	106	104	102	103	100	101
Operario B	103	102	107	101	105	102	103

Determine :

a) Producción media de cada operario.
b) Moda del operario A.
c) Mediana del operario B.

2) Se hace una encuesta entre 100 personas acerca del número de horas diarias que se dedican a ver televisión, obteniéndose la siguiente información :

N° de horas	f_i
$0-1$	30
$2-3$	20
$4-5$	15
$6-7$	32
$8-9$	1
$10-11$	2
Total	100

Calcular la media, la mediana y la moda (caso a y b).

3) De un total de 100 datos, 20 son 4, 40 son 5, 30 son 6 y el resto 7. Hallar la media y la moda.

4) Cuatro grupos de estudiantes, consistentes en $15, 20, 10$ y 18 individuos, dieron pesos de $60, 72, 55$ y 65 kilos. Hallar el peso medio de los estudiantes.

5) Las notas de un estudiante en sus certámenes han sido $84, 91, 72, 68, 87$ y 78. Hallar la media, la mediana y la moda.

6) La siguiente tabla corresponde a la estatura de 80 estudiantes de una determinada carrera.

Estatura	f_i
$1,65-1,69$	6
$1,70-1,74$	12
$1,75-1,79$	30
$1,80-1,84$	22
$1,85-1,89$	8
$1,90-1,94$	2
Total	80

Hallar la media, mediana y moda (caso a y b) de la estatura.

7) La oficina de Censo, proporcionó las edades de hombres y mujeres divorciados (en miles de personas de 15 años de edad o más).

Edad	Hombre	Mujer
$15-19$	2	2
$20-24$	80	210
$25-29$	174	303
$30-34$	210	315
$35-39$	385	656
$40-44$	450	656
$45-49$	295	409
$50-54$	174	200
Total	1770	2751

Obtener las medidas de tendencia central.

Solución

1)

$a)\, \overline{x}_{\text{A}} = 103$ $\qquad$ $\overline{x}_{\text{B}} = 103,3$

$b)$ No hay moda, todos los datos tienen frecuencia uno.

$c)\, Me_{\text{B}} = 103$

2) $\overline{x} = 3,7$ $\qquad$ $Me = 3,5$ $\qquad$ (Caso a) $Mo = 6,21$
(Caso b) $Mo = 5,625$

3) $\overline{x} = 5,3$ $\qquad$ $Mo = 5$

4) El peso promedio de los estudiantes es $64,44$ kilos.

5) $\overline{x} = 80$ $\qquad$ $Me = 81$ $\qquad$ Mo no existe

6) $\overline{x} = 1,78$ $\qquad$ $Me = 1,78$ $\qquad$ $Mo = 1,78$

7)

	Hombre	Mujer
$\overline{x}$	$39,14$	$37,94$
Me	$39,88$	$38,66$
Mo (caso a)	$40,98$	$39,5$
Mo (caso b)	$41,67$	$37,88$ y $41,42$

Medidas de dispersión

Una segunda propiedad que describe a un conjunto de datos es la dispersión. Dispersión es el grado de variación o diseminación de los datos. Dos conjuntos de datos pueden diferir tanto en tendencia central como en dispersión o dos conjuntos de datos pueden tener las mismas medidas de tendencia central, pero diferir mucho en términos de dispersión.

Ejemplo: 1) $2,\ 2,\ 2,\ 2,\ 2 \quad \overline{x} = 2$

2) $1,\ 1,\ 2,\ 3,\ 3 \quad \overline{x} = 2$

Los estadígrafos de dispersión nos indican si la distribución o conjunto de datos forma grupos homogéneos o heterogéneos. Las medidas de dispersión a estudiar son: rango, desviación media, varianza y desviación estándar.

Rango

Indica el número de valores que toma la variable. El rango es la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos.

$R = x_{\text{máx}} - x_{\text{mín}}$

Si los datos están agrupados en una tabla de frecuencias, el recorrido es la diferencia entre el límite real superior del último intervalo y el límite real inferior del primer intervalo.

$R = L_{\text{máx}} - L_{\text{mín}}$

Ejemplo:

1) Sea el siguiente conjunto de datos :

$12 \quad 15 \quad 17 \quad 23 \quad 25 \quad 28$

$x_{\text{máx}} = 28 \qquad x_{\text{mín}} = 12 \qquad R = 28 - 12 = 16$

2) Sea la siguiente tabla:

Peso (*Kg.*)	f_i
$55,0 - 63,0$	5
$63,1 - 71,1$	15
$71,2 - 79,2$	12
$79,3 - 87,3$	5
$87,4 - 95,4$	3
Total	40

$L_{min} = 54,95 \qquad L_{max} = 95,45$

$R = 95,45 - 54,95 \qquad R = 40,5\,Kg.$

El rango mide "la dispersión total" del conjunto de datos. Aunque el rango es una medida de dispersión simple y que se calcula con facilidad, su debilidad preponderante es que no toma en consideración la forma en que se distribuyen los datos entre los valores más pequeños y los más grandes.

Desviación Media

Es la media aritmética de los valores absolutos de las desviaciones de todos los datos respecto a la media aritmética. Su símbolo es DM.

a) Desviación media para datos no agrupados

$$DM = \frac{\sum\limits_{i=1}^{n} |\, x_i - \overline{x} \,|}{n}$$

<u>Ejemplo</u> : Obtener la desviación media para los datos $5,\ 7,\ 8,\ 10,\ 16$

$$\overline{x} = \frac{5+7+8+10+16}{5} = 9,2$$

$$DM = \frac{|\,5-9,2\,| + |\,7-9,2\,| + |\,8-9,2\,| + |\,10-9,2\,| + |\,16-9,2\,|}{5}$$

$$DM = \frac{15,2}{5}$$

$$DM = 3,04$$

b) Desviación media para datos agrupados

$$DM = \frac{\sum\limits_{i=1}^{m} |\, x_i - \overline{x} \,|\, f_i}{n}$$ donde x_i es la marca de clase

Ejemplo : Determine la desviación media de los siguientes datos agrupados :

Pesos (*Kg.*)	f_i
$60-62$	5
$63-65$	18
$66-68$	42
$69-71$	27
$72-74$	8
Total	100

Pesos (*Kg.*)	x_i	f_i	$x_i \cdot f_i$	$x_i - \overline{x}$	$\lvert x_i - \overline{x} \rvert f_i$
$60 - 62$	61	5	305	$6,45$	$32,25$
$63 - 65$	64	18	1152	$3,45$	$62,10$
$66 - 68$	67	42	2814	$0,45$	$18,90$
$69 - 71$	70	27	1890	$2,55$	$68,85$
$72 - 74$	73	8	584	$5,55$	$44,40$
Total		100	6745		$226,5$

$$\overline{x} = \frac{6745}{100} = 67,45$$

$$DM = \frac{226,5}{100} = 2,265$$

Varianza y Desviación Estándar

Dos medidas de dispersión que se utilizan con frecuencia y que sí toman en consideración la forma en que se distribuyen los valores son la varianza y su raíz cuadrada, la desviación estándar. Estas medidas establecen la forma en que los valores fluctúan con respecto a la media.

Varianza

La varianza se define como el promedio aritmético de las diferencias entre cada uno de los valores del conjunto de datos y la media aritmética del conjunto elevadas al cuadrado.

Su símbolo es S^2 si estamos trabajando con una muestra y σ^2 si estamos trabajando con una población.

a) Varianza para datos no agrupados

$$S^2 = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$ donde x_i representa los datos de la muestra.

$$\sigma^2 = \frac{\sum\limits_{i=1}^{N} (x_i - \mu)^2}{N-1}$$ donde x_i representa los datos de la población.

Ejemplo : Determine la varianza del siguiente conjunto de datos:

25 12 23 28 17 15

$$\overline{x} = \frac{25+12+23+28+17+15}{6} = 20$$

$$S^2 = \frac{(25-20)^2+(12-20)^2+(23-20)^2+(28-20)^2+(17-20)^2+(15-20)^2}{6-1}$$

$$S^2 = \frac{196}{5} \qquad \Rightarrow \qquad S^2 = 39,2 \text{ (en unidades al cuadrado)}$$

b) Varianza para datos agrupados

Muestra

$$S^2 = \frac{\sum_{i=1}^{m} (x_i - \overline{x})^2 f_i}{n-1}$$

Población

$$\sigma^2 = \frac{\sum_{i=1}^{m} (x_i - \mu)^2 f_i}{N-1}$$

donde x_i es la marca de clase.

Ejemplo : Considere la tabla con los datos de los edades de 26 personas

Edades (años)	f_i
$15-20$	2
$21-26$	7
$27-32$	8
$33-38$	5
$39-44$	4
Total	26

Edades (años)	x_i	f_i	$x_i \cdot f_i$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 f_i$
$15-20$	$17,5$	2	$35,0$	$155,2516$	$310,5032$
$21-26$	$23,5$	7	$164,5$	$41,7316$	$292,1212$
$27-32$	$29,5$	8	$236,0$	$0,2116$	$1,6928$
$33-38$	$35,5$	5	$177,5$	$30,6916$	$153,458$
$39-44$	$41,5$	4	$166,0$	$133,1716$	$532,6864$
Total		26	$779,0$		$1290,4616$

$$\overline{x} = \frac{779,0}{26} = 29,96 \text{ años}$$

$$S^2 = \frac{1290,4616}{25} = 51,618 \text{ (en años}^2\text{)}$$

Las fórmulas anteriores para calcular la Varianza muestral tienen una forma abreviada:

Para datos no agrupados

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2}{n-1}$$

donde: x_i representa los datos

Para datos agrupados

$$S^2 = \frac{\sum_{i=1}^{m} (f_i x_i^2) - n(\overline{x})^2}{n-1}$$

donde: x_i representa la marca de clase

Propiedades de la Varianza

1) $Var(x) = S_x^2 \geq 0$

2) $Var(x) = 0$ si $x =$ constante

3) $Var(ax) = a^2 Var(x)$

4) $Var(x+b) = Var(x)$

5) $Var(ax+b) = a^2 Var(x)$

6) Las unidades de medida de la varianza son las unidades al cuadrado de los datos.

Ejemplo: De un grupo de contribuyentes se determinó que el promedio de impuestos es de \$32.200, con una varianza de \$7.600. Determinar en cada uno de los siguientes casos, la nueva varianza:

a) Los impuestos aumentan en un 2 %

b) A los impuestos se les disminuye la cantidad de \$2.300

c) A cada contribuyente, se le disminuye un 3 % y además se le condona \$2.550

Solución:

a) $Var(x) = 7.600 \cdot (1,02)^2 = 7.907$ La nueva varianza es \$ 7.907

b) $Var(x) = 7.600$ La nueva varianza es \$ 7.600

c) $Var(x) = 7.600 \cdot (0,97)^2 = 7.150,8$ La nueva varianza es \$ $7.150,8$

Desviación Típica o Desviación Estándar

Es la raíz cuadrada positiva de la Varianza. Su símbolo es S si se está trabajando con una muestra y es σ si se está trabajando con una población.

a) Desviación estándar para datos no agrupados

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$ donde x_i representa los datos de la muestra.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N-1}}$$ donde x_i representa los datos de la población.

Ejemplo : Para el conjunto de datos $25, 12, 23, 28, 17, 15$ donde se obtuvo que su varianza era $S^2 = 39,2$; tendremos entonces que su desviación estándar es :

$S = \sqrt{39,2} = 6,26$ (unidades)

b) Desviación estándar para datos agrupados

Muestra	**Población**
$S = \sqrt{\frac{\sum_{i=1}^{m} (x_i - \overline{x})^2 f_i}{n-1}}$	$\sigma = \sqrt{\frac{\sum_{i=1}^{m} (x_i - \mu)^2 f_i}{N-1}}$

donde x_i es la marca de clase.

Ejemplo : Para el ejemplo de los datos tabulados sobre las edades de 26 personas se obtuvo como varianza $S^2 = 51,618$; luego su desviación estándar será :

$S = \sqrt{51,618} = 7,18$ (años)

¿Qué indican la Varianza y la Desviación Estándar?

La varianza y la desviación estándar miden la dispersión "promedio" en torno a la media aritmética, es decir, cómo fluctúan las observaciones mayores por encima de la media aritmética y cómo se distribuyen las observaciones menores por debajo de ella.

La varianza tiene ciertas propiedades matemáticas útiles. Sin embargo, al calcularla se obtienen unidades al cuadrado : cm^2, $pulgadas^2$, mm^2, $(edades)^2$, $(horas)^2$, etc. por ello, en la práctica, la principal medida de dispersión que se utiliza es la desviación estándar, cuyo valor está dado en las unidades originales : cm, pulgadas, mm, edades, horas, etc.

En los ejemplos anteriores:

a) Para la muestra de datos : $25, 12, 23, 28, 17, 15$ se obtuvo por desviación estándar : $S = 6,26$ (unidades). Esto indica que la mayor parte de los datos de esta muestra se agrupan dentro de $6,26$ unidades por encima y por debajo de la media aritmética, es decir, entre $20 - 6,26 = 13,74$ y $20 + 6,26 = 26,26$

b) Para el caso de los datos tabulados correspondientes a las edades de 26 personas, se obtuvo una desviación estándar de $S = 7,18$ años. Esto indica que la mayor parte de los datos están agrupados entre $29,96 - 7,18 = 22,78$ años y $29,96 + 7,18 = 37,14$ años.

Edades (años)	f_i
$15 - 20$	2
$\mathbf{21 - 26}$	$\mathbf{7}$
$\mathbf{27 - 32}$	$\mathbf{8}$
$\mathbf{33 - 38}$	$\mathbf{5}$
$39 - 44$	4
Total	26

Criterio de Homogeneidad

Una distribución se considera homogénea, si la desviación estándar se encuentra entre la quinta y la cuarta parte del rango. Si no es así, entonces se considera que la muestra es heterogénea.

a) Para la muestra de datos : $25,\ 12,\ 23,\ 28,\ 17,\ 15$

$R = 28 - 12 = 16$ $\qquad\qquad$ $S = 6,26$

$$\left[\frac{R}{5}, \frac{R}{4}\right] = [\,3,2\,;4,0\,]$$

$S\ \notin [\,3,2\,;\,4,0\,]$

Por lo tanto, la muestra es heterogénea.

b) Para el caso de los datos tabulados de las edades de 26 personas

Edades (años)	f_i
$15 - 20$	2
$21 - 26$	7
$27 - 32$	8
$33 - 38$	5
$39 - 44$	4
Total	26

$R = 44,5 - 14,5 = 30$ (años) $\qquad\qquad$ $S = 7,18$ (años)

$$\left[\frac{R}{5}, \frac{R}{4}\right] = [\,6\,;7,5\,]$$

$S\ \in [\,6\,;\,7,5\,]$

Por lo tanto, la muestra es homogénea.

Observaciones :

1) Cuanto más separados o dispersos estén los datos, es decir, para muestras heterogéneas, tanto mayores serán el rango, la varianza y la desviación estándar.

2) Si los datos están más concentrados, es decir, para muestras homogéneas, tanto menores serán el rango, la varianza y la desviación estándar.

3) Si todas las observaciones son iguales (de manera que no haya variación en los datos), el rango, la varianza y la desviación estándar serán iguales a cero.

Ejercicios

1) En una industria dos operarios en siete días de trabajo, son capaces de producir, por día, y en forma individual la siguiente cantidad de árboles para fresa de 250 mm de longitud por 300 mm de diámetro.

Operario A	105	106	104	102	103	100	101
Operario B	103	102	107	101	105	102	103

Determine :

a) Rango del operario A y del operario B
b) Varianza del operario A.
c) Desviación estándar de ambos operarios.
d) ¿Son las muestras homogéneas?.

2) Se hace una encuesta entre 100 personas acerca del número de horas diarias que se dedican a ver televisión, obteniéndose la siguiente información :

N° de horas	f_i
$0-1$	30
$2-3$	20
$4-5$	15
$6-7$	32
$8-9$	1
$10-11$	2
Total	100

Calcular la varianza y la desviación estándar.

3) De un total de 100 datos, 20 son 4, 40 son 5, 30 son 6 y el resto 7. Hallar la desviación estándar.

4) Cuatro grupos de estudiantes, consistentes en $15, 20, 10$ y 18 individuos, dieron pesos de $60, 72, 55$ y 65 kilos. Hallar la varianza de los estudiantes.

5) Las notas de un estudiante en sus certámenes han sido $84, 91, 72, 68, 87$ y 78. Hallar la desviación estándar. Las notas , ¿son homogéneas?.

6) La siguiente tabla corresponde a la estatura de 80 estudiantes de una determinada carrera:

Estatura	f_i
$1,65 - 1,69$	6
$1,70 - 1,74$	12
$1,75 - 1,79$	30
$1,80 - 1,84$	22
$1,85 - 1,89$	8
$1,90 - 1,94$	2
Total	80

Hallar rango, varianza y desviación estándar de la estatura.

7) La oficina de Censo, proporcionó las edades de hombres y mujeres divorciados (en miles de personas de 18 años de edad o más).

Edad	Hombre	Mujer
$15 - 19$	5	9
$20 - 24$	80	210
$25 - 29$	174	303
$30 - 34$	210	315
$35 - 39$	385	656
$40 - 44$	450	656
$45 - 49$	295	409
$50 - 54$	174	200
Total	1773	2758

Obtener las medidas de dispersión (rango, varianza y desviación estándar) tanto para los hombres como para las mujeres. Determine, además si las muestras son homogéneas o no.

Solución

1)

$a)\ R_{\text{A}} = 6$ $\qquad$ $R_{\text{B}} = 6$

$b)\ s^2_{\text{A}} = 4,67$

$c)\ s_{\text{A}} = 2,16$ $\qquad$ $s_{\text{B}} = 2,06$

$d)$ Ambas muestras no son homogéneas.

$2)\ s^2 = 7,19$ $\qquad$ $s = 2,68$

$3)\ s = 0,9$

$4)\ s^2 = 37,67$

$5)\ s = 8,92$ $\qquad$ Las notas no son homogéneas.

$6)\ R = 0,3$ $\qquad$ $s^2 = 0,0033$ $\qquad$ $s = 0,057$

7)

	Hombres	Mujeres
R	40	40
s^2	$64,03$	$67,68$
s	8	$8,23$

Ambas muestras son homogéneas.

Autoevaluación

1) En una encuesta realizada a 25 personas en la ciudad de Chillán, sobre su equipo de fútbol preferido, se obtuvieron los siguientes resultados:

U. de Chile, Colo Colo, U. Católica, Ñublense, Colo Colo, U. de Chile, Colo Colo Colo Colo, U. de Chile, Colo Colo, U. Católica, Ñublense, Colo Colo, U. de Chile, U. de Chile, U. de Chile, Colo Colo, U. Católica, Ñublense, Colo Colo, U. de Chile, U. Católica, Colo Colo, U. de Chile, Concepción

a) Construya una tabla para la información obtenida
b) Construya un gráfico adecuado para la información dada
c) ¿Cuántas personas son hinchas de Colo Colo?
d) ¿Qué porcentaje de personas prefiere a U. de Chile?
e) ¿Qué porcentaje de encuestados no es hincha de Ñublense?

2) Los salarios ofrecidos a 16 personas son (en miles de pesos):

165	149	166	167	154	165	144	135
155	170	150	151	142	148	149	100

Determine e interprete para la muestra:
a) Media aritmética
b) Moda
c) Mediana

3) Los impuestos pagados por un grupo de contribuyentes han dado origen a la siguiente tabla de frecuencia:

Monto de impuestos en miles	Nº personas
1 - 20	4
21 - 40	15
41 - 60	21
61 - 80	18
81 - 100	2
Total	60

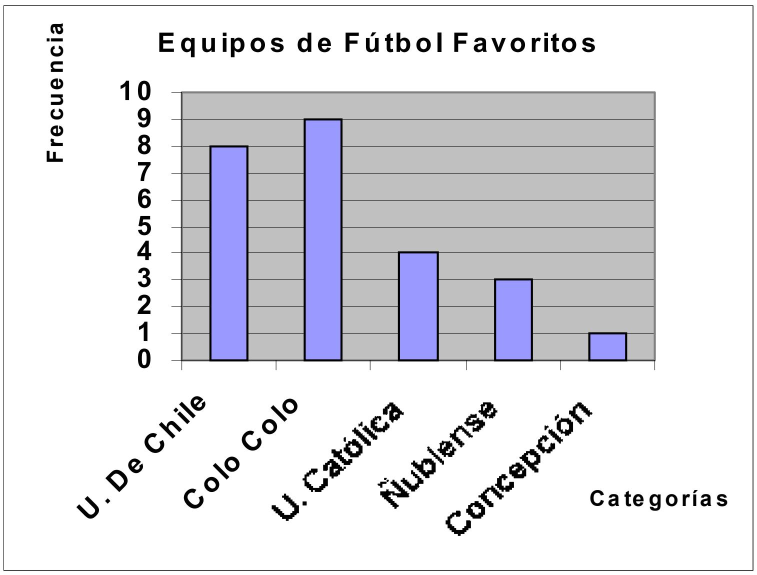
Determine:
a) Desviación Estándar Muestral y explique su significado
b) Determine si la muestra es homogénea o heterogénea. Justifique su respuesta.

Solución:

1) a)

Categorías	f_i	F_i	h_i	H_i
U. de Chile	8	8	$0,32$	$0,32$
Colo Colo	9	17	$0,36$	$0,68$
U. Católica	4	21	$0,16$	$0,84$
Ñublense	3	24	$0,12$	$0,96$
Concepción	1	25	$0,04$	$1,00$
Total	25		$1,00$	

b)



c) Las personas hinchas de Colo Colo son 9

d) El porcentaje de personas que prefiere a U. de Chile es $32\,\%$

e) El porcentaje de personas que no prefiere a Ñublense es $88\ \%$

2) a) $\overline{x} = 150,625$ El salario promedio es de \$ 150.625

b) $Me = 150,5$ El 50 % de las personas tiene un salario superior a \$ 150.500

c) $Mo = 149$ y 165 Los salarios más comunes son \$ 149.000 y \$ 165.000

3) a) $S = 19,655$ La desviación estándar es un estadístico que nos indica que tan dispersos están los datos, con respecto a la media aritmética.

b) Los datos no son homogéneos.

Unidad N°2: Probabilidades

Elementos de Probabilidades

Los primeros estudios de probabilidad fueron motivados por la posibilidad de acierto o fracaso en los juegos de azar. La probabilidad es un mecanismo por medio del cual pueden estudiarse sucesos aleatorios, es decir, operaciones cuyo resultado no puede ser predicho de antemano con seguridad. Por ejemplo, el lanzamiento de una moneda.

Enfoques de probabilidad

1) ***Experimento aleatorio o experimento***: cualquiera operación cuyo resultado no puede ser predicho de anterioridad con seguridad.

Ejemplo:
a) lanzamiento de una moneda
b) lanzamiento de un dado
c) extracción de una carta de una baraja de 52 cartas

2) ***Espacio muestral***: es el conjunto de todos los posibles resultados asociados a un experimento. Su símbolo es Ω . Si el espacio muestral tiene un número finito de elementos o infinito numerable, entonces se dice que éste es ***discreto*** y si el espacio muestral tiene como elementos todos los puntos de algún intervalo real, entonces se dice que éste es ***continuo*** .

Ejemplo:
a) experimento:lanzamiento de un dado
$\Omega = \{1, 2, 3, 4, 5, 6\}$

b) experimento: tiempo de duración de un tubo fluorescente
$\Omega = \{\, t\,, t \geq 0 \,\}$

3) ***Evento o suceso***: es cualquier subconjunto de un espacio muestral. Todo subconjunto es un evento, en particular Ω mismo es un evento, llamado ***suceso seguro*** y el conjunto vacío, $\emptyset$, también es un evento, llamado ***suceso imposible*** .

Ejemplo:
A $=$ {obtener un número impar al lanzar un dado}
A $= \{1, 3, 5\,\}$

B $=$ {obtener al menos una cara al lanzar una moneda dos veces}
B $= \{cs, sc, cc\}$

Como los eventos son subconjuntos de Ω, entonces es posible aplicar la teoría de conjuntos para obtener nuevos eventos.
Si A y B son eventos, entonces también lo son $A \cup B$, $A \cap B$, A^c
$A \cup B$ ocurre si, y sólo si sólo ocurre A o sólo ocurre B u ocurren A y B a la vez.
$A \cap B$ ocurre si, y sólo si ocurre A y ocurre B a la vez.
A^cocurre si, y sólo si no ocurre A.

En todo experimento aleatorio Ω se considera el conjunto universal, por lo tanto, todos los complementos son tomados respecto a Ω.

Ejemplo

Considere el experimento lanzamiento de dos dados.
a) Determine el espacio muestral

b) Obtenga los siguientes eventos:
A = {la suma de los dos números es un múltiplo de dos}
B = {ambos dados muestran la misma cara}
C = {los dos números son primos}
D = {la resta de los dos números es divisible por tres}

c) Encuentre, si es posible, $A \cup B$, $C \cap D$, B^c, $B^c \cap C^c$

$a)$

$$\Omega = \left\langle \begin{pmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{pmatrix} \right\rangle$$

$b)$

$$A = \left\{ \begin{matrix} (1,1) & (1,3) & (1,5) & (2,2) & (2,4) & (2,6) \\ (3,1) & (3,3) & (3,5) & (4,2) & (4,4) & (4,6) \\ (5,1) & (5,3) & (5,5) & (6,2) & (6,4) & (6,6) \end{matrix} \right\}$$

$$B = \{\, (1,1) \quad (2,2) \quad (3,3) \quad (4,4) \quad (5,5) \quad (6,6) \,\}$$

$$C = \left\{ \begin{matrix} (2,2) & (2,3) & (2,5) \\ (3,2) & (3,3) & (3,5) \\ (5,2) & (5,3) & (5,5) \end{matrix} \right\}$$

$$D = \{\, (1,4) \quad (2,5) \quad (3,6) \quad (4,1) \quad (5,2) \quad (6,3) \,\}$$

$c)$

$A \cup B = A$

$C \cap D = \{\, (2,5) \quad (5,2) \,\}$

$B^c = \{(x,y)/x \neq y\}$

$$B^c \cap C^c = \begin{pmatrix} (1,2) & (1,3) & (1,4) & (1,5) & (1,6) & (2,1) \\ (2,4) & (2,6) & (3,1) & (3,4) & (3,6) & (4,1) \\ (4,2) & (4,3) & (4,5) & (4,6) & (5,1) & (5,4) \\ (5,6) & (6,1) & (6,2) & (6,3) & (6,4) & (6,5) \end{pmatrix}$$

Concepto de probabilidad en espacio finito equiprobable

Si Ω es un espacio muestral con n elementos, entonces la probabilidad de un evento A es el cuociente $\frac{m}{n}$, donde m es el número de elementos de A

Esto se denota: $\mathrm{P}(\mathrm{A}) = \frac{m}{n}$

Ejemplo :

$\Omega = \{$lanzamiento de un dado$\}$ $\Rightarrow \Omega = \{1, 2, 3, 4, 5, 6\}$
A $= \{$aparece un múltiplo de tres$\}$ $\Rightarrow \mathrm{A} = \{3, 6\}$

$\mathrm{P}(\mathrm{A}) = \frac{2}{6} = \frac{1}{3}$

Definición: Diremos que dos eventos A y B son ***mutuamente excluyentes o disjuntos*** si no pueden ocurrir juntos, es decir $\mathrm{A} \cap \mathrm{B} = \emptyset$

Por ejemplo, $\Omega = \{$ lanzamiento de un dado$\}$ $\Rightarrow \Omega = \{ 1, 2, 3, 4, 5, 6\}$
A $= \{$ aparece un múltiplo de tres $\}$ $\Rightarrow A = \{ 3, 6 \}$
B $= \{$ aparece un múltiplo de cuatro $\}$ $\Rightarrow B = \{4\}$

Luego, A y B son eventos disjuntos, porque $\mathrm{A} \cap \mathrm{B} = \emptyset$

Axiomas de probabilidad

Sea Ω un espacio muestral y sean A y B dos eventos cualesquiera de este:

Axioma1 : $\mathrm{P}(\Omega) = 1$

Axioma2 : $\mathrm{P}(\mathrm{A}) \geq 0 \qquad \forall \, \mathrm{A} \subseteq \Omega$

Axioma3 : $\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B})$ si $\mathrm{A} \cap \mathrm{B} = \emptyset$
En general, $\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i \right) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + ... + \mathrm{P}(A_i)$ con

$A_i \cap A_j = \emptyset \;\; \forall \, i \neq j$

De estos tres axiomas fundamentales es posible determinar algunas propiedades y consecuencias:

Teorema1 :

a) $\mathrm{P}(\emptyset) = 0$

Demostración

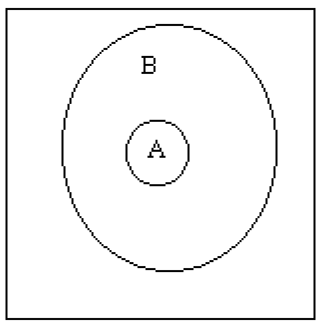
$\Omega = \Omega \cup \emptyset$
$\mathrm{P}(\Omega) = \mathrm{P}(\Omega \cup \emptyset)$
$\mathrm{P}(\Omega) = \mathrm{P}(\Omega) + \mathrm{P}(\emptyset)$ pues $\Omega \cap \emptyset = \emptyset$
$1 = 1 + 0$
$0 = \mathrm{P}(\emptyset)$

b) $P(A^c) = 1 - P(A)$

Demostración

$\Omega = A \cup A^c$
$P(\Omega) = P(A \cup A^c)$
$P(\Omega) = P(A) + P(A^c)$ pues $A \cap A^c = \emptyset$
$1 = P(A) + P(A^c)$
$1 - P(A) = P(A^c)$

c) Si $A \subseteq B$, entonces $P(A) \leq P(B)$



Demostración

$B = A \cup (B - A)$
$P(B) = P[A \cup (B - A)]$
$P(B) = P(A) + P(B - A)$ pues $A \cap (B - A) = \emptyset$
Luego, $P(A) \leq P(B)$

Corolario

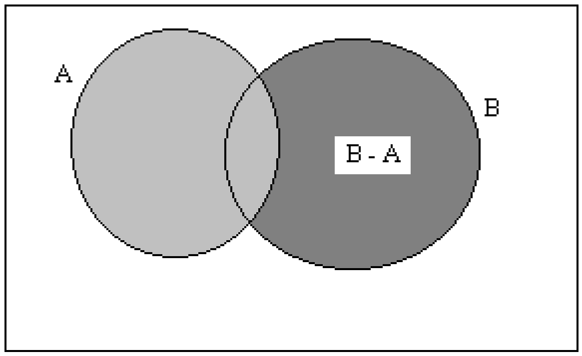
$0 \leq P(A) \leq 1$

Demostración

$\emptyset \subseteq A \subseteq \Omega$
$P(\emptyset) \leq P(A) \leq P(\Omega)$
$0 \leq P(A) \leq 1$

Teorema 2 :

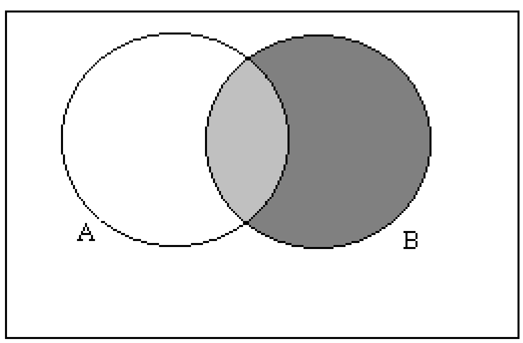
a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Demostración

$A \cup B = A \cup (B - A)$
$P(A \cup B) = P[A \cup (B - A)]$
$P(A \cup B) = P(A) + P(B - A)$ pues $A \cap (B - A) = \emptyset$
$P(A \cup B) - P(A) = P(B - A)$ (1)

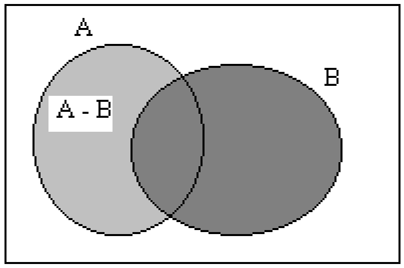
Por otro lado



$B = (A \cap B) \cup (B - A)$
$P(B) = P(A \cap B) + P(B - A)$ pues $(A \cap B) \cap (B - A) = \emptyset$
$P(B) - P(A \cap B) = P(B - A)$ (2)

de (1) y (2)
$P(A \cup B) - P(A) = P(B) - P(A \cap B)$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

b) $P(A - B) = P(A) - P(A \cap B)$

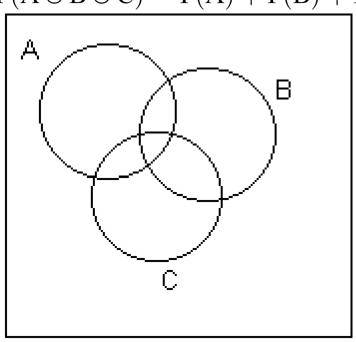


Demostración

$A \cup B = (A - B) \cup B$
$P(A \cup B) = P[(A - B) \cup B]$
$P(A) + P(B) - P(A \cap B) = P(A - B) + P(B)$ pues $(A - B) \cap B = \emptyset$
$P(A) - P(A \cap B) = P(A - B)$

Corolario

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



Demostración

$A \cup B \cup C = (A \cup B) \cup C$
$P(A \cup B \cup C) = P[(A \cup B) \cup C]$
$= P(A \cup B) + P(C) - P[(A \cup B) \cap C]$
$= P(A) + P(B) - P(A \cap B) + P(C) - P[(A \cap C) \cup (B \cap C)]$
$= P(A) + P(B) - P(A \cap B) + P(C) - [P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)]$
$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Teorema3 :

Sea Ω un espacio muestral y A un evento de Ω , $A \subseteq \Omega$, entonces

$$P(A) = P(A_1) + P(A_2) + P(A_3) + ... + P(A_k)$$
$$= \sum_{i=1}^{k} P(A_i)$$

Donde A_i son eventos disjuntos cuya unión es A

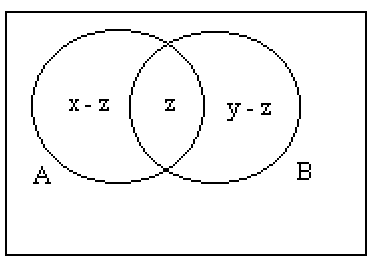
Demostración

$A = A_1 \cup A_2 \cup A_3 \cup ... \cup A_k$
$P(A) = P(A_1 \cup A_2 \cup A_3 \cup ... \cup A_k)$
$P(A) = P(A_1) + P(A_2) + P(A_3) + ... + P(A_k)$ pues $A_i \cap A_j = \emptyset$
$P(A) = \sum_{i=1}^{k} P(A_i)$

Ejemplos

1) Suponga que A y B son eventos para los cuales $P(A) = x$; $P(B) = y$ y $P(A \cap B) = z$. Determine:

a) $P(A^c \cup B^c)$
b) $P(A^c \cup B)$
c) $P(A^c \cap B)$
d) $P(A^c \cap B^c)$



Solución

a) $P(A^c \cup B^c) = P[(A \cap B)^c]$
$= 1 - P(A \cap B)$
$= 1 - z$

b) $P(A^c \cup B) = P[(A - B)^c]$
$= 1 - P(A - B)$
$= 1 - [P(A) - P(A \cap B)]$
$= 1 - x + z$

c) $P(A^c \cap B) = P(B - A)$
$= P(B) - P(A \cap B)$
$= y - z$

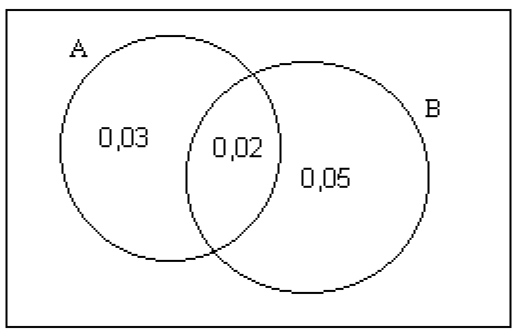
d) $P(A^c \cap B^c) = P[(A \cup B)^c]$
$= 1 - P(A \cup B)$
$= 1 - P(A) - P(B) + P(A \cap B)$
$= 1 - x - y + z$

2) De la producción de tornillos de cierta magnitud resulta que el 5 % de ellos no tienen el largo especificado, el 7 % no tienen el diámetro especificado y el 2 % tiene ambos defectos. Se elige un tornillo al azar de la producción de estas magnitudes. ¿Cuál es la probabilidad que:

a) tenga al menos uno de los dos defectos?.
b) tenga sólo el defecto del largo?
c) tenga sólo uno de los dos defectos?
d) no tenga defectos?

Solución

$A = \{\text{tornillos con defecto del largo}\}$
$B = \{\text{tornillos con defecto del diámetro}\}$



a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0,05 + 0,07 - 0,02$
$= 0,10$

La probabilidad de que tenga al menos uno de los dos defectos es de 0,10

b) $P(A - B) = P(A) - P(A \cap B)$
$= 0,05 - 0,02$
$= 0,03$

La probabilidad de que tenga sólo el defecto del largo es de 0,03

c) $P(A - B) + P(B - A) = (0,03) + [P(B) - P(A \cap B)]$
$= 0,03 + 0,07 - 0,02$
$= 0,08$

La probabilidad de que tenga sólo uno de los dos defectos es de 0,08

d) $P(A \cup B)^c = 1 - P(A \cup B)$
$= 1 - 0,10$
$= 0,90$

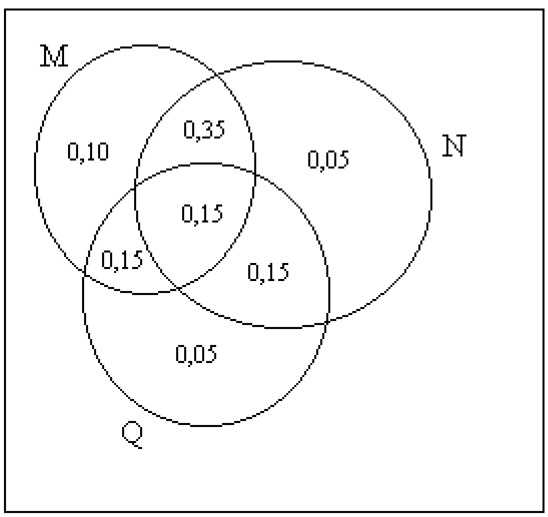
La probabilidad de que no tenga defectos es de 0,90

3) La alimentación de cierta especie se considera completa si cada individuo consume tres tipos de alimentos en cantidades adecuadas. En una población se encontró que el 75 % consume alimento tipo A, el 70 % alimento tipoB, el 50 % alimento tipo C, el 50 % alimento tipo A y B, el 30 % alimento tipo A y C, el 30 % alimento tipo B y C y el 15 % consume de los tres tipos de alimentos. Se elige un individuo al azar en la población, calcular la probabilidad que:

a) consuma sólo alimento tipo C.
b) consuma sólo un tipo de alimento.
c) consuma al menos dos tipos de alimentos

Solución

M = {individuo de la población que consume alimento tipo A}
N = {individuo de la población que consume alimento tipo B}
Q = {individuo de la población que consume alimento tipo C}



a) La probabilidad de que un individuo sólo consuma alimento tipo C es de 0,05
b) La probabilidad de que un individuo consuma sólo un tipo de alimento es de 0,20 .
c) La probabilidad de que un individuoconsuma al menos dos tipos de alimentos es de 0,80.

Ejercicios

1) Si A,B y C son eventos mutuamente excluyentes, y $P(A) = 0,2$; $P(B) = 0,3$; $P(C) = 0,2$
Encuentre :

a) P(A U B U C)

b) $P\left[A^c \cap (B \cup C) \right]$

c) P(B U C)

2) Sean A y B eventos tales que $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$; $P(A \cap B) = \frac{1}{4}$, calcule :

a) $P(A^c)$

b) $P(B^c)$

c) $P(A \cup B)$

d) $P(A - B)$

e) $P(A^c \cup B^c)$

f) $P(A^c \cap B^c)$

3) De un total de 500 estudiantes, se encuentra que 210 fuman, que 258 toman bebidas alcohólicas, que 216 toman alimentos entre comidas, que 122 fuman y toman bebidas alcohólicas, que 83 toman alimentos entre comidas y también bebidas alcohólicas, que 97 fuman y toman alimentos entre comidas y que 52 practican estos tres dañinos hábitos. Si se escoge aleatoriamente a un miembro de esta generación, encuentre la probabilidad de que el estudiante :

a) fumen, pero no tome bebidas alcohólicas.
b) tome alimentos entre comidas e ingiera bebidas alcohólicas, pero no fume.
c) no fume y no tome alimentos entre comidas.

4) La probabilidad de que una industria XX se ubique en la ciudad A es de 0,7; de que se localice en la cuidad B es de 0,4 y de que se encuentre en A o en B, o en ambas es de 0,8. ¿Cuál es la probabilidad de que la industria se localice :

a) en ambas cuidades?.
b) en ninguna de ellas?.

5) En una bolsa hay 36 fichas numeradas del 1 al 36, respectivamente. Si se extrae una ficha, calcular la probabilidad de que la ficha extraída sea :

a) un número par

b) un número primo

c) un múltiplo de 5

d) un número terminado en 2

e) un número divisible por 6

f) un número impar mayor que 20.

Solución

1)
a) P(A U B U C) $= 0,7$ b) P$\left[\, A^c \cap (B \cup C)\,\right] = 0,5$
c) P(B U C) $= 0,5$

2)
a) $P(A^c) = \dfrac{1}{2}$ b) $P(B^c) = \dfrac{2}{3}$

c) $P(A \cup B) = \dfrac{7}{12}$ d) $P(A - B) = \dfrac{1}{4}$

e) $P(A^c \cup B^c) = \dfrac{3}{4}$ f) $P(A^c \cap B^c) = \dfrac{5}{12}$

3)
a) La probabilidad de que fumen, pero no tome bebidas alcohólicas es $\dfrac{88}{500}$

b) La probabilidad de que tome alimentos entre comidas e ingiera bebidas alcohólicas, pero no fume es $\dfrac{31}{500}$

c) La probabilidad de que no fume y no tome alimentos entre comidas es $\dfrac{171}{500}$

4)
a) La probabilidad de que la industria se localice en ambas ciudades es $0,3$

b) La probabilidad de que la industria no se localice en ninguna de ellas es $0,2$

5)
a) La probabilidad de que la ficha extraída sea un número par es $\dfrac{1}{2}$

b) La probabilidad de que la ficha extraída sea un número primo es $\dfrac{11}{36}$

c) La probabilidad de que la ficha extraída sea un múltiplo de 5 es $\dfrac{7}{36}$

d) La probabilidad de que la ficha extraída sea un número terminado en 2 es $\dfrac{1}{9}$

e) La probabilidad de que la ficha extraída sea un número divisible por 6 es $\dfrac{1}{6}$

f) La probabilidad de que la ficha extraída sea un número impar mayor que 20 es $\dfrac{2}{9}$

Probabilidad Condicional

Cuando se está calculando la probabilidad de un evento A en particular, y se tiene información sobre la ocurrencia de otro evento B, esta probabilidad se conoce como ***probabilidad condicional***, la cual se denota por $P(A/B)$, se lee "probabilidad de A dado B" y se define como:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{con } P(B) \neq 0$$

Las probabilidades condicionales satisfacen los axionas de probabilidad

1) $P(\Omega/B) = 1$

$$\begin{aligned} P(\Omega/B) &= \frac{P(\Omega \cap B)}{P(B)} \\ &= \frac{P(B)}{P(B)} \\ &= 1 \end{aligned}$$

2) $P[(A \cup C)/B] = P(A/B) + P(C/B)$ $\qquad A \cap C = \emptyset$

$$\begin{aligned} P[(A \cup C)/B] &= \frac{P[(A \cup C) \cap B]}{P(B)} \\ &= \frac{P[(A \cap B) \cup (C \cap B)]}{P(B)} \\ &= \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)} \\ &= P(A/B) + P(C/B) \end{aligned}$$

Ejemplos

1) La probabilidad de que un vuelo de programación regular despegue a tiempo es $P(D) = 0,83$; la que llegue a tiempo es $P(A) = 0,82$ y la que despegue y llegue a tiempo es $P(D \cap A) = 0,78$. Encuentre la probabilidad de que el avión:

a) llegue a tiempo dado que despegó a tiempo.
b) despegue a tiempo dado que llegó a tiempo

Solución

$D = \{$ despegar a tiempo$\}$

$A = \{$ llegar a tiempo$\}$

a) $$\begin{aligned} P(A/D) &= \frac{P(A \cap D)}{P(D)} \\ &= \frac{0,78}{0,83} = 0,94 \end{aligned}$$

La probabilidad de que el avión llegue a tiempo dado que despegó a tiempo es de $0,94$.

b) $\mathrm{P(D/A)} = \dfrac{\mathrm{P(D \cap A)}}{\mathrm{P(A)}}$
$= \dfrac{0,78}{0,82}$
$= 0,95$

La probabilidad de que el avión despegue a tiempo dado que llegó a tiempo es de 0,95 .

2) En una oficina hay 100 máquinas calculadoras, algunas de ellas son eléctricas (E) mientras que otras son manuales (M). De ellas unas son nuevas (N) y otras usadas (U). El número de máquinas por categoría está dada en la siguiente tabla:

	E	M	Total
N	40	30	70
U	20	10	30

Una persona entra a la oficina y escoge una máquina al azar, descubre que es nueva. ¿Cuál es la probabilidad que sea eléctrica?

$$\begin{aligned}\mathrm{P(E/N)} &= \frac{\mathrm{P(E \cap N)}}{\mathrm{P(N)}}\\ &= \frac{\frac{40}{100}}{\frac{70}{100}}\\ &= \frac{4}{7}\end{aligned}$$

La probabilidad es de 0,57 .

3) Un grupo de 500 ejecutivos es clasificado de acuerdo a las características del peso y a la insidencia del peso en la hipertensión. Se da la siguiente tabla:

	Sobre peso(SP)	Peso normal(PN)	Bajo peso(BP)	Total
Hipertenso(H)	50	40	10	100
No hipertenso($\mathrm{H^c}$)	75	225	100	400
Total	125	265	110	500

a) ¿Cuál es la probabilidad de que una persona elegida al azar sea hipertensa?

b) Una persona elegida al azar tiene sobrepeso. ¿Cuál es la probabilidad que también sea hipertensa?

c) Una persona elegida al azar no es hipertensa. ¿Cuál es la probabilidad de que tenga peso normal?

a) $P(H)=\dfrac{100}{500}=\dfrac{1}{5}$

La probabilidad de que una persona sea hipertensa es de $0{,}20$.

b) $$\begin{aligned} P(H/SP) &= \frac{P(H\cap SP)}{P(SP)} \\ &= \frac{\frac{50}{500}}{\frac{125}{500}} \\ &= \frac{2}{5} \end{aligned}$$

La probabilidad de que una persona con sobrepeso sea también hipertensa es de $0{,}40$.

c) $$\begin{aligned} P(N/H^c) &= \frac{P(N\cap H^c)}{P(H^c)} \\ &= \frac{\frac{225}{500}}{\frac{400}{500}} \\ &= \frac{9}{16} \end{aligned}$$

La probabilidad de que una persona no hipertensa tenga también peso normal es de $0{,}5625$.

Uno de los usos más frecuentes de la probabilidad condicional es dar un procedimiento fácil para asignar probabilidades a intersecciones de eventos. Del concepto de probabilidad condicional es posible encontrar una expresión útil, llamada regla del producto, para la probabilidad de intersección de eventos, esta es:

$$P(A/B)=\frac{P(A\cap B)}{P(B)}$$

$$\boxed{\mathbf{P(AB)=P(A/B)\cdot P(B)}}$$

Así,

$$\begin{aligned} P(A\cap B\cap C) &= P(A/B\cap C)\cdot P(B\cap C) \\ &= P(A/B\cap C)\cdot P(B/C)\cdot P(C) \end{aligned}$$

$$\begin{aligned} P(A\cap B\cap C\cap D) &= P(A/B\cap C\cap D)\cdot P(B\cap C\cap D) \\ &= P(A/B\cap C\cap D)\cdot P(B/C\cap D)\cdot P(C\cap D) \\ &= P(A/B\cap C\cap D)\cdot P(B/C\cap D)\cdot P(C/D)\cdot P(D) \end{aligned}$$

Ejemplos:

1) Se seleccionan 2 fichas al azar, sin reemplazo, de una urna que contiene 4 blancas y 8 negras. Calcular la probabilidad de que:

a) ambas sean blancas.

b) la segunda sea blanca.

a) $\mathrm{B} = \{\text{fichas blancas}\}$

$\mathrm{N} = \{\text{fichas negras}\}$

$\mathrm{P(B)} = \dfrac{4}{12}$ $\qquad\qquad$ $\mathrm{P(N)} = \dfrac{8}{12}$

$$\mathrm{P(B_1 \cap B_2) = P(B_1) \cdot P(B_2/B_1)}$$

$$= \frac{4}{12} \cdot \frac{3}{11}$$

$$= \frac{1}{11}$$

La probabilidad de ambas fichas sean blancas es de 0,09 .

b) $\mathrm{P(B_1 \cap B_2) + P(N_1 \cap B_2)} \qquad = \dfrac{1}{11} + \mathrm{P(N_1) \cdot P(B_2/N_1)}$

$$= \frac{1}{11} + \frac{8}{12} \cdot \frac{4}{11}$$

$$= \frac{1}{3}$$

La probabilidad de que la segunda ficha sea blanca es de 0,33 .

2) Una caja de fusibles contiene 20 unidades, de las cuales 5 son defectuosas. Si tres de estos fusibles son tomados al azar, en sucesión y sin reemplazo.

a) ¿Cuál es la probabilidad que los tres sean defectuosos?

b) Si en cada una de las dos primeras se extrajo un defectuoso.¿Cuál es la probabilidad que el tercero extraido sea bueno?

c) Si los dos primeros estaban buenos. ¿Cuál es la probabilidad que el tercero extraído sea defectuoso?

d) ¿Cuál es la probabilidad que los dos primeros sean buenos y el tercero defectuoso?

$\mathrm{D} = \{\text{fusible defectuoso}\}$

$\mathrm{D^c} = \{\text{fusible no defectuoso}\}$

$\mathrm{P(D)} = \dfrac{5}{20}$ $\qquad\qquad$ $\mathrm{P(D^c)} = \dfrac{15}{20}$

a) $\mathrm{P(D_1 \cap D_2 \cap D_3)} \qquad = \mathrm{P(D_1) \cdot P(D_2/D_1) \cdot P(D_3/D_1 \cap D_2)}$

$$= \frac{5}{20} \cdot \frac{4}{19} \cdot \frac{3}{18} \quad = \frac{1}{144}$$

La probabilidad es de $\dfrac{1}{144}$

b) $\mathrm{P}(\mathrm{D}_3^c/\mathrm{D}_1 \cap \mathrm{D}_2) = \dfrac{15}{18}$

La probabilidad es de un $0,83$.

c) $\mathrm{P}(\mathrm{D}_3/\mathrm{D}_1^c \cap \mathrm{D}_2^c) = \dfrac{5}{18}$

La probabilidad es de un $0,27$.

d) $$\begin{aligned} \mathrm{P}(\mathrm{D}_1^c \cap \mathrm{D}_2^c \cap \mathrm{D}_3) &= \mathrm{P}(\mathrm{D}_1^c) \cdot \mathrm{P}(\mathrm{D}_2^c/\mathrm{D}_1^c) \cdot \mathrm{P}(\mathrm{D}_3/\mathrm{D}_1^c \cap \mathrm{D}_2^c) \\ &= \frac{15}{20} \cdot \frac{14}{19} \cdot \frac{5}{18} \\ &= \frac{35}{228} \end{aligned}$$

La probabilidad es de un $0,1535$.

Ejercicios

1) La probabilidad de que un automóvil al que se le llena el tanque de gasolina necesite también un cambio de aceite es de 0,25 ; la de que requiera un nuevo filtro de aceite es de 0,40 y de que le haga falta tanto cambio de aceite como de filtro es de 0,14.

a) Si se debe cambiar el aceite, ¿cuál es la probabilidad de que necesite un filtro nuevo?.

b) Si se necesita un filtro nuevo, ¿cuál es la probabilidad de que requiera un cambio de aceite?.

2) Para parejas de casados que viven en una cierta ciudad de los suburbios., la probabilidad de que el esposo vote en alguna elección es de 0,21, la de que su esposa lo haga, de 0,28 y la de que ambos voten, de 0,15. ¿Cuál es la probabilidad de

a) al menos un miembro de la pareja de casados vote?.

b) vote la esposa, dado que su esposo lo hace?.

c) vote un esposo, dado que su esposa no lo hace?.

3) De una caja que contiene 6 pelotas negras y 4 verdes, se sacan tres en sucesión, reemplazándose cada pelota en la caja antes de extraer la siguiente.

a) ¿Cuál es la probabilidad de que las tres sean del mismo color?.

b) ¿Cuál es la probabilidad de que primera pelota sea negra, la segunda verde y la tercera negra?.

c) Repita las mismas preguntas anteriores, pero asuma que no hay reemplazo.

4) Una urna contiene 7 bolas rojas y 3 bolas blancas. Se sacan 3 bolas de la urna . Hallar la probabilidad de que las dos primeras sean rojas y la tercera blanca.

a) las bolas se devuelven a la urna.

b) las bolas no se devuelven a la urna.

5) En cierta facultad, 25 % de los estudiantes perdieron matemáticas, 15 % perdieron química y 10 % perdieron las dos. Se selecciona un estudiante al azar.

a) Si perdió química, ¿cuál es probabilidad de que perdió matemáticas?

b) Si perdió matemáticas, ¿cuál es probabilidad de que perdió química?

c) ¿Cuál es probabilidad de que perdió matemáticas o química?

6) Sean A y B eventos con $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ y $P(A \cap B) = \frac{1}{4}$. Hallar :

a) $P(A/B)$

b) $P(B/A)$

c) $P(A \cup B)$

d) $P(A^c/B^c)$

e) $P(B^c/A^c)$

7) A un jugador le reparten 5 cartas de una baraja corriente de 52 cartas. ¿Cuál es la probabilidad de que todas sean corazones?.

8) Una clase tiene 15 niñas y 19 niños. Si se escogen tres estudiantes al azar.¿Cuál es probabilidad de que :

a) todos sean niños.

b) todos sean niñas.

c) al menos uno sea niño

d) dos sean mujeres.

e) al menos dos sean niños.

9) Se estima que la probabilidad de que aumenten las ventas de automóviles en el siguiente mes es de 0,40. Se estima que la probabilidad de que aumenten las ventas de refacciones es de 0,30. Se estima que la probabilidad de que ambas industrias experimenten un aumento en ventas es de 0,10. ¿Cuál es la probabilidad de que :

a) hayan aumentado las ventas de automóviles durante el mes, dado que existe información de que han aumentado las ventas de refacciones?

b) hayan aumentado las ventas de refacciones, dado que existe información de que aumentaron las ventas de automóviles durante el mes?

Solución

1) $A = \{\text{cambio de aceite}\}$ $B = \{\text{nuevo filtro}\}$

a) $P(B/A) = 0,56$ b) $P(A/B) = 0,35$

2) $A = \{\text{esposo vota}\}$ $B = \{\text{esposa vota}\}$

a) $P(A \cup B) = 0,34$ b) $P(B/A) = 0,71$
c)$P(A/B^c) = 0,08$

3) $N = \{\text{pelota negra}\}$ $V = \{\text{pelota verde}\}$

a) $P(N_1 \cap N_2 \cap N_3) + P(V_1 \cap V_2 \cap V_3) = \frac{7}{25} = 0,28$

b) $P(N_1 \cap V_2 \cap N_3) = \frac{18}{125} = 0,144$

c) $P(N_1 \cap N_2 \cap N_3) + P(V_1 \cap V_2 \cap V_3) = \frac{1}{5} = 0,20$

$P(N_1 \cap V_2 \cap N_3) = \frac{1}{6} = 0,1666$

4) $R = \{\text{pelota roja}\}$ $B = \{\text{pelota blanca}\}$

a) $P(R_1 \cap R_2 \cap B_3) = \frac{147}{1000}$

b) $P(R_1 \cap R_2 \cap B_3) = \frac{126}{720}$

5) $A = \{\text{perder matemáticas}\}$ $B = \{\text{perder química}\}$

a)$P(A/B) = \frac{2}{3}$ b) $P(B/A) = \frac{2}{5}$
c) $P(A \cup B) = 0,3$

6)
a) $P(A/B) = \frac{3}{4}$ b) $P(B/A) = \frac{1}{2}$

c) $P(A \cup B) = \frac{7}{12}$ d) $P(A^c/B^c) = \frac{5}{8}$

e) $P(B^c/A^c) = \frac{5}{6}$

7) $P(C_1 \cap C_2 \cap C_3 \cap C_4 \cap C_5) = 0,005$

8) $A = \{\text{niñas}\}$ $\qquad$ $B = \{\text{niños}\}$

a) $P(B_1 \cap B_2 \cap B_3) = 0,16$

b) $P(A_1 \cap A_2 \cap A_3) = 0,08$

c) $3P(B_1 \cap A_2 \cap A_3) + 3P(B_1 \cap B_2 \cap A_3) + P(B_1 \cap B_2 \cap B_3) = 0,92$

d) $3P(B_1 \cap A_2 \cap A_3) = 0,33$

e) $3P(B_1 \cap B_2 \cap A_3) + P(B_1 \cap B_2 \cap B_3) = 0,59$

9) $A = \{\text{aumento venta de automóviles}\}$
$B = \{\text{aumento ventas de refacciones}\}$

a) $P(A/B) = \dfrac{1}{3}$ $\qquad$ b) $P(B/A) = \dfrac{1}{4}$

Teorema: (Probabilidad total) Suponga que los eventos $A_1, A_2, ..., A_k$ forman una partición de Ω, es decir, $A_1 \cup A_2 \cup ... \cup A_k = \Omega$, $A_i \neq \emptyset$ y $A_i \cap A_j = \emptyset \quad \forall i \neq j$. Entonces para cualquier evento $E \subset \Omega$ se tiene:

$$P(E) = \sum_{i=1}^{k} P(A_i) \cdot P(E/A_i)$$

Teorema de Bayes:

Si $A_1, A_2, ..., A_k$ es una partición de Ω, es decir, $A_1 \cup A_2 \cup ... \cup A_k = \Omega$, $A_i \neq \emptyset$ y $A_i \cap A_j = \emptyset$ $\forall i \neq j$. Entonces para cualquier evento $B \subset \Omega$ se tiene:

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

$$P(A_i/B) = \frac{P(B \cap A_i)}{P(B \cap A_1) + P(B \cap A_2) + ... + P(B \cap A_k)}$$

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{P(B/A_1) \cdot P(A_1) + P(B/A_2) \cdot P(A_2) + ... + P(B/A_k) \cdot P(A_k)}$$

Ejemplos:

1) La probabilidad de que Alicia estudie para su examen final de Estadística es 0,2 . Si estudia la probabilidad de que apruebe el examen es 0,8, en tanto que si no estudia la probabilidad es 0,5.

a) ¿Cuál es la probabilidad que Alicia apruebe estadística?.

b) Dado que Alicia aprobó su examen. ¿Cuál es la probabilidad de que haya estudiado?.

$E = \{\text{Alicia estudia}\}$
$E^c = \{\text{Alicia no estudia}\}$
$A = \{\text{Alicia aprueba estadística}\}$

$P(E) = 0,2 \qquad P(E^c) = 0,8 \qquad P(A/E) = 0,8 \qquad P(A/E^c) = 0,5$

a) $P(A) = P(A \cap E) + P(A \cap E^c)$
$P(A) = P(A/E) \cdot P(E) + P(A/E^c) \cdot P(E^c)$
$P(A) = (0,8)(0,2) + (0,5)(0,8)$
$P(A) = 0,56$

La probabilidad de que Alicia apruebe estadística es de 0,56 .

b) $\mathrm{P(E/A)} = \dfrac{\mathrm{P(E \cap A)}}{\mathrm{P(A)}}$

$$= \frac{\mathrm{P(A \cap E)}}{\mathrm{P(A)}}$$

$$= \frac{\mathrm{P(A/E) \cdot P(E)}}{\mathrm{P(A)}}$$

$$= \frac{(0,8)(0,2)}{0,56}$$

$$= 0,29$$

La probabilidad de que Alicia haya estudiado dado que aprobó estadística es de 0,29 .

2) Componentes complejas son ensambladas en una planta que usa dos líneas de ensamblado A y B. La línea A usa equipos más viejos que la línea B de manera que es algo más lenta y menos confiable. Suponga que en un día dado, la línea A ha ensamblado 8 componentes de los cuales 2 son defectuosos y 6 son no defectuosos, mientras que la línea B ha producido 1 componente defectuoso y 9 componentes no defectuosos. El encargado de ventas selecciona al azar una de estas 18 componentes para una demostración y encuentra que es defectuosa. ¿Cuál es la probabilidad que esta componente haya sido ensamblada por la línea A?.

A = {línea A}
B = {línea B}
D = {artículo defectuoso}

$\mathrm{P(A)} = \mathrm{P(B)} = \dfrac{1}{2}$ $\qquad \mathrm{P(D/A)} = \dfrac{2}{8}$ $\qquad \mathrm{P(D/B)} = \dfrac{1}{10}$

$$\mathrm{P(A/D)} = \frac{\mathrm{P(A \cap D)}}{\mathrm{P(D)}}$$

$$= \frac{\mathrm{P(D \cap A)}}{\mathrm{P(D \cap A) + P(D \cap B)}}$$

$$= \frac{\mathrm{P(D/A) \cdot P(A)}}{\mathrm{P(D/A) \cdot P(A) + P(D/B) \cdot P(B)}}$$

$$= \frac{\frac{2}{8} \cdot \frac{1}{2}}{\frac{2}{8} \cdot \frac{1}{2} + \frac{1}{10} \cdot \frac{1}{2}}$$

$$= \frac{5}{7}$$

La probabilidad de que la componente defectuosa la haya producido la línea A es de 0,71 .

3) De un grupo gande de habitantes de una ciudad que tiene igual número de personas en administración, comercio, servicio de salud y servicio municipal se encontró que el 35 % de los administrativos, el 25 % de los comerciantes, el 20 % del servicio de salud y el 15 % del servicio municipal eran mujeres.

a) ¿Cuál es la probabilidad que una mujer escogida al azar del grupo sea administrativa?

b) ¿Cuál es la probabilidad que un individuo del grupo elegido al azar sea hombre?

$A = \{\text{administrativo}\}$ $\quad$ $B = \{\text{comerciante}\}$

$C = \{\text{servicio salud}\}$ $\quad$ $D = \{\text{servicio municipal}\}$

$M = \{\text{mujer}\}$ $\quad$ $M^c = \{\text{hombre}\}$

$$P(A) = P(B) = P(C) = P(D) = \frac{1}{4}$$

$P(M/A) = 0,35$ $\quad$ $P(M/B) = 0,25$

$P(M/C) = 0,20$ $\quad$ $P(M/D) = 0,15$

a) $$P(A/M) = \frac{P(A \cap M)}{P(M)}$$

$$= \frac{P(M \cap A)}{P(M \cap A) + P(M \cap B) + P(M \cap C) + P(M \cap D)}$$

$$= \frac{P(M/A) \cdot P(A)}{P(M/A) \cdot P(A) + P(M/B) \cdot P(B) + P(M/C) \cdot P(C) + P(M/D) \cdot P(D)}$$

$$= \frac{(0,35)(0,25)}{(0,35)(0,25) + (0,25)(0,25) + (0,20)(0,25) + (0,15)(0,25)}$$

$$= 0,37$$

La probabilidad de que la mujer sea administrativa es de 0,37 .

b) $$\begin{aligned} P(M^c) &= 1 - P(M) \\ &= 1 - 0,2375 \\ &= 0,7625 \end{aligned}$$

La probabilidad de que el individuo sea un hombre es de 0,7625 .

Ejercicios

1) La policía planea reforzar el respeto a los límites de velocidad mediante la utilización de sistemas de radar en cuatro diferentes sitios dentro de la ciudad. Los sistemas de radar en cada sitio L_1, L_2, L_3 y L_4 se ponen a funcionar, respectivamente, el 40 %, 30 %, 20 % y 30 % del tiempo, y si una persona que conduce a gran velocidad rumbo a su trabajo tiene, respectivamente, las probabilidades de 0,2 ; 0,1 ; 0,5 y 0,2 de pasar por alguno de estos sitios y que le multen. ¿Cuál es la probabilidad de que le levanten una multa?.

2) Suponga que se distribuyen pelotas de colores en tres cajas idénticas de la siguiente manera :

	Caja 1	Caja 2	Caja 3
Roja	2	4	3
Blanca	3	1	4
Azul	5	3	3

Una caja se selecciona aleatoriamente, de ella se saca una pelota, también aleatoriamente, y se observa que es roja. ¿Cuál es la probabilidad de que la caja 3 sea la que se escogió?.

3) Tres máquinas A, B y C producen respectivamente 60 %, 30 % y 10 % del número total de artículos de una fábrica. Los porcentajes de desperfectos de producción de estas máquinas son respectivamente 2 %, 3 % y 4 %. Seleccionando un artículo al azar resultó defectuoso. Hallar la probabilidad de que el artículo hubiera sido producido por la máquina C.

4) Una compañía necesita tomar la decisión de patrocinar en la TV uno de los siguientes programas : juegos de futbol(F), una serie del oeste(O) o un programa musical(M). Las probabilidades de que decidan por F, O o M son 0,40 ;0,35 y 0,25 respectivamente. Las probabilidades de que las ganancias aumenten sustancialmente si escogen F, O o M son 0,50 ;0,40 y 0,30 respectivamente. Si las ganancias aumetan sustancialmente, encontrar la probabilidad de que la compañía haya escogido la serie del oeste.

5) Existen tres teorías económicas principales : I, que la inflación va a desaparecer pronto; D, que ocurrirá la depresión, y R, que ocurrirá la recesión. Las probabilidades de que I, D o R ocurran son 0,40 ; 0,35 y 0,25 , respectivamente. Las probabilidades de que las acciones de la Compañía Goldmine tripliquen su valor si ocurre I, D o R son 0,90 ;0,60 y 0,20 respectivamente. Si las acciones triplican su valor, ¿cuál es la probabilidad de que la inflación haya desaparecido?.

6) Tres máquinas A, B y C producen componentes mecánicos similares. A produce el 45 % del total de componentes, B el 30 % y C el 25 %. Para el programa de producción usual, el 8 % de los componentes producidos por A no cumplen con las especificaciones establecidas, para B y C, las cifras correspondientes son 6 % y 3 % , respectivamente; un componente es extraído al azar de la producción total y se encuentra defectuoso. Encontrar la probabilidad de que el componente seleccionado fuera producido por la máquina A.

Solución

1) $\mathrm{M} = \{\text{multa}\}$

$\mathrm{P(M)} = 0,27$

2) $\mathrm{R} = \{\text{roja}\}$ $\mathrm{B} = \{\text{blanca}\}$ $\mathrm{A} = \{\text{azul}\}$

$\mathrm{C_1} = \{\text{caja 1}\}$ $\mathrm{C_2} = \{\text{caja 2}\}$ $\mathrm{C_3} = \{\text{caja 3}\}$

$\mathrm{P(C_3/R)} = \dfrac{3}{10}$

3) $\mathrm{A} = \{\text{máquina A}\}$ $\mathrm{B} = \{\text{máquina B}\}$

$\mathrm{C} = \{\text{máquina C}\}$ $\mathrm{D} = \{\text{artículo defectuoso}\}$

$\mathrm{P(C/D)} = 0,16$

4) $\mathrm{F} = \{\text{juego de fútbol}\}$ $\mathrm{O} = \{\text{serie del oeste}\}$

$\mathrm{G} = \{\text{programa musical}\}$ $\mathrm{G} = \{\text{aumento de ganacias}\}$

$\mathrm{P(O/G)} = 0,35$

5) $\mathrm{I} = \{\text{inflación va a desaparecer}\}$ $\mathrm{D} = \{\text{ocurrirá depresión}\}$

$\mathrm{C} = \{\text{ocurrirá recesión}\}$ $\mathrm{A} = \{\text{acciones triplicadas}\}$

$\mathrm{P(I/A)} = 0,58$

6) $\mathrm{A} = \{\text{máquina A}\}$ $\mathrm{B} = \{\text{máquina B}\}$

$\mathrm{C} = \{\text{máquina C}\}$ $\mathrm{D} = \{\text{artículo defectuoso}\}$

$\mathrm{P(A/D)} = 0,58$

Eventos Independientes

Concepto: Los eventos A y B se dicen independientes si, y sólo si $\mathrm{P(A \cap B) = P(A) \cdot P(B)}$

Teorema: Suponga que $\mathrm{P(A)} \neq 0$ y $\mathrm{P(B)} \neq 0$, entonces A y B independientes implica que ellos no son excluyentes y A , B mutuamente excluyentes implica que ellos no son independientes.

Ejemplos

1) Si dos dados son lanzados una vez y sean los siguientes eventos
$\mathrm{A} = \{\text{la suma es 7}\}$
$\mathrm{B} = \{\text{los dos dados muestran el mismo número}\}$
$\mathrm{C} = \{\text{el primer dado es par}\}$

¿Son A y B , A y C independientes?

$$\mathrm{A} = \{(1,6)\,;(2,5)\,;(3,4)\,;(4,3)\,;(5,2)\,;(6,1)\} \qquad \mathrm{P(A)} = \frac{1}{6}$$

$$\mathrm{B} = \{(1,1)\,;(2,2)\,;(3,3)\,;(4,4)\,;(5,5)\,;(6,6)\} \qquad \mathrm{P(B)} = \frac{1}{6}$$

$$\mathrm{C} = \left\{ \begin{array}{cccccc} (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\} \qquad \mathrm{P(C)} = \frac{1}{2}$$

$$\mathrm{A \cap B} = \emptyset \qquad \Rightarrow \mathrm{P(A \cap B)} = 0$$

$\mathrm{P(A \cap B) \neq P(A) \cdot P(B)}$ A y B no son independientes

$$\mathrm{A \cap C} = \{(2,5)\,;(4,3)\,;(6,1)\} \qquad \Rightarrow \mathrm{P(A \cap C)} = \frac{1}{12}$$

$$\mathrm{P(A) \cdot P(C)} = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

$\mathrm{P(A \cap C) = P(A) \cdot P(C)}$ A y C son independientes

2) Dada la siguiente tabla

	con cáncer(C)	sin cáncer($\mathrm{C^c}$)
fumador(F)	0,5	0,2
no fumador($\mathrm{F^c}$)	0,1	0,2

¿Son F y C eventos independientes?

$\mathrm{P(F \cap C)} = 0,5 \qquad \mathrm{P(F)} = 0,7 \qquad \mathrm{P(C)} = 0,6$

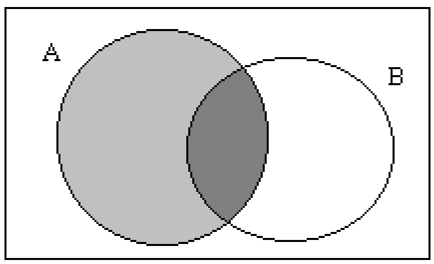
$P(F) \cdot P(C) = (0,7) \cdot (0,6) = 0,42$

$P(F \cap C) \neq P(F) \cdot P(C)$ F y C no son independientes

3) Sabiendo que A y B son eventos independientes, demuestre que:

a) A y B^c son independientes
b) A^c y B son independientes

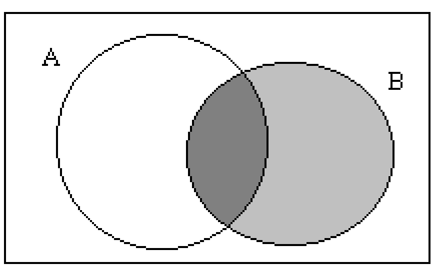
a) A y B independientes si, y sólo si $P(A \cap B) = P(A) \cdot P(B)$



$A = (A \cap B) \cup (A - B)$
$P(A) = P[(A \cap B) \cup (A - B)]$
$P(A) = P(A \cap B) + P(A - B)$
$P(A) = P(A) \cdot P(B) + P(A \cap B^c)$
$P(A) - P(A) \cdot P(B) = P(A \cap B^c)$
$P(A)[1 - P(B)] = P(A \cap B^c)$
$P(A) \cdot P(B^c) = P(A \cap B^c)$

Por lo tanto, si A y B son independientes, entonces A y B^c también lo son.

b)



$B = (A \cap B) \cup (B - A)$
$P(B) = P[(A \cap B) \cup (B - A)]$
$P(B) = P(A \cap B) + P(B - A)$
$P(B) = P(A) \cdot P(B) + P(B \cap A^c)$
$P(B) - P(A) \cdot P(B) = P(B \cap A^c)$
$P(B)[1 - P(A)] = P(B \cap A^c)$
$P(B) \cdot P(A^c) = P(B \cap A^c)$

Por lo tanto, si A y B son independientes, entonces B y A^c también lo son.

Ejercicios

1) Sea el caso de lanzar dos monedas corrientes al aire. Sean los eventos :
A = {todas caras o todas sellos}
B = {aparece una cara}
C = {aparece a lo menos una cara}

a) ¿ Son A y B, A y C, B y C independientes?

2) Se lanzan dos dados. Sean los eventos :
A = {la suma de cinco}
B = {el primer número es impar}
C = {el segundo número es divisible por tres}
D = {la suma es mayor que siete}

¿Cuáles eventos son indepentientes tomados en parejas?

3) Si A y B son eventos independientes, pruebe que A^c y B^c también lo son.

Solución

1) Ninguno es independiente.

2) Sólo son independientes A y B ; A y C ; B y C

3) La demostración es verdadera, es decir, si $P(A \cap B) = P(A) \cdot P(B)$, entonces $P(A^c \cap B^c) = P(A^c) \cdot P(B^c)$

Variables Aleatorias(v.a)

Concepto: una variable aleatoria es una función que asocia un número real a cada elemento del espacio muestral.

Se usarán letras mayúsculas para denotar a una v.a y letras minúsculas para denotar los valores que ella adquiere.

Ejemplos:

1) Se sacan dos pelotas en sucesión, sin reemplazo, de una urna que contiene 4 pelotas rojas y 3 negras. Los resultados posibles y los valores x de la v.a X, donde X es el número de pelotas rojas son:

Espacio muestral	x
RR	2
RN	1
NR	1
NN	0

2) El encargado de un almacén le devuelve tres cascos de seguridad, seleccionados aleatoriamente, a tres obreros del taller, quienes ya se lo habían probado previamente. Suponiendo que el orden de los obreros Pérez, González y Muñoz es el correcto para recibir su casco original, señale los posibles órdenes en que los tres obreros reciben un casco y encuentre los valors m de la v.a M que representa el número de asociaciones correctas.

Espacio muestral	m
PGM	3
PMG	1
MPG	0
MGP	1
GPM	1
GMP	0

En los ejemplos anteriores, el espacio muestral tiene un número finito de elementos.

Conceptos:

1) Si en espacio muestral contiene un número finito de posibilidades o una secuencia interminable con tantos elementos como números naturales existen, entonces se llama ***espacio muestral discreto***.

Los dos ejemplos anteriores corresponden a espacio muestral discreto.

2) Si en espacio muestral contiene un número infinito de posibilidades igual al número de puntos de un segmento de línea, entonces se llama ***espacio muestral continuo***.

Por ejemplo: tiempo necesario para ejecutar una reacción química.

Una v.a se llama ***v.a discreta*** si se puede contar su conjunto de resultados posibles.Una v.a se llama ***v.a continua*** si se puede tomar en una escala continua.

En la mayoría de los problemas prácticos, las v.a continuas representan ***datos medidos*** , tales como alturas, pesos, temperatursa, distancias o períodos de vida; mientras que las v.a discretas representan ***datos que se cuentan*** ,tales como el número de artículos defectuosos de una muestra de k artículos o el número de accidentes por año en una vía rápida en una determinada ciudad.

Distribuciones discretas de probabilidad

Una v.a discreta asume cada uno de sus valores con una cierta probabilidad.

Con mucha frecuencia es conveniente representar con una fórmula todas las probabilidades de una v.a X. Dicha fórmula, necesariamente, debe ser función de los valores numéricos x, y que se representa por $f(x)$, $g(x)$, $h(x)$, etc.Por lo tanto, $f(x) = \mathrm{P}(X = x)$. Al conjunto de pares ordenados $(x, f(x))$ se le llama ***función de probabilidad o distribución de probabilidad*** de la v.a discreta X.

Concepto : El conjunto de pares ordenados $(x, f(x))$ es una ***función de probabilidad, función masa de probabilidad o distribución de probabilidad*** de la v.a discreta X si satisface las siguientes condiciones :

1) $f(x) \geq 0 \qquad \forall\, x \in \mathbb{R}$

2) $\sum\limits_{x} f(x) = 1$

3) $\mathrm{P}(X = x) = f(x)$

Ejemplos

1) Una moneda se lanza dos veces, entonces $\Omega = \{(c,c)\,;(c,s)\,;(s,c)\,;(s,s)\}$. Sea X la v.a que consiste en observar el número de caras.

Espacio muestral	x
$c\,c$	2
$c\,s$	1
$s\,c$	1
$s\,s$	0

$Rec\,X = \{0, 1, 2\}$

La función de probabilidad es:

$f(0) = \mathrm{P}(x = 0) = \dfrac{1}{4}$ $\qquad$ $f(1) = \mathrm{P}(x = 1) = \dfrac{2}{4}$

$f(2) = \mathrm{P}(x = 2) = \dfrac{1}{4}$

x	0	1	2
$\mathrm{P}(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

2) De un lote de 25 artículos de los cuales 5 son defectuosos se eligen 4 al azar. Sea Y la v.a que representa el número de artículos defectuosos encontrados. Obtener la distribución de probabilidades de la v.a Y si los artículos se eligen sin sustitución.

$RecY = \{0, 1, 2, 3, 4\}$

Sea $D =$ artículo defectuoso, por lo tanto, $D^c =$ arículo no defectuoso

$\text{P}(D) = \dfrac{5}{25}$ $\qquad\qquad$ $\text{P}(D^c) = \dfrac{20}{25}$

$$f(0) = \text{P}(y = 0) = \text{P}(D_1^c \cap D_2^c \cap D_3^c \cap D_4^c) = \frac{20}{25} \cdot \frac{19}{24} \cdot \frac{18}{23} \cdot \frac{17}{22} = \frac{4845}{12650}$$

$$f(1) = \text{P}(y = 1) = 4\text{P}(\text{D}_1 \cap D_2^c \cap D_3^c \cap D_4^c) = 4 \cdot \frac{5}{25} \cdot \frac{20}{24} \cdot \frac{19}{23} \cdot \frac{18}{22} = \frac{5700}{12650}$$

$$f(2) = \text{P}(y = 2) = 6\text{P}(\text{D}_1 \cap D_2 \cap D_3^c \cap D_4^c) = 6 \cdot \frac{5}{25} \cdot \frac{4}{24} \cdot \frac{20}{23} \cdot \frac{19}{22} = \frac{1900}{12650}$$

$$f(3) = \text{P}(y = 3) = 4\text{P}(\text{D}_1 \cap D_2 \cap D_3 \cap D_4^c) = 4 \cdot \frac{5}{25} \cdot \frac{4}{24} \cdot \frac{3}{23} \cdot \frac{20}{22} = \frac{200}{12650}$$

$$f(4) = \text{P}(y = 4) = \text{P}(\text{D}_1 \cap D_2 \cap D_3 \cap D_4) = \frac{5}{25} \cdot \frac{4}{24} \cdot \frac{3}{23} \cdot \frac{2}{22} = \frac{5}{12650}$$

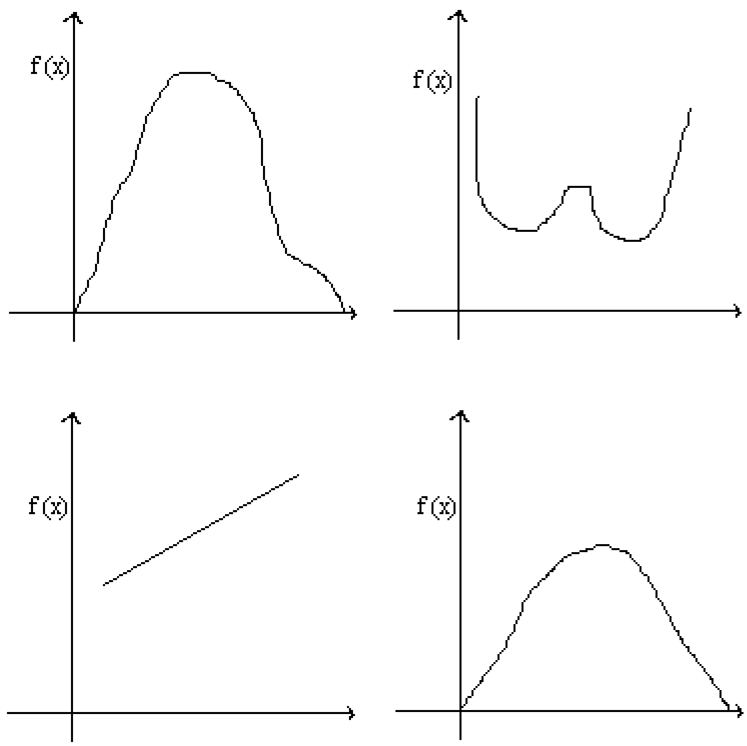
y	0	1	2	3	4
$\text{P}(Y = y)$	$\frac{4845}{12650}$	$\frac{5700}{12650}$	$\frac{1900}{12650}$	$\frac{200}{12650}$	$\frac{5}{12650}$

Distribuciones continuas de probabilidades

Una v.a continua tiene probabilidad cero de asumir cualquiera de sus valores. Luego, su distribución de probabilidad no puede darse en forma tabular.

Como una distribución de probabilidad de una v.a continua no puede presentarse en forma tabular, si puede tener una fórmula. Esta fórmula es una función, es decir, $f(x)$ y para este tipo de variables se llama ***función de densidad de probabilidad o función de densidad*.**

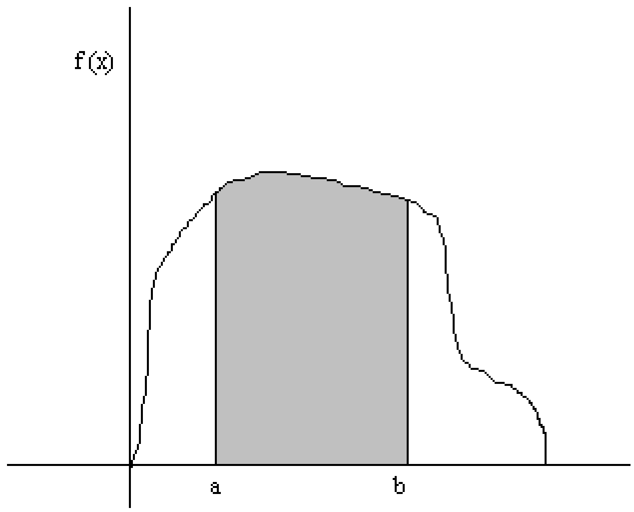
Algunas de las formas de la función de densidad son



Las áreas bajo la curva representarán las probabilidades, por lo tanto, el gráfico de la función de densidad se ubica siempre sobre el eje X

Una función de densidad se construye de tal forma que el área comprendida bajo la curva es siempre igual a uno, cuando se calcula sobre todo el recorrido de la v.a X.

Así P$(a < x < b) = \displaystyle\int_a^b f(x)\,dx$



Concepto : La función $f(x)$ es una ***función de densidad de probabilidad*** para la v.a continua X, definida en el conjunto de los números reales, si:

1) $f(x) \geq 0 \qquad\qquad \forall x \in \mathbb{R}$

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

3) $\mathrm{P}(a < x < b) = \displaystyle\int_{a}^{b} f(x)\,dx$

Ejemplos

1) Suponga que el error en la temperatura de reacción, en grados celcius, para un experimento controlado de laboratorio es una v.a continua Y, que tiene función de densidad

$$f(y) = \begin{cases} \dfrac{y^2}{3} & -1 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Muestre que cumple las dos primeras condiciones de una función de densidad y además determine $\mathrm{P}(0 < y < 1)$

$a)\ \dfrac{y^2}{3} \geq 0$

$$b)\int_{-\infty}^{+\infty} f(y)\,dy = \int_{-\infty}^{-1} f(y)\,dy + \int_{-1}^{2} f(y)\,dy + \int_{2}^{+\infty} f(y)\,dy$$

$$\int_{-\infty}^{+\infty} f(y)\,dy\ = \int_{-\infty}^{-1} 0\,dy + \int_{-1}^{2} \frac{y^2}{3}\,dy + \int_{2}^{+\infty} 0\,dy$$

$$\int_{-\infty}^{+\infty} f(y)\,dy = \left.\frac{y^3}{9}\right|_{-1}^{2}$$

$$\int_{-\infty}^{+\infty} f(y)\,dy = \frac{8}{9} + \frac{1}{9}$$

$$\int_{-\infty}^{+\infty} f(y)\,dy = 1$$

Por lo tanto $f(y)$ cumple con las dos primeras condiciones de una función de densidad.

$$\mathrm{P}(0 < y < 1) = \int_{0}^{1} \frac{y^2}{3}\,dy = \left.\frac{y^3}{9}\right|_{0}^{1} = \frac{1}{9}$$

2) Para la función $f(t) = \begin{cases} e^{-t} & 0 < t < 20 \\ 0 & \text{en otro caso} \end{cases}$ Determine:

a) $\mathrm{P}(t \leq 3)$ b) $\mathrm{P}(1 < t \leq 4)$ c) $\mathrm{P}(t = 1)$

$$a)\,\mathrm{P}(t \leq 3) = \int_{-\infty}^{3} f(t)\,dt$$

$$= \int_{-\infty}^{0} 0\,dt + \int_{0}^{3} e^{-t}\,dt$$

$$= 0 - \left. e^{-t}\right|_{0}^{3}$$

$$= -e^{-3} + 1$$

$$= \frac{e^3 - 1}{e^3}$$

$$b)\,\mathrm{P}(1 < t \leq 4) = \int_{1}^{4} f(t)\,dt$$

$$= \int_{1}^{4} e^{-t}\,dt$$

$$= -\left. e^{-t}\right|_{1}^{4}$$

$$= -e^{-4} + e^{-1}$$
$$= \frac{e^3 - 1}{e^4}$$

$c)$ P$(t = 1) = \int_1^1 f(t)\,dt = 0$

Ejercicios

1) De una caja que contiene 4 monedas de \$ 100 y 2 de \$ 50 , se seleccionan tres de ellas al azar sin reemplazo. Determine la distribución de probabilidad para el total T de las tres monedas.

2) De una caja que contiene 4 pelotas negras y 2 verdes, se seleccionan 3 de ellas en sucesión con reemplazo. Encuentre la distribución de probabilidad para el número de pelotas verdes.

3) Una v.a. X continua tiene función de densidad :

$$f(x) = \begin{cases} \frac{1}{2} & 2 < x < 3 \\ 0 & \text{e.o.c.} \end{cases}$$

Encuentre :
a) P$(2 < x < 2{,}5)$

b) P$(x \leq 1{,}6)$

4) Una v.a. Y continua tiene función de densidad :

$$f(y) = \begin{cases} \frac{2(1+y)}{27} & 2 < y < 5 \\ 0 & \text{e.o.c.} \end{cases}$$

Encuentre :
a)P$(y < 4)$

b) P$(3 < y < 4)$

Solución

1) T = total de las tres monedas A = moneda de \$100 B = moneda de \$50

Monedas	t
AAA	300
AAB	250
ABA	250
BAA	250
ABB	200
BAB	200
BBA	200
BBB	150

t	150	200	250	300
$f(t)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

2) X = número de pelotas verdes V = pelota verde N = pelota negra

Monedas	t
VVV	3
VVN	2
VNV	2
NVV	2
VNN	1
NVN	1
NNV	1
NNN	0

x	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

3) a) $P(2 < x < 2{,}5) = 0,25$

b) $P(x \leq 1{,}6) = 0$

4) a) $P(y < 4) = \dfrac{16}{27}$

b) $P(3 < y < 4) = \dfrac{1}{3}$

Esperanza o valor esperado

El valor esperado se usa como una medida de centro de una distribución de probabilidad de una v.a

Concepto

Sea X una v.a con función de probabilidad o función de densidad $f(x)$. Sea $g(X)$ una función de la v.a X . El valor esperado de X, simbolizado por $E(X)$ es:

$$E[g(X)] = \begin{cases} \sum_x h(x) \cdot f(x) & \text{si } X \text{ es una v.a discreta} \\ \int_{-\infty}^{+\infty} h(x) \cdot f(x)\, dx & \text{si } X \text{ es una v.a continua} \end{cases}$$

Observaciones

1) Si $g(X) = X$, entonces se está calculando la esperanza de la v.a X

$$E(X) = \begin{cases} \sum_x x \cdot f(x) & \text{si } X \text{ es una v.a discreta} \\ \int_{-\infty}^{+\infty} x \cdot f(x)\, dx & \text{si } X \text{ es una v.a continua} \end{cases}$$

$E(X) = \mu$

2) Si $g(x) = (X-\mu)^2$, entonces $E[g(x)]$ se llama ***varianza*** de la v.a X y se simboliza como σ^2

$$\sigma^2 = E[g(X)] = \begin{cases} \sum_x (X-\mu)^2 \cdot f(x) & \text{si } X \text{ es una v.a discreta} \\ \int_{-\infty}^{+\infty} (X-\mu)^2 \cdot f(x)\, dx & \text{si } X \text{ es una v.a continua} \end{cases}$$

$\sigma = \sqrt{Var(X)}$ se conoce como desviación estándar.

σ mide la dispersión de los valores de la v.a X con respecto a su media (μ)

Propiedades de la esperanza

Sea X una v.a, entonces

1) $E(c) = c$

2) $E(cX) = c \cdot E(X)$

3) $E[h(X) \pm g(X)] = E[h(X)] \pm E[g(X)]$

Usando las propiedades de la esperanza es posible determinar una forma más simple para calcular $\sigma^2 = Var(X)$

$$\begin{aligned}\sigma^2 &= E(X-\mu)^2\\ &= E(X^2 - 2X\mu + \mu^2)\\ &= E(X^2) - E(2X\mu) + E(\mu^2)\\ &= E(X^2) - 2\mu E(X) + \mu^2 \qquad \text{pero } E(X)=\mu\\ &= E(X^2) - 2E(X)E(X) + [E(X)]^2\\ &= E(X^2) - [E(X)]^2\end{aligned}$$

Luego, $Var(X) = \sigma^2 = E(X^2) - [E(X)]^2$

Así,

$$Var(X) = \begin{cases} \displaystyle\sum_x x^2 \cdot f(x) - \left[\sum_x x \cdot f(x)\right]^2 & \text{si } X \text{ es una v.a discreta} \\ \displaystyle\int_{-\infty}^{+\infty} x^2 \cdot f(x) - \left[\int_{-\infty}^{+\infty} x \cdot f(x)\right]^2 & \text{si } X \text{ es una v.a continua} \end{cases}$$

Propiedades de varianza

Sea X una v.a, entonces

1) $Var(c) = 0$

2) $Var(cX) = c^2 \cdot Var(X)$

3) $Var(X + a) = Var(X)$

Ejemplos:

1) Se lanza una moneda tres veces, si las tres veces aparece cara o parece sello un jugador gana \$5, pero si no es así pierde \$3.¿Cuál es la esperanza de este juego?

Sea X la v.a que denota ganancia o pérdida

$\Omega = \{(c,c,c)\,;(c,s,c)\,;(c,c,s);(s,c,c);(s,s,c)\,;(s,c,s)\,;(c,s,s)\,;(s,s,s)\}$

x	5	-3
$f(x)$	$\frac{2}{8}$	$\frac{6}{8}$

$$E(X) = 5\left(\frac{1}{4}\right) - 3\left(\frac{3}{4}\right) = -1$$

El jugador pierde, en promedio \$1 por lanzamiento de las tres monedas.

2) Sea Y la v.a que representa la vida en horas de un cierto dispositivo electrónico. La función de densidad es

$$f(Y) = \begin{cases} \dfrac{20000}{Y^3} & y > 100 \\ 0 & \text{e.o.c.} \end{cases}$$ Encuentre la vida esperada de este dispositivo.

$$E(Y) = \int_{-\infty}^{+\infty} Y \cdot f(Y)\, dY = \int_{100}^{+\infty} Y \cdot \frac{20000}{Y^3}\, dy$$

$$= \lim_{b \to +\infty} \int_{100}^{b} 20000 Y^{-2} dy$$

$$= \lim_{b \to +\infty} -\frac{20000}{Y}\Big|_{100}^{b}$$

$$= \lim_{b \to +\infty} -\frac{20000}{b} + \frac{20000}{100}$$

$$= 200$$

La duración promedio de este dispositivo es de 200 horas.

3) Las ventas por hora de una máquina automática puede ser 20 , 21 o 22 cajetillas de cigarros con probabilidad 0,3 ; 0,5 y 0,2 respectivamente . ¿Cuál es la venta esperada por hora para esta máquina? ¿Cuál es la varianza de ventas por hora ?

$X =$ ventas por hora de cigarrillos

x	20	21	22
$f(x)$	$0,3$	$0,5$	$0,2$

$E(X) = (20)(0,3) + (21)(0,5) + (22)(0,2)$
$E(X) = 20,9$

La venta esperada por hora es de 20,9 cajetillas.

$Var(X) = E(X^2) - [E(X)]^2$

$E(X^2) = (400)(0,3) + (441)(0,5) + (484)(0,2)$
$E(X^2) = 437,3$

$Var(X) = 437,3 - 436,81$
$Var(X) = 0,49$

La varianza de ventas por hora es de 0,49 cajetillas^2.

4) Sea V la velocidad del viento, em Km/hr.,y suponga que V tiene función de densidad

$$f(V) = \begin{cases} \dfrac{1}{10} & 0 < v < 10 \\ 0 & \text{e.o.c.} \end{cases}$$

La presión W en libras/pie^2, sobre la superficie del ala de un aeroplano está dada por la relación: $W = 0,003V^2$. Determine el valor esperado y la varianza de la presión.

$$\begin{aligned} E(W) &= E(0,003V^2) \\ &= 0,003E(V^2) \\ &= 0,003\int_0^{10} V^2 \cdot \frac{1}{10}\, dV \\ &= 0,003 \cdot \frac{1}{10} \cdot \frac{V^3}{3}\Big|_0^{10} \\ &= \frac{1}{10} \end{aligned}$$

La presión promedio es de 0,1 libra/pie^2.

$$Var(W) = E(W^2) - [E(W)]^2$$

$$\begin{aligned} E(W^2) &= E[(0,003)^2 V^4] \\ &= (0,003)^2 \int_0^{10} V^4 \cdot \frac{1}{10}\, dV \\ &= (0,003) \cdot \frac{1}{10} \cdot \frac{V^5}{5}\Big|_0^{10} \\ &= 0,018 \end{aligned}$$

$$Var(W) = 0,018 - 0,01$$
$$Var(W) = 0,008$$

La varianza es de 0,008 (libra/pie^2)^2

5) Sea X una v.a con $\mu = 5$ y $\sigma^2 = 9$. Calcule el valor esperado de la v.a $Y = \frac{1}{3}(X-5)$ y la varianza

$E(X) = 5$ $\qquad Var(X) = 9$

$$\begin{aligned} E\left[\frac{1}{3}(X-5)\right] &= \frac{1}{3} \cdot E(X) - E\left(\frac{5}{3}\right) \\ &= \frac{1}{3} \cdot 5 - \frac{5}{3} \\ &= 0 \end{aligned}$$

$$Var\left[\frac{1}{3}(X-5)\right] = \left(\frac{1}{3}\right)^2 \cdot Var(X)$$

$$= \frac{1}{9} \cdot 9$$

$$= 1$$

6) Suponga que el número de autos Z, que pasan a través de una máquina lavadora, entre las 4:00 P.M. y las 5:00 P.M. de un viernes, tiene la siguiente distribución de probabilidades

z	4	5	6	7	8	9
$f(z)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Sea $g(Z) = 2Z - 1$ que representa la cantidad de dinero, en dólares, que el gerente del negocio le paga al encargado. Encuentre las ganancias esperadas del encargado en este período en particular.

$$E(Z) = (4)\left(\frac{1}{12}\right) + (5)\left(\frac{1}{12}\right) + (6)\left(\frac{1}{4}\right) + (7)\left(\frac{1}{4}\right) + (8)\left(\frac{1}{6}\right) + (9)\left(\frac{1}{6}\right)$$

$$E(Z) = \frac{41}{6}$$

$$E(2Z-1) = 2E(Z) - 1$$

$$= 2 \cdot \frac{41}{6} - 1$$

$$= \frac{38}{3}$$

La ganancia esperada del encargado es 12,67 dólares estre las 4:00 P.M. y las 5:00 P.M.

7) Sea X una v.a con función de densidad

$$f(X) = \begin{cases} \dfrac{X^2}{3} & -1 < x < 2 \\ 0 & \text{e.o.c.} \end{cases}$$

Encuentre el valor esperado de g$(X) = 4X + 3$

$$E(X) = \int_{-1}^{2} X \cdot \frac{X^2}{3} dX$$

$$E(X) = \frac{1}{3} \cdot \frac{1}{4} X^4 \Big|_{-1}^{2}$$

$$E(X) = \frac{15}{12}$$

$$E(4X+3) = 4E(X)+3$$

$$= 4\cdot\frac{15}{12}+3$$

$$= 8$$

8) Calcule la varianza de $h(X) = 2X+3$ donde X es una v.a con distribución de probabilidad

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

$$E(X) = (0)\left(\frac{1}{4}\right)+(1)\left(\frac{1}{8}\right)+(2)\left(\frac{1}{2}\right)+(3)\left(\frac{1}{8}\right)$$

$$E(X) = \frac{3}{2}$$

$$Var(X) = E(X^2)-[E(X)]^2$$

$$E(X^2) = (0)\left(\frac{1}{4}\right)+(1)\left(\frac{1}{8}\right)+(4)\left(\frac{1}{2}\right)+(9)\left(\frac{1}{8}\right)$$

$$E(X^2) = \frac{13}{4}$$

$$Var(X) = \frac{13}{4}-\frac{9}{4}$$

$$Var(X) = \frac{5}{4}$$

$$Var(2X+3) = 4\cdot Var(X)$$

$$= 4\cdot\frac{5}{4}$$

$$= 5$$

Ejercicios

1) Por invertir en unas acciones en particular, una persona puede obtener ganancias de \$ 4.000 con una probabilidad de 0,3; o una pérdida de \$ 1.000 con una probabilidad de 0,7. ¿Cuál es la ganancia que espera esta persona?.

2) Suponga que un distribuidor de joyas antiguas está interesado en comprar un collar de oro para el cual las probabilidades son 0,22; 0,36; 0,28 y 0,14 respectivamente, de que la poseedora estaría dispuesta a venderla en \$ 250.000, en \$ 150.000, al costo(\$100.000) o con una pérdida de \$ 150.000. ¿Cuál es la utilidad que ella espera?.

3) Si la utilidad de un distribuidor , en unidades de \$ 1.000, en un nuevo automóvil puede considerarse como una v.a. X con función de densidad :

$$f(X) = \begin{cases} 2(1-X) & 0 < x < 1 \\ 0 & \text{e.o.c.} \end{cases}$$

Encuentre la utilidad promedio por automóvil.

4) La función de densidad de la v.a. Y, el número total de horas, en unidades de 100 horas, de que una familia utilice una aspiradora durante un año es :

$$f(Y) = \begin{cases} Y & 0 < y < 1 \\ 2-Y & 1 \leq y < 2 \\ 0 & \text{e.o.c.} \end{cases}$$

Encuentre el número promedio de horas por año que la familia utiliza la aspiradora.

5) Si X representa el resultado cuando se lanza un dado balanceado. Encuentre la esperanza y la varianza de la variable g(X) $= 3\text{X}^2 + 4$

6) Una v.a. continua Z tiene función de densidad :

$$f(Z) = \begin{cases} e^{-Z} & z > 0 \\ 0 & \text{e.o.c.} \end{cases}$$

Encuentre el valor esperado y la varianza de h(Z) $= 2Z - 3$

Solución

1) Esta persona espera una ganancia de \$500 .

2) Espera una utilidad de \$116000 .

3) La utilidad promedio del automóvil es \$667 .

4) La familia utiliza la aspiradora , en promedio, 100 horas .

5) $E[\,g(x)\,] = \dfrac{99}{2}$ $\qquad Var[\,g(x)\,] = \dfrac{16107}{2}$

6) $E[\,h(Z)] = \,-1$ $\qquad Var[\,h(Z)] = 4$

Distribuciones discretas de probabilidad

El comportamiento de una v.a queda descrito por su distribución de probabilidad.

1) Distribución Bernoulli

El experimento más sencillo es aquel que puede resultar en uno de dos resultados posibles.

Ejemplo

a) aprobar o reprobar una asignatura
b) obtener cara o sello al lanzar una moneda
c) sexo de un niño al nacer

El experimento con dos resultados posibles se denomina ensayo Bernoulli

Cualquier experimento puede usarse para definir un ensayo Bernoulli, simplemente denotando algún evento A como éxito y su complemento A^c como fracaso.

La distribución de probabilidad para un ensayo Bernoulli depende sólo de un parámetro p, probabilidad de éxito, y entonces $1-p$ es la probabilidad de fracaso $(1-p=q)$, donde $0<p<1$

Concepto: Sea Ω el espacio muestral de un experimento, sea $A \subseteq \Omega$ cualquier evento con $P(A)=p$, $0<p<1$ y sea X la v.a definida por

$$X(x)=\begin{cases}1 & x\in A\\ 0 & x\in A^c\end{cases}$$

Entonces X se llama v.a Bernoulli con parámetro p.

La distribución de probabilidad de una v.a Bernoulli es de la siguiente forma

$P(x=1)=P(A)=p$
$P(x=0)=P(A^c)=1-p=q$

x	1	0
$f(x)$	p	q

la cual se puede resumir de la siguiente forma

$\boxed{f(x)=p^x \cdot q^{1-x} \quad x=0,1}$ y se denota $X \sim$ Bernoulli (p)

El símbolo $\sim$ denota distribución

El proceso Bernoulli debe tener las siguientes propiedades:

a) El experimento consiste en n intentos repetidos.

b) Los resultados de cada uno de los intentos pueden clasificarse como un éxito o un fracaso.

c) La probabilidad de éxito, p, permanece constante para todos los intentos.

d) Los intentos repetidos son independientes .

e) $E(X) = p$, $Var(X) = p \cdot q$

2) Distribución Binomial

Concepto: un experimento que consiste de n ensayos Bernoulli independientes, cada uno con probabilidad de éxito p , se llama experimento binomial con n ensayos y parámetro p .

Ensayos independientes indica que los ensayos son eventos independientes, esto es, lo que ocurre en un ensayo no influye en el resultado de cualquier otro ensayo.

El espacio muestral para un experimento binomial es el producto cartesiano de los espacios muestrales de los ensayos Bernoulli consigo mismo n veces

$\Omega = \Omega_1 \times \Omega_2 \times \Omega_3 \times ... \times \Omega_n$ donde $\Omega_i = \{\text{éxito}(E), \text{fracaso(F)}\} \forall\, i = 1, 2, ..., n$

Cada elemento de Ω es una $n-$ upla, $w = (w_1, w_2, w_3, ..., w_n)$ donde $w_i =$ E o F
Luego $\mathrm{P}_i(\mathrm{E}) = p$, $\mathrm{P}_i(\mathrm{F}) = 1 - p = q$ $\qquad \forall\, i = 1, 2, ..., n$

Concepto: Sea X el número total de éxitos en un experimento binomial con n ensayos y parámetro p .Entonces X se llama v.a binomial con parámetro n y p .Luego $X \sim b(n, p)$ y su distribución de probabilidades es:

$$\boxed{\boldsymbol{f(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} \qquad x = 0, 1, ..., n}}$$

$$\boxed{\boldsymbol{f(x) = \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot q^{n-x} \qquad x = 0, 1, ..., n}}$$

$E(X) = n \cdot p$, $Var(X) = n \cdot p \cdot q$

Ejemplos

1) Cinco dados son lanzados una vez
a) ¿Cuál es la probabilidad de obtener al menos un tres?
b) ¿Cuál es la probabilidad de obtener al menos dos tres?

$n = 5$
X es la v.a que denota el número de tres al lanzar cinco dados
$X = \{0, 1, 2, 3, 4, 5\}$
$\mathrm{P(E)} = \mathrm{P}(\text{obtener un número tres}) = \frac{1}{6}$
$\mathrm{P(F)} = \mathrm{P}(\text{no obtener un número tres}) = \frac{5}{6}$

$X \sim b\left(5, \frac{1}{6}\right)$

$$f(x) = \binom{5}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{5-x} \qquad x = 0, 1, 2, 3, 4, 5$$

$a)\ \mathrm{P}(x \geq 1) = 1 - \mathrm{P}(x = 0)$

$$\mathrm{P}(x = 0) = \binom{5}{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{5-0} = \left(\frac{5}{6}\right)^{5}$$

$$\mathrm{P}(x \geq 1) = 1 - \left(\frac{5}{6}\right)^{5} = 0,5981$$

La probabilidad de obtener al menos un tres es de 0,5981 .

$b)\ \mathrm{P}(x \geq 2) = 1 - \mathrm{P}(x < 2) = 1 - \mathrm{P}(x = 0) - \mathrm{P}(x = 1)$

$$\mathrm{P}(x = 1) = \binom{5}{1}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{5-1} = 0,4018$$

$$\mathrm{P}(x \geq 2) = 0,5981 - 0,4018 = 0,19$$

La probabilidad de obtener al menos dos tres es de 0,19 .

2) La probabilidad de que una cierta clase de componente pase con éxito una determinada prueba de impacto es $3/4$. Encuentre la probabilidad de que exactamente dos de los siguientes cuatro componentes que se prueben pasen la prueba.

$n = 4$
$X =$ pasar con éxito la prueba de impacto
$X = \{0, 1, 2, 3, 4\}$

$$p = \frac{3}{4} \qquad q = \frac{1}{4}$$

$$X \sim b\left(4, \frac{3}{4}\right)$$

$$f(x) = \binom{4}{x}\left(\frac{3}{4}\right)^{x}\left(\frac{1}{4}\right)^{4-x} \qquad x = 0, 1, 2, 3, 4$$

$$\mathrm{P}(x = 2) = \binom{4}{2}\left(\frac{3}{4}\right)^{2}\left(\frac{1}{4}\right)^{2} = \frac{27}{128} = 0,2109$$

La probabilidad de que exactamente dos de las siguientes piezas cuatro componentes que se prueben pasen la prueba es de 0,2109 .

3) La probabilidad de que un paciente se recupere de una cierta enfermedad a la sangre es 0,4 . Si se sabe que 15 personas han contraído esta enfermedad.

a) ¿Cuál es la probabilidad de que al menos 10 sobrevivan?

b) ¿Cuál es la probabilidad de que sobrevivan entre 3 y 8 personas?

c) ¿Cuál es la probabilidad de que sobrevivan 5 personas?

$X =$ persona que sobreviva a la enfermedad

$X = \{0, 1, 2, 3, .., 15\}$

$p = 0,4 \qquad\qquad q = 0,6$

$X \sim b(15;\, 0,4)$

$$f(x) = \binom{15}{x}(0,4)^x(0,6)^{15-x}$$

$a)$ $\mathrm{P}(x \geq 10) = \mathrm{P}(x = 10) + \mathrm{P}(x = 11) + \mathrm{P}(x = 12) + \mathrm{P}(x = 13) + \mathrm{P}(x = 14) + \mathrm{P}(x = 15)$
$= 0,0338$

La probabilidad de que al menos 10 sobrevivan es de 0,0338 .

$b)$ $\mathrm{P}(3 \leq x \leq 8) = \mathrm{P}(x = 3) + \mathrm{P}(x = 4) + \mathrm{P}(x = 5) + \mathrm{P}(x = 6) + \mathrm{P}(x = 7) + \mathrm{P}(x = 8)$
$= 0,8779$

La probabilidad de que sobrevivan tres y ocho personas es de 0,8779 .

$$c)\ \mathrm{P}(x = 5) = \binom{15}{5}(0,4)^5(0,6)^{10} = 0,1859$$

La probabilidad de que sobrevivan cinco personas es de 0,1859

4) Se sabe que el 30 % de las piezas defectuosas de un proceso de manufactura pueden quedar bien mediante un trabajo de reprocesado.

a) ¿Cuál es la probabilidad de que en un lote de seis piezas defectuosas se puedan reprocesar por lo menos tres de ellas?

b) ¿Cuál es la probabilidad de que ninguna de ellas se puedan reprocesar?

c) ¿Cuál es la probabilidad de que todas ellas se puedan reprocesar?

$X =$ piezas reprocesadas

$X = \{0, 1, 2, 3, 4, 5, 6\}$

$p = 0,3 \qquad\qquad q = 0,7$

$X \sim b(6;\, 0,3)$

$$f(x) = \binom{6}{x}(0,3)^x(0,7)^{6-x}$$

$a)$ $\mathrm{P}(x \geq 3) = \mathrm{P}(x = 3) + \mathrm{P}(x = 4) + \mathrm{P}(x = 5) + \mathrm{P}(x = 6) = 0,2557$

La probabilidad de que se puedan reprocesar al menos tres piezas es de 0,2557 .

$b)\ \mathrm{P}(x=0) = \binom{6}{0}(0,3)^{0}(0,7)^{6} = \ 0,1176$

La probabilidad de que ninguna de las piezas se pueda reprocesar es de 0,1176 .

$c)\ \mathrm{P}(x \geq 1) = 1 - \mathrm{P}(x=0) = 1 - 0,1176 = 0,8823$

La probabilidad de que todas las piezas se pueda reprocesar es de 0,8823 .

Ejercicios

1) Al probar una cierta clase de neumático para camión en un terreno escabroso se encontró que el 25 % de los camiones terminaban la prueba con los neumáticos dañados. De los siguientes 6 camiones probados, encuentre la probabilidad de que :

a) de 3 a 6 tengan los neumáticos dañados.
b) Menos de 2 tengan los neumáticos dañados.
c) más de cinco tengan los neumáticos dañados.

2) La probabilidad de que un paciente se recupere de una delicada operación de corazón es 0,9 . ¿Cuál es la probabilidad de que exactamente 5 de los próximos 7 pacientes que se sometan a esta intervención sobrevivan?.

3) Un ingeniero de control de tráfico reporta que el 75 % de los vehículos que pasan por un punto de verificación tienen matrículas del estado. ¿Cuál es la probabilidad de que más de 4 de los siguientes 9 vehículos no sean del estado?.

4) Una investigación demostró que el 20 % de los habitantes de una ciudad prefieren un teléfono blanco que cualquier otro. ¿Cuál es la probabilidad de que más de la mitad de los siguientes 8 teléfonos que se instalen en esta cuidad sean de color blanco?.

5) Se sabe que el 40 % de los ratones inyectados con un suero quedan protegidos contra una cierta enfermedad. Si 5 ratones son inyectados, encuentre la probabilidad de que :

a) Ninguno contraiga la enfermedad
b) menos de 2 la contraigan.
c) más de tres la contraigan.

Solución

1)

$a)\,\mathrm{P}(3 \leq x \leq 6) = 0,166$

$b)\,\mathrm{P}(x < 2) = 0,54$

$c)\,\mathrm{P}(x > 5) = 0,002$

$2)\,\mathrm{P}(x = 5) = 0,12$

$3)\,\mathrm{P}(x > 4) = 0,063$

$4)\,\mathrm{P}(x \geq 4) = 0,06$

5)

$a)\,\mathrm{P}(x = 0) = 0,08$

$b)\,\mathrm{P}(x < 2) = 0,34$

$c)\,\mathrm{P}(x > 3) = 0,09$

3) Distribución Hipergeométrica

Tanto la distribución binomial como la distribución hipergeométrica persiguen un mismo objetivo: el número de éxitos en una muestra que contiene n observaciones. Lo que establece una diferencia entre estas dos distribuciones de probabilidad discreta es la forma en que se obtiene la información. Para el caso de la distribución binomial la información de la muestra se toma con reposición de una muestra finita, o sin reposición de una población infinita. Para el modelo hipergeométrico la información de la muestra se toma sin reposición de una población finita. Por lo tanto, la probabilidad de éxito, p ,es constante a lo largo de todas las observaciones de un experimento binomial, en cambio, en una distribución hipergeométrica el resultado de una observación afecta el resultado de las obseravciones previas.

En general, el interés que se tiene es en la probabilidad de seleccionar x éxitos de los k posibles resultados o artículos también considerados éxitos y $n-x$ fracasos de los $N-k$ posibles resultados o artículos también considerados fracasos, cuando una muestra aleatoria de tamaño n se selecciona de N resultados o artículos totales. Esto se conoce como un experimento hipergeométrico.

La función de probabilidad de una v.a X con distribución hipergeométrica es

$$\boxed{\boldsymbol{f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x = 0, 1, 2, \ldots, n}}$$

$X \sim H(N, n, k)$

$$E(X) = n \cdot \frac{k}{N} \qquad\qquad Var(X) = n \cdot \frac{k}{N}\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Ejemplos

1) Un comité compuesto por cinco personas se selecciona aleatoriamente de un grupo formado por tres químicos y cinco físicos. Encuentre la distribución de probabilidad para el número de químicos en el comité.

$X =$ v.a que indica el número de químicos
$X = \{0, 1, 2, 3\}$
$N = 8$ $\qquad n = 5$ $\qquad k = 3$ $\qquad N - k = 5$

$X \sim H(8, 5, 3)$

$$f(x) = \frac{\binom{3}{x}\binom{5}{5-x}}{\binom{8}{5}} \qquad x = 0, 1, 2, 3$$

$$\text{P}(x = 0) = \frac{\binom{3}{0}\binom{5}{5}}{\binom{8}{5}} = \frac{1}{56} \qquad \text{P}(x = 1) = \frac{\binom{3}{1}\binom{5}{4}}{\binom{8}{5}} = \frac{15}{56}$$

$$\text{P}(x = 2) = \frac{\binom{3}{2}\binom{5}{3}}{\binom{8}{5}} = \frac{30}{56} \qquad \text{P}(x = 3) = \frac{\binom{3}{3}\binom{5}{2}}{\binom{8}{5}} = \frac{10}{56}$$

x	0	1	2	3
$f(x)$	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

2) Entre 16 postulantes para un trabajo, 10 tenían un grado universitario. Si tres de los postulantes son elegidos al azar para una entrevista. ¿Cuál es la probabilidad de que:

a) ninguno tenga grado universitario?.
b) exactamente uno tenga grado universitario?.
c) dos tengan grado universitario?.
d) los tres tengan grado universitario?.

$X =$ v.a que indica postulante con grado universiatrio
$X = \{0, 1, 2, 3\}$
$N = 16 \qquad n = 3 \qquad k = 10 \qquad N - k = 6$

$X \sim H(8, 3, 10)$

$$f(x) = \frac{\binom{10}{x}\binom{6}{3-x}}{\binom{16}{3}} \qquad x = 0, 1, 2, 3$$

$$\text{P}(x = 0) = \frac{\binom{10}{0}\binom{6}{3}}{\binom{16}{3}} = \frac{1}{28}$$

La probabilidad de que ninguno tenga grado universitario es de 0,0357 .

$$P(x=1) = \frac{\binom{10}{1}\binom{6}{2}}{\binom{16}{3}} = \frac{15}{56}$$

La probabilidad de que uno tenga grado universitario es de 0,2679 .

$$P(x=2) = \frac{\binom{10}{2}\binom{6}{1}}{\binom{16}{3}} = \frac{27}{56}$$

La probabilidad de que dos tengan grado universitario es de 0,4821 .

$$P(x=3) = \frac{\binom{10}{3}\binom{6}{0}}{\binom{16}{3}} = \frac{3}{14}$$

La probabilidad de que los tres tengan grado universitario es de 0,2143 .

3) Lotes de 40 componentes cada uno se consideran acptables si no contienen más de tres defectuosos. El procedimiento de muestreo del lote consiste en seleccionar 5 componentes aleatoriamente y rechazar el lote si se encuentra un componente defectuoso. ¿Cuál es la probabilidad de que exactamente un defectuoso se encuentre en la muestra si hay tres defectuosos en todo el lote?.

$X =$ artículos defectuosos
$X = \{0, 1, 2, 3\}$
$N = 40$ $n = 5$ $k = 3$ $N - k = 37$

$X \sim H(40, 5, 3)$

$$f(x) = \frac{\binom{3}{x}\binom{37}{5-x}}{\binom{40}{5}}$$

$$P(x=1) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0,3011$$

La probabilidad es de 0,3011 .

Ejercicios

1) Para evitar que lo descubran en la aduana, un viajero ha colocado 6 tabletas de narcótico en una botella que contiene 9 píldoras de vitamina que son similares en apariencia. Si el oficial de la aduana selecciona 3 tabletas aleatoriamente para analizarlas, ¿cuál es la probabilidad de que el viajero sea arrestado por posesión ilegal de narcóticos?.

2) El dueño de una casa planta 6 tallos que selecciona al azar de una caja que contiene 5 tallos de tulipán y 4 de narciso. ¿Cuál es la probabilidad de que plante 2 tallos de narciso y 4 de tulipán?.

3) De un lote de 10 proyectiles, 4 se seleccionan al azar y se disparan. Si el lote contiene 3 proyectiles defectuosos que no explotarán, ¿cuál es la probabilidad de que :

a) los 4 exploten?.

b) al menos 2 no exploten?.

4) ¿Cuál es la probabilidad de que una mesera se rehúse a servir bebidas alcohólicas únicamente a 2 menores de edad, si verifica aleatoriamente sólo 5 identificaciones de entre 9 estudiantes, de los cuales 4 no tienen la edad suficiente?.

5) Una compañía manufacturera utiliza un esquema para aceptación de los artículos producidos antes de ser embarcados. El plan es de dos etapas. Se preparan cajas de 25 para embarque y se selecciona una muestra de tres para verificar si tiene algún artículo defectuoso. Si se encuentra uno, la caja entera se regresa para verificarla al 100 %. Si no se encuentra ningún artículo defectuoso la caja se embarca.

a) ¿Cuál es la probabilidad de que se embarque una caja que contiene 3 artículos defectuosos?.

b) ¿Cuál es la probabilidad de que una caja que contiene sólo un artículo defectuoso regrese para la verificación?.

Solución

$1)\ \mathrm{P}(x \leq 3) = \dfrac{83}{84}$

$2)\ \mathrm{P}(x = 2) = \dfrac{5}{14}$

$3)$

$a)\ \mathrm{P}(x = 0) = \dfrac{1}{6}$

$b)\ \mathrm{P}(x < 2) = \dfrac{3}{10}$

$4)\ \mathrm{P}(x = 2) = \dfrac{10}{21}$

$5)$

$a)\ \mathrm{P}(x = 3) = \dfrac{1}{2300}$

$b)\ \mathrm{P}(x = 1) = \dfrac{693}{2300}$

4) Distribución Poisson

Los experimentos que resultan en valores numéricos de una v.a X y que representan el número de resultados durante un intervalo de tiempo dado o en una región específica frecuentemente se llaman ***experimentos Poisson*** . El intervalo de tiempo dado puede ser de cualquier duración, por ejemplo, un minuto, un día, una semana, un mes o inclusive un año. Por tal motivo un experimento Poisson puede generar observaciones para una cierta v.a X que representen el número de llamadas telefónicas por hora que se recibe en una oficina, el número de días en que una determinada escuela se cierra en invierno debido a la nieve, o al número de juegos pospuestos debido a la lluvia durante una temporada de fútbol.

El número x de resultados que ocurren en un experimento Poisson se llama ***v.a. de Poisson*** . El número promedio de resultados se calcula de la forma $\mu = E(X) = \lambda t$, donde t es el tiempo o región específicos de interés.

La función de distribución es de la forma:

$$f(X) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \qquad x = 0, 1, 2, ...$$

$$f(X) = \frac{e^{-\mu}\,(\mu)^x}{x!}$$

$$X \sim P(\mu)$$

Ejemplos

1) El número promedio de partículas radioactivas que pasan a través de un contador durante un milisegundo en un experimento de laboratorio es 4. ¿Cuál es la probabilidad de que entren 6 partículas al contador en un milisegundo determinado?.

$\mu = 4$ $\qquad$ $X =$ Nº de partículas que entran en el contador.

$$X \sim P(4)$$

$$f(X) = \frac{e^{-4}\,(4)^x}{x!} \qquad x = 0, 1, 2, ...$$

$$\text{P}(x = 6) = \frac{e^{-4}\,(4)^6}{6!} = 0,1042$$

La probabilidad de que entren 6 partículas al contador en un milisegundo determinado es de 0,1042 .

2) Se sabe que 10 es el número promedio de camiones tanque de aceite que llegan por día a una cierta ciudad portuaria. Las instalaciones del puerto pueden atender cuando mucho a 15 camiones tanque en un día. ¿Cuál es la probabilidad de que en un día determinado tengan que regresar los camiones tanque?.

$\mu = 10$ $\quad$ $X =$ Nº de camiones tanque por día.

$X \sim P(10)$

$$f(X) = \frac{e^{-10}\,(10)^x}{x!} \qquad x = 0, 1, 2, ...$$

$$\begin{aligned} \mathrm{P}(x > 15) &= 1 - \mathrm{P}(x \leq 15) \\ &= 1 - [\mathrm{P}(x = 0) + \mathrm{P}(x = 1) + \mathrm{P}(x = 2) + ... + \mathrm{P}(x = 15)] \\ &= 0,00487 \end{aligned}$$

La probabilidad de que en un día determinado tengan que regresar los camiones tanque es de 0,00487 .

3) Suponga que los clientes llegan a una fila de espera a una tasa de 4 por minuto. Suponiendo que el número de personas que llegan a la fila en cualquier intervalo de tiempo dado tiene distribución Poisson.¿Cuál es la probabilidad de que al menos una persona llegue a la fila en un intervalo de $\frac{1}{2}$ minuto?.

4 clientes $\quad$ 1minuto

μ clientes $\quad$ $\frac{1}{2}$ minuto $\quad$ $\Rightarrow \mu = 2$ clientes

$X =$ Nº de clientes que llegan en $\frac{1}{2}$minuto

$X \sim P(2)$

$$f(X) = \frac{e^{-2}\,(2)^x}{x!} \qquad x = 0, 1, 2, ...$$

$$\begin{aligned} \mathrm{P}(x \geq 1) &= 1 - \mathrm{P}(x = 0) \\ &= 1 - \frac{e^{-2}\,(2)^0}{0!} \\ &= 0,8647 \end{aligned}$$

La probabilidad de que al menos una persona llegue a la fila en un intervalo de $\frac{1}{2}$ minuto es de un 0,8647 .

Ejercicios

1) En promedio, en una cierta intersección ocurren 3 accidentes viales por mes. ¿Cuál es la probabilidad de que en un determinado mes en esta intersección :

a) ocurran exactamente 5 accidentes?.

b) ocurran menos de 3 accidentes?.

2) Una cierta área de la ciudad XX es afectada en promedio por 6 huracanes al año. Encuentre la probabilidad de que en un determinado año esta área sea afectada por :

a) menos de 4 huracanes.

b) cualquier cantidad entre 6 y 8 huracanes.

3) El número promedio de ratas de campo por acre en un campo de trigo de 5 acres se estima que es de 12. Encuentre la probabilidad de que menos de 4 ratas de campo se encuentren en este campo de trigo.

4) Un restaurante prepara una ensalada que contiene en promedio 5 verduras diferentes. Encuentre la probabilidad de que la ensalada contenga más de 5 verduras en un determinado día.

5) En un estudio de un inventario se determinó que, en promedio, la demanda por un artículo en particular en una bodega era de 5 veces al día. ¿Cuál es la probabilidad de que en un determinado día este artículo sea requerido :

a) más de cinco veces?.

b) ni una sola vez?.

Solución

1)

$a)$ $\mathrm{P}(x = 5) = 0,1$

$b)$ $\mathrm{P}(x < 3) = 0,42$

2)

$a)$ $\mathrm{P}(x < 4) = 0,15$

$b)$ $\mathrm{P}(6 \leq x \leq 8) = 0,40$

3) $\mathrm{P}(x < 4) = 0,002$

4) $\mathrm{P}(x > 5) = 0,47$

5)

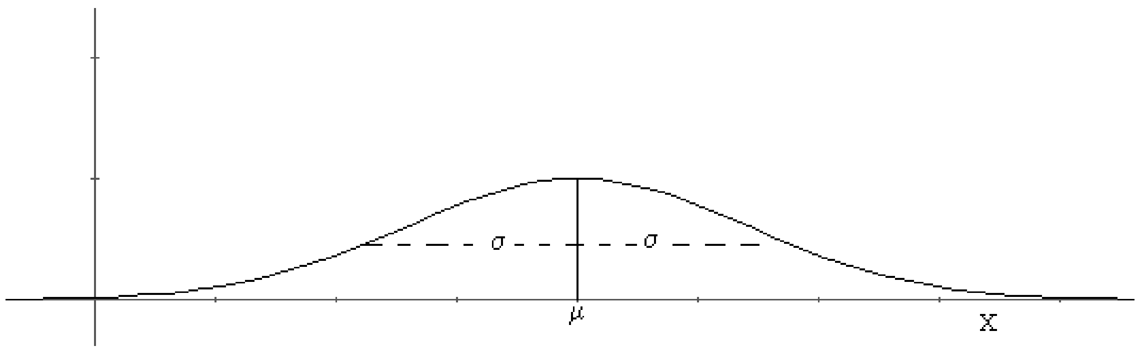
$a)$ $\mathrm{P}(x > 5) = 0,47$

$b)$ $\mathrm{P}(x = 0) = 0,007$

Distribuciones continuas de probabilidad

1) Distribución Normal

Es la distribución continua de probabilidad más importante en el campo de la estadística. Su gráfica recibe el nombre de curva normal, su forma es la de una campana



Esta curva permite describir muchos fenómenos que ocurren en la naturaleza, la industria y la investigación.

Una v.a continua X que tiene distribución en forma de campana se llama v.a. normal.

Concepto: la función de densidad de la v.a normal X , con media μ y varianza σ^2, es:

$$f(X) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

$X \sim N(\mu\,, \sigma^2)$

Propiedades de la curva normal

1) El máximo valor de la curva es en $x = \mu$

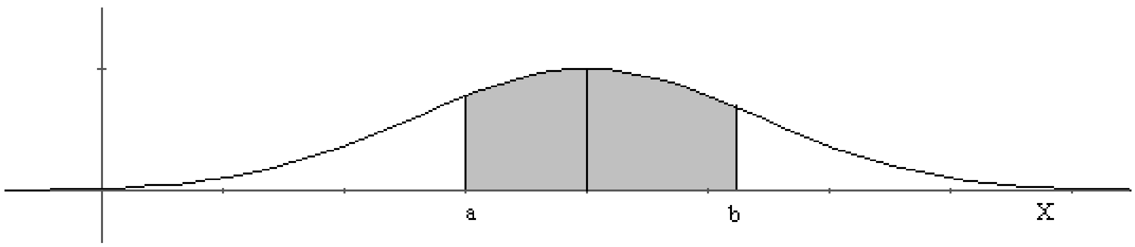
2) La curva es simétrica respecto a la recta $x = \mu$

3) La curva es cóncava hacia arriba en $]-\infty, \mu - \sigma\ [\cup]\ \mu + \sigma,\ +\infty[$ y es cóncava hacia abajo en $]\ \mu - \sigma\,, \mu + \sigma\ [$.

4) La curva es asintótica al eje X.

5) El área bajo la curva y sobre el eje X es uno.

Areas bajo la curva normal



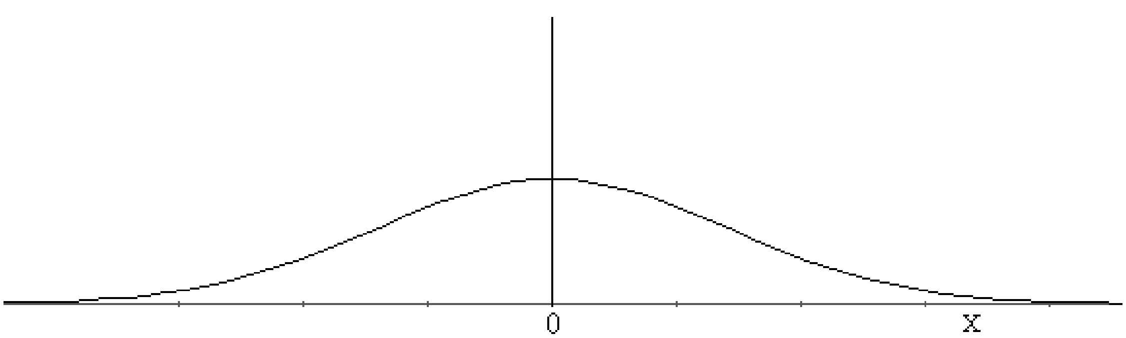
$$P(a < X < b) = \int_a^b f(X)\, dX$$

Sin embargo, resolver esta integral con la función de densidad de la v.a normal no es tan simple. Por tal motivo, se recurre a un proceso denominado ***estandarización*** basándose en una v.a normal z que tiene $\mu = 0$ y $\sigma^2 = 1$ y que se denomina ***distribución normal estándar***

Concepto: Si z es una v.a normal con $\mu = 0$ y $\sigma^2 = 1$, tiene función de densidad:

$$f(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}Z^2} \qquad -\infty < x < \infty$$

$Z \sim N(0,1)$



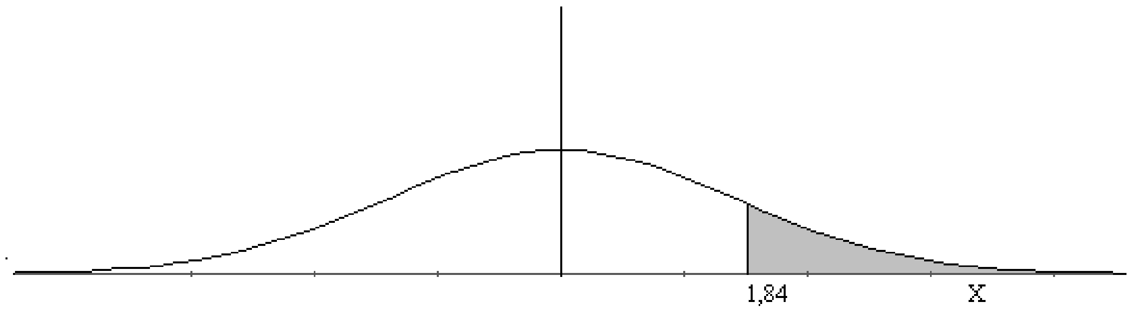
El proceso de estandarización se realiza de la siguiente forma:

Si $X \sim N(\mu\,, \sigma^2)$, entonces $Z = \dfrac{X-\mu}{\sigma}$

Los valores de la v.a normal z se encuentran tabulados

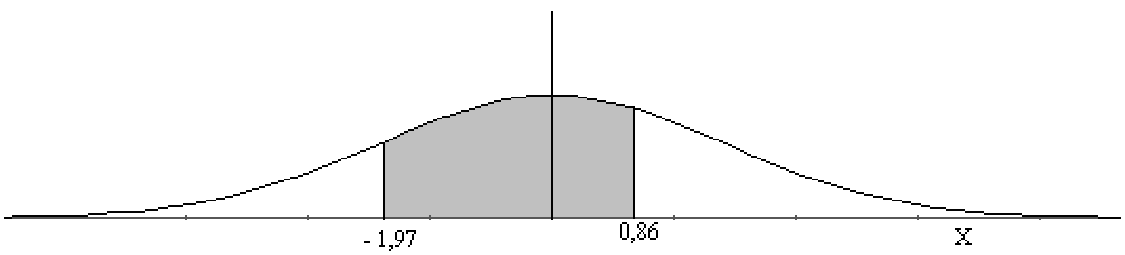
Ejemplos:

1) $\mathrm{P}(z > 1,84)$



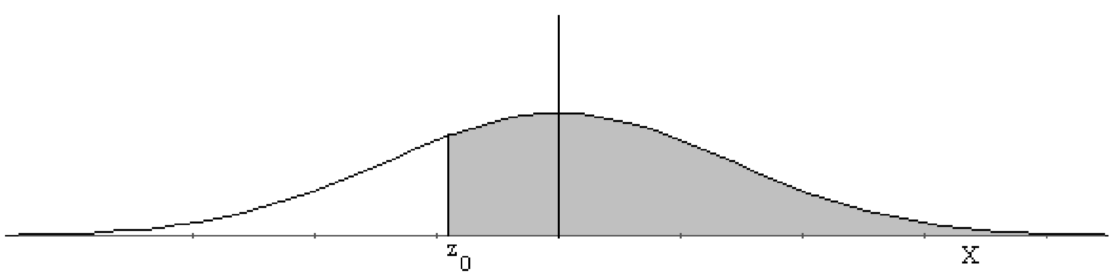
$$\begin{aligned}\mathrm{P}(z > 1,84) &= 1 - \mathrm{P}(z \leq 1,84)\\ &= 1 - 0,9671\\ &= 0,0329\end{aligned}$$

2) $\mathrm{P}(-1,97 < z < 0,86)$



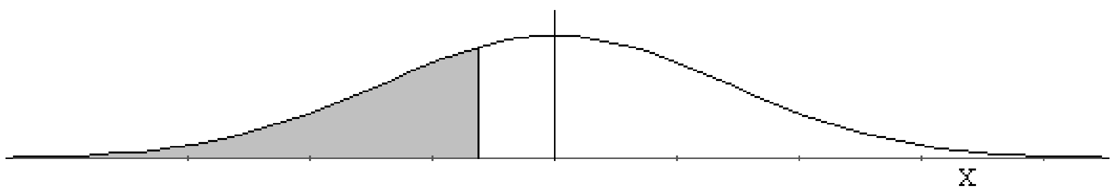
$$\begin{aligned}\mathrm{P}(-1,97 < z < 0,86) &= \mathrm{P}(z < 0,86) - \mathrm{P}(z < -1,97)\\ &= 0,8051 - 0,0244\\ &= 0,7807\end{aligned}$$

3) $\mathrm{P}(z > z_0) = 0,7486$



$\mathrm{P}(z > z_0) = 0,7486$
$1 - \mathrm{P}(z \leq z_0) = 0,7486$
$1 - 0,7486 = \mathrm{P}(z \leq z_0)$
$0,2514 = \mathrm{P}(z \leq z_0) \Rightarrow z_0 = \ -0,67$

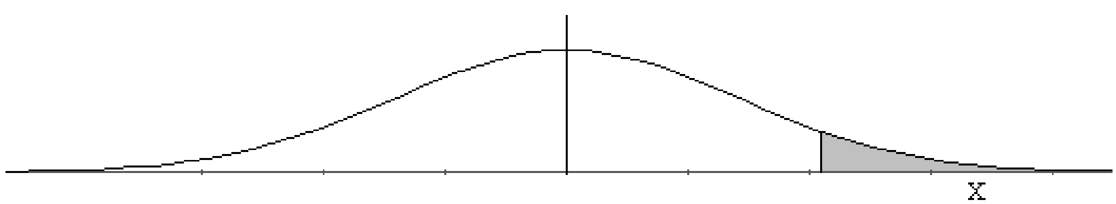
4) Sea X una v.a normal con $\mu = 40$ y $\sigma = 6$, determine
a) $\mathrm{P}(X \leq x) = 0,45$



$$\mathrm{P}\left(z \leq \frac{x-40}{6}\right) = 0,45$$

$$\frac{x-40}{6} = \ -0,13 \Rightarrow x = 39,22$$

b) $\mathrm{P}(X > x) = 0,14$



$$1 - \mathrm{P}\left(z \leq \frac{x-40}{6}\right) = 0,14$$

$$\mathrm{P}\left(z \leq \frac{x-40}{6}\right) = 0,86$$

$$\frac{x-40}{6} = 1,08 \Rightarrow x = 46,48$$

Ejercicios

I) Usando la tabla determine :

a) P$(z < 0,83)$	Resp.:0,7967
b) P$(z < -1,27)$	Resp.:0,1020
c) P$(z > 0,83)$	Resp.:0,2033
d) P$(z > -1,27)$	Resp.:0,898
e) P$(0,47 < z < 1,08)$	Resp.:0,1791
f) P$(\mid z \mid < 1,39)$	Resp.:0,5354
g) P$(z > z_1) = 0,06$	Resp.:$z_1 =$ 1,55
h) P$(-0,93 < z < z_1) = 0,7235$	Resp.:$z_1 =$ 1,28

II) Dada la v.a. X distribuida normalmente con media 18 y desviación estándar 2,5. Encuentre :

a) P$(x < 15)$	Resp.:0,1151
b) P$(x < x_1) = 0,2236$	Resp.:$x_1 =$ 16,1
c) P$(x > x_1) = 0,1814$	Resp.:$x_1 =$ 20,28
d) P$(17 < x < 21)$	Resp.:0,5403

Problemas de aplicación

1) Cierto tipo de batería dura un promedio de tres años, con una desviación estándar de 0,5 años. Suponiendo que las duraciones de las baterías son normalmente distribuidas, encuentre la probabilidad de que una determinada batería dure menos de 2,3 años.

$X \sim N(3\ ;\ 0,25)$ $\qquad$ $X =$ duración de la batería

$$\begin{aligned} \mathrm{P}(x < 2,3) &= \mathrm{P}\left(z < \frac{2,3-3}{0,5}\right) \\ &= \mathrm{P}(z < -1,4) \\ &= 0,0808 \end{aligned}$$

La probabilidad de que una determinada batería dure menos de 2,3 años es de un $0,808$.

2) Una compañía fabrica focos cuya duración es normalmente distribuida con una media de 800 horas y una desviación estándar de 40 horas. Encuentre la probabilidad de que un foco dure entre las 778 y 834 horas de uso.

$X \sim N(800\ , 1600)$ $\qquad$ $X =$ duración de los focos

$$\begin{aligned} \mathrm{P}(778 < x < 834) &= \mathrm{P}\left(\frac{778-800}{40} < z < \frac{834-800}{40}\right) \\ &= \mathrm{P}(-0,55 < z < 0,85) \\ &= \mathrm{P}(z < 0,85) - \mathrm{P}(z < -0,55) \\ &= 0,8023 - 0,2912 \\ &= 0,5111 \end{aligned}$$

La probabilidad de que un foco dure entre las 778 y 834 horas de uso es de un $0,5111$.

3) Una cierta máquina produce resistencias eléctricas que tienen un valor medio de 40 ohms y una desviación stándar de 2 ohms. Suponiendo que los valores de las resistencias siguen una distribución normal y que pueden medirse con cualquier grado de precisión. ¿Qué porcentaje de las resistencias tendrá un valor que exceda a 43 ohms?

$X \sim N(40, 4)$ $\qquad$ $X =$ valor de las resistencias eléctricas

$$\begin{aligned} \mathrm{P}(x > 43) &= 1 - \mathrm{P}\left(z \leq \frac{43-40}{2}\right) \\ &= 1 - \mathrm{P}(z \leq 1,5) \\ &= 1 - 0,9332 \\ &= 0,0668 \end{aligned}$$

El $0,0668$ de las resistencias tendrá un valor que exceda a 43 ohms.

4) En una empresa las edades de los trabajadores se distribuye normalmente con media 50 años y desviación estándar es de 5 años.

a) ¿Qué porcentaje de los trabajadores tiene entre 50 y 52,5 años?

b) ¿Cuál es la probabilidad que un trabajador cualquiera no sea mayor de 45 años?

c) ¿Cuál es la probabilidad que un trabajador tenga entre 41 y 58 años?

d) El 20 % de los trabajadores están bajo cierta edad ¿Cuál es esa edad?

$X \sim N(50, 25)$ $\qquad$ $X =$ edad de los trabajadores

$$
\begin{aligned}
a)\ \mathrm{P}(50 < x < 52,5) &= \mathrm{P}\left(\frac{50-50}{5} < z < \frac{52,5-50}{5}\right) \\
&= \mathrm{P}(0 < z < 0,5) \\
&= \mathrm{P}(z < 0,5) - \mathrm{P}(z < 0) \\
&= 0,6915 - 0,5 \\
&= 0,1915
\end{aligned}
$$

El 19,15 % de los trabajadores tiene entre 50 y 52,5 años.

$$
\begin{aligned}
b)\ \mathrm{P}(x \leq 45) &= \mathrm{P}\left(z \leq \frac{45-50}{5}\right) \\
&= \mathrm{P}(z \leq -1) \\
&= 0,1587
\end{aligned}
$$

La probabilidad que un trabajador cualquiera no sea mayor de 45 años es de $0,1587$.

$$
\begin{aligned}
c)\ \mathrm{P}(41 < x < 58) &= \mathrm{P}\left(\frac{41-50}{5} < z < \frac{58-50}{5}\right) \\
&= \mathrm{P}(-1,8 < z < 1,6) \\
&= \mathrm{P}(z < 1,6) - \mathrm{P}(z < -1,8) \\
&= 0,9452 - 0,0359 \\
&= 0,9093
\end{aligned}
$$

La probabilidad que un trabajador tenga entre 41 y 58 años es de $0,9093$

$d)$ $\mathrm{P}(X \leq x) = 0,20$

$$\mathrm{P}\left(z \leq \frac{x-50}{5}\right) = 0,20$$

$$\frac{x-50}{5} = -0,85 \Rightarrow x = 45,75$$

El 20 % de los trabajadores tiene una edad menor o igual a 45,75 años.

Ejercicios

Resuelva los siguientes problemas :

1) Las piezas de pan de centeno distribuidas a las tiendas locales por una cierta pastelería tienen una longitud promedio de 30 cm. y una desviación estándar de 2 cm. Suponiendo que las longitudes están normalmente distribuidas, ¿qué porcentaje de las piezas son

a) de más de 31,7 cm. de longitud?.
b) entre 29,3 y 33,5 cm. de longitud?.
c) de una longitud menor que 25,5 cm.?.

2) Una máquina despachadora de refrescos está ajustada para servir un promedio de 200 milílitros por vaso. Si la cantidad de refresco es normalmente distribuida con una desviación estándar de 15 milílitros.

a) ¿Qué fracción de los vasos contendrá más de 224 milílitros?.
b) ¿Cuál es la probabilidad de un vaso contenga entre 191 y 209 milílitros?.

3) El diámetro interno ya terminado de un anillo de pistón está normalmente distribuido con una media de 10 cm. y una desviación estándar de 0,03 cm.

a) ¿Qué proporción de los anillos tendrá un diámetro interno que exceda de 10,075 cm.?.
b) ¿Cuál es la probabilidad de que un anillo de pistón tenga un diámetro interno entre 9,97 y 10,03 cm.?.
c) ¿Para que valor el diámetro interno de un anillo de pistón será menor que el 15 %?.

4) La resistencia a la tensión de cierto componente metálico está normalmente distribuida con una media de 10.000 Kg./cm^2 y una desviación estándar de 100 Kg./cm^2.

a) ¿Cuál es la proporción de estos componentes que exceden de 10.150 Kg./cm^2 de resistensia a la tensión?.
b) Si las especificaciones requieren que todos los componentes tengan una resistencia a la tensión entre 9.800 y 10.200 Kg./cm^2 inclusive, ¿qué porcentaje de piezas se esperaría que se desecharan?.

5) La vida promedio de cierto tipo de motor pequeño es de 10 años con una desviación estándar de 2 años. El fabricante repone sin cargo todos los motores que fallen dentro del período de garantía. Si está dispuesto a reponer sólo 3 % de los motores que fallan, ¿qué tan larga deberá ser la garantía que otorgue?. Suponga que las vidas de los motores siguen una distribución normal.

6) Suponga que un consultor está investigando cuánto tiempo necesitarían los obreros de la fábrica para montar cierta pieza en una planta de automóviles Volvo, y determinó que la información(tiempo en segundos) estaba distribuida normalmente con una media de 75 segundos y una desviación estándar de 6 segundos.

a) ¿Cuál es la probabilidad de que un obrero seleccionado aleatoriamente pueda montar la pieza en menos de 75 segundos o en más de 81 segundos?

b) ¿Cuál es la probabilidad de que un obrero seleccionado aleatoriamente pueda montar la pieza de 69 a 81 segundos?.

c) ¿Cuál es la probabilidad de que un obrero seleccionado aleatoriamente pueda montar la pieza en menos de 62 segundos?.

d) ¿Cuál es la probabilidad de que un obrero seleccionado aleatoriamente pueda montar la pieza de 62 a 69 segundos?.

e) ¿Cuántos segundos deben pasar antes de que el 50 % de los obreros monten la pieza?.

f) ¿Cuántos segundos deben pasar antes de que el 10 % de los obreros monten la pieza?.

7) El espesor de un lote de 10.000 arandelas de bronce de un cierto tipo fabricadas por una gran compañía tiene una distribución normal con media 0,0191 pulgadas y desviación estándar 0,000425 pulgadas. Compruebe que se puede esperar que el 99,04 % de estas arandelas tenga un espesor entre 0,0180 y 0,0202 pulgadas.

8) El tiempo de reacción para un cierto experimento psicológico está distribuido normalmente con media 20 segundos y desviación estándar 4 segundos.

a) ¿Cuál es la probabilidad de que una persona tenga un tiempo de reacción entre 14 y 30 segundos?.

b) ¿Cuál es la probabilidad de que una persona tenga un tiempo de reacción entre 25 y 30 segundos?.

c) ¿Qué porcentaje de personas tienen un tiempo de reacción de más de 14 segundos?.

d) ¿Cuál es el tiempo de reacción de modo que sólo el 1 % de todas las personas reaccionen con mayor rapidez?.

9) Un procesador de alimentos envasa café en pequeños tarros, los pesos de los tarros están normalmente distribuidos con una desviación estándar de 0,3 onzas. Si el 5 % de los tarros pesa más de 12,492 onzas. ¿Cuál es el promedio de los tarros?.

Solución

1)

a) El 19,77 % de las piezas tiene una longitud de más de 31,7 cm.

b) El 59,67 % de las piezas tiene una longitud entre 29,3 y 33,5 cm.

c) El 1,22 % de las piezas tiene una longitud de menor a 25,5 cm.

2)

a) El 0,0548 de los vasos contendrá más de 224 milímetros.

b) El 0,4514 de los vasos tendrá entre 191 y 209 milímetros.

3)

a) El 0,0062 de los anillos tendrá un diámetro superior a 10,075 cm.

b) El 0,6826 de los anillos tendrá un diámetro entre 9,97 y 10,03 cm.

c) El 15 % de los anillos tendrá un diámetro de 9,9688 cm.

4)

a) El 0,0668 de los componentes exceden de 10150 Kg/cm^2 de resistencia a la tensión.

b) El 4,56 % de las piezas se desecharán.

5) Debe tener una garantía de a lo más 6,24 años.

6)

a) Existe un 0,6587 de probabilidad que un obrero pueda montar una pieza en menos de 75 seg. o en más de 81 seg.

b) Existe un 0,6826 de probabilidad que un obrero pueda montar una pieza entre 69 y 81 seg.

c) Existe un 0,015 de probabilidad que un obrero pueda montar una pieza en menos de 62 seg.

d) Existe un 0,1437 de probabilidad que un obrero pueda montar una pieza entre 62 y 69 seg.

e) Deben pasar 75 segundos antes de que el 50 % de los obreros monten la pieza.
f) Deben pasar 67,26 segundos antes de que el 10 % de los obreros monten la pieza.

7) Se cumple que el 99,04 % de las arandelas tiene un espesor entre 0,0180 y 0,0202 pulgadas.

8)

a) El 0,927 de las personas tiene un tiempo de reacción entre 14 y 30 segundos.

b) El 0,0994 de las personas tiene un tiempo de reacción entre 25 y 30 segundos.

c) El 93,32 % de las personas tiene un tiempo de reacción de más de 14 segundos.

d) El tiempo de reacción es de 29,28 segundos.

9) El promedio de los tarros es 12 onzas.

Autoevaluación 1

1) En una ciudad se publican los periódicos A, B y C. Una encuesta reciente a 800 lectores indica lo siguiente : 208 lee A, 240 lee B, 192 lee C, 64 lee A y B; 40 lee A y C; 32 lee B y C; 24 lee A, B y C. Para un adulto escogido al azar, calcular la probabilidad de que:

a) no lea ninguno de los periódicos.
b) lea exactamente uno de los periódicos
c) lea B y C, pero no A
d) lea sólo A o sólo C

2) Si $P(A) = \frac{3}{7}$, $P(B) = \frac{2}{3}$ y $P(A \cap B) = \frac{1}{6}$ Determine :

a) $P(A \cup B)$ b) $P(B^c)$
c) $P(B - A)$ d) $P(A^c/B^c)$
e) ¿A y B independientes?. Justifique

3) Si se sacan al azar y sin reemplazo cuatro pelotas de una bolsa que contiene 6 pelotas rojas y 7 negras.

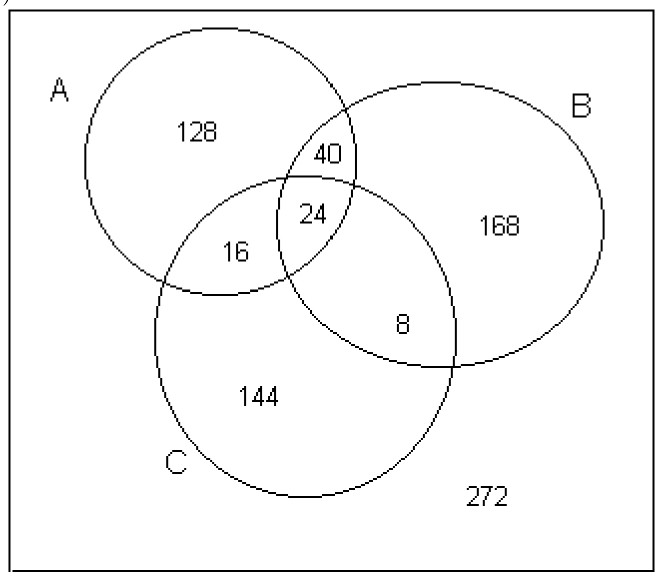
a) ¿Cuál es la probabilidad de que la primera pelota sea negra y tres restantes rojas?.
b) Si las tres primeras pelotas fueron rojas. ¿Cuál es la probabilidad de que la cuarta pelota sea negra?
c) Si las dos primeras pelotas fueron rojas. ¿Cuál es la probabilidad que la tercera sea negra y la cuarta roja?
d) ¿Cuál es la probabilidad de que las dos primeras pelotas sean rojas y las dos últimas negras?
e) ¿Cuál es la probabilidad de que aparezca una de cada color?

4) Se recibieron dos cajas de camisas para hombre, provenientes de la fábrica. La caja uno contenía 15 camisas deportivas y 25 camisas de vestir. La caja dos contenía 10 camisas deportivas y 30 camisas de vestir.

a) ¿Cuál es la probabilidad de elegir una camisa deportiva?
b) ¿Cuál es la probabilidad de elegir una camisa de vestir?
c) Se seleccionó al azar una de las dos cajas y se eligió aleatoriamente una camisa de esa caja para inspeccionarla. La camisa era de vestir. Dada esta información, ¿cuál es la probabilidad de que la caja de la que proviene la camisa deportiva sea la uno?

Solución

1)



a) La probabilidad de no lea ningún periódico es $\frac{272}{800}$

b) La probabilidad de que lea exactamente uno de los periódicos es $\frac{440}{800}$

c) La probabilidad de que lea B y C, pero no A es $\frac{8}{800}$

d) La probabilidad de que lea sólo A o sólo C es $\frac{272}{800}$

2)

$a)\ \mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \frac{39}{42}$ $\qquad b)\ \mathrm{P}(\mathrm{B}^c) = \frac{1}{3}$

$c)\ \mathrm{P}(\mathrm{B} - \mathrm{A}) = \frac{1}{2}$ $\qquad d)\ \mathrm{P}(\mathrm{A}^c/\mathrm{B}^c) = \frac{3}{14}$

$e)\ \mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \frac{1}{6}$ $\qquad \mathrm{P}(\mathrm{A}) \cdot \mathrm{P}(\mathrm{B}) = \frac{2}{7}$

Por lo tanto, A y B no son independientes.

3)

a) $\mathrm{P}(\mathrm{N}_1 \cap \mathrm{R}_2 \cap \mathrm{R}_3 \cap \mathrm{R}_4) = \frac{7}{13} \cdot \frac{6}{12} \cdot \frac{5}{11} \cdot \frac{4}{10} = \frac{7}{143}$

b) $P(N_4/R_1 \cap R_2 \cap R_3) = \frac{7}{10}$

c) $P(N_3 \cap N_4/R_1 \cap R_2) = \frac{7}{11} \cdot \frac{4}{10} = \frac{14}{55}$

d) $P(R_1 \cap R_2 \cap N_3 \cap N_4) = \frac{6}{13} \cdot \frac{5}{12} \cdot \frac{7}{11} \cdot \frac{6}{10} = \frac{21}{286}$

e) $P(R_1 \cap N_2 \cap R_3 \cap N_4) + P(N_1 \cap R_2 \cap N_3 \cap R_4)$
$= \frac{6}{13} \cdot \frac{7}{12} \cdot \frac{5}{11} \cdot \frac{6}{10} + \frac{7}{13} \cdot \frac{6}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} = \frac{21}{143}$

4)

Sea $D =$ camisa deportiva y $V =$ camisa de vestir

Sea $C_1 =$ caja uno y $C_2 =$ caja dos

a) $P(D) = \frac{25}{80}$

b) $P(V) = \frac{55}{80}$

c) $P(C_1/V) = \frac{5}{11}$

Autoevaluación 2

1) Dada la función de densidad de la v. a. X :

$$f(x) = \begin{cases} kx(1-x) & 0 < x < 1 \\ 0 & \text{e.o.c.} \end{cases}$$

a) Determine el valor de k .
b) Obtener :

b.1) P$(x = 1)$ b.2) P$\left(\frac{1}{2} < x < \frac{3}{2}\right)$

b.3) P$\left(x \geq \frac{2}{3}\right)$

c) Calcule E$(5x + 7)$

2) De acuerdo con un estudio publicado por un grupo de sociólogos de una cierta universidad, aproximadamente el 60 % de los adictos al Valium en el estado XX, lo tomaron por primera vez debido a problemas sicológicos. Encuentre la probabilidad de que de los siguientes 8 adictos entrevistados :

a) exactamente 3 hayan comenzado a usarlo debido a problemas sicológicos.
b) al menos 3 de ellos comenzaron a tomarlo por problemas que no fueron sicológicos.

3) Un comité de 3 integrantes se forma aleatoriamente seleccionado de entre 5 doctores y 3 enfermeras. Encuentre la distribución de probabilidades para el número de enfermeras y determine P$(2 \leq x \leq 3)$.

4) En un estudio de un inventario se determinó que, en promedio, la demanda por un artículo en particular en una bodega era de 5 veces al día. ¿Cuál es la probabilidad que en un determinado día este artículo sea requerido :

a) más de 4 veces?.
b) ni una sola vez?.

5) Se encontró que un grupo de calificaciones de exámenes finales en un curso de estadística elemental estaba normalmente distribuida con una media de 80 y una desviación estándar de 8.

a) ¿Cuál es la probabilidad de obtener cuando mucho una calificación de 81 en este examen ?.
b) ¿Qué porcentaje de los estudiantes alcanzaron calificaciones entre 55 y 89 ?.
c) ¿Cuál es la probabilidad de obtener a lo menos un 47?
d) ¿Cuál es la probabilidad de obtener una nota superior a 51, pero inferior a 85?
c) ¿Cuál es la calificación del examen final si sólo el 5 % de los estudiantes que pasaron la prueba tuvieron calificaciones más altas?.

Solución

1)

$a)\, k = 6$

$b.1)\, \mathrm{P}(x = 1) = 0$

$b.2)\, \mathrm{P}\left(\frac{1}{2} < x < \frac{3}{2}\right) = \frac{1}{2}$

$b.3)\, \mathrm{P}\left(x \geq \frac{2}{3}\right) = \frac{7}{27}$

$c)\, E(5x + 7) = \frac{19}{2}$

2)

$a)\, f(x) = \binom{8}{x}(0,60)^x(0,40)^{8-x}$

$\mathrm{P}(x = 3) = 0,124$

$b)\, f(x) = \binom{8}{x}(0,40)^x(0,60)^{8-x}$

$\mathrm{P}(x \geq 3) = 1 - \mathrm{P}(x = 0) + \mathrm{P}(x = 1) + \mathrm{P}(x = 2) = 0,685$

3)

$$f(x) = \frac{\binom{3}{x}\binom{5}{3-x}}{\binom{8}{3}}$$

x	0	1	2	3
$f(x)$	$\frac{10}{56}$	$\frac{30}{56}$	$\frac{15}{56}$	$\frac{1}{56}$

$\mathrm{P}(2 \leq x \leq 3) = \frac{2}{7}$

4)

$$f(x) = \frac{e^{-5} \cdot 5^x}{x!}$$

$a)\, \mathrm{P}(x > 4) = 0,56$

$b)\, \mathrm{P}(x = 0) = 0,007$

$5)$

$\mu = 80 \qquad \sigma = 8 \qquad X \sim N(80, 64)$

a) La probabilidad de obtener cuando mucho una calificación 81 es $0,5517$

b) El $87,08$ % de los estudiantes obtuvieron calificaciones entre 55 y 89

c) La probabilidad de obtener a lo menos un 47 es 1

d) La probabilidad de obtener una nota superior a 51, pero no inferior a 85 es $0,7357$

e) La calificación es de un $93,12$

Unidad N°3:Intervalos de Confianza

Inferencia Estadística

La teoría de ***Inferencia Estadística*** consiste en aquellos métodos con los cuales se pueden realizar inferencias o generalizaciones acerca de una población.

La Inferencia Estadística puede dividirse en 2 áreas:

a) Estimación de Parámetros

b) Pruebas de Hipótesis

Estimación de parámetros

Los parámetros a estudiar son parámetros poblacionales como la ***media*** y la ***varianza.***

Si θ es un parámetro desconocido, entonces $\hat{\theta}$ será su estimador.

Así, $\overline{x}$ es un estimador de μ y s^2 es un estimador de σ^2 y ellos cumplen con la propiedad de **insesgamiento.**

Estimación por intervalo

Una estimación por intervalo de un parámetro poblacional θ es un intervalo de la forma $\hat{\theta}_1 < \theta < \hat{\theta}_2$, donde $\hat{\theta}_1$ y $\hat{\theta}_2$ dependen del valor de $\hat{\theta}$ para una muestra particular y también de la distribución muestral de $\hat{\theta}$.

Basado en la distribución muestral de $\hat{\theta}$ se puede determinar si el intervalo $\left(\hat{\theta}_1, \hat{\theta}_2\right)$ con una probabilidad dada contiene realmente el parámetro que se supone que va a estimar.

Esto es P$\left(\hat{\theta}_1 < \theta < \hat{\theta}_2\right) = 1 - \alpha$ donde $0 < \alpha < 1$

El intervalo $\left(\hat{\theta}_1, \hat{\theta}_2\right)$ calculado de una muestra particular se llama ***intervalo de confianza*** del $(1-\alpha)100\,\%$, la fracción $1-\alpha$ se denomina ***coeficiente de confianza, grado de confianza*** o ***nivel de confianza*** y los puntos $\hat{\theta}_1$ y $\hat{\theta}_2$ se llaman ***límites de confianza.***

Por ejemplo:

a) Si $\alpha = 0,05$, entonces se tiene un intervalo de confianza del $95\,\%$

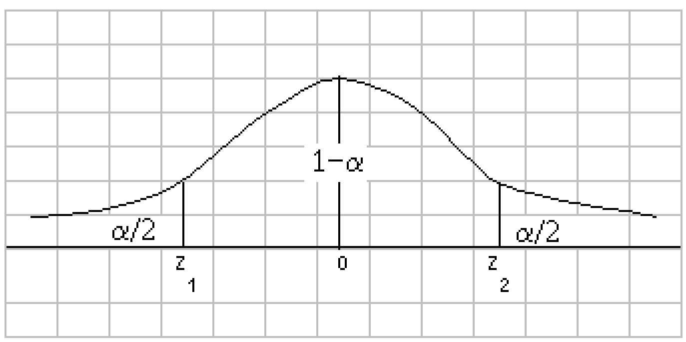
b) Si $\alpha = 0,01$, entonces el intervalo de confianza es del $99\ \%$

A) Intervalo de confianza para la media (μ) de una población normal

A$_1$) Se conoce su varianza

Obs.: Si $X \sim N(\mu\,,\sigma^2)$, entonces $\overline{x} \sim N\left(\mu\,,\dfrac{\sigma^2}{n}\right)$

Como $\overline{x} \sim N\left(\mu\,,\dfrac{\sigma^2}{n}\right)$, entonces $Z = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$



$\mathrm{P}(Z_1 < Z < Z_2) = 1 - \alpha$

$\mathrm{P}(Z < Z_2) = \dfrac{\alpha}{2} + 1 - \alpha = 1 - \dfrac{\alpha}{2}$

$\mathrm{P}(Z < Z_1) = \dfrac{\alpha}{2}$

Luego: $Z_2 = Z_{1-\frac{\alpha}{2}}$ por construcción

$Z_1 = Z_{\frac{\alpha}{2}}$ pero $Z_1 = -Z_2$

Luego: $Z_1 = -Z_{1-\frac{\alpha}{2}}$

Así, $\quad P(Z_1 < Z < Z_2) = 1 - \alpha$

$$P\left(-Z_{1-\frac{\alpha}{2}} < \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \overline{x}-\mu < Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(-\overline{x}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{x}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\overline{x}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Si $\overline{x}$ es la media de una muestra aleatoria de tamaño n de una población normal con varianza σ^2, el intervalo de confianza de $(1-\alpha)100$ % para μ es :

$$\boldsymbol{\left(\overline{x}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\overline{x}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)}$$

Ejemplo: Si una muestra aleatoria de tamaño 20 de una población normal con varianza 225 tiene una media muestral de $64,3$. Construya un intervalo de confianza del 95% de confianza para μ

$(1-\alpha)100\% = 95\% \Rightarrow \alpha = 0,05$
$n = 20$
$\sigma^2 = 225 \Rightarrow \sigma = 15$
$\overline{x} = 64,3$

$$\left(64,3 - Z_{1-\frac{0,05}{2}}\frac{15}{\sqrt{20}}, 64,3 + Z_{1-\frac{0,05}{2}}\frac{15}{\sqrt{20}}\right)$$

$$\left(64,3 - (1,96)\frac{15}{\sqrt{20}}; 64,3 + (1,96)\frac{15}{\sqrt{20}}\right)$$

$(57,7; 70,9)$

Teorema: Si se usa $\overline{x}$ como una estimación de μ , se puede tener una confianza del $(1-\alpha)100\%$ de que el ***error*** no excederá de: $Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

En el ejemplo anterior:

$$Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$$

$$(1,96)\frac{15}{\sqrt{20}} = 6,57$$

Se puede tener una confianza del 95% de que $\overline{x}$ difiere de μ por una cantidad menor que $6,57$.

Teorema: Si se utiliza $\overline{x}$ como una estimación de μ , entonces se puede tener una confianza del $(1-\alpha)100\%$ de que el ***error*** no excederá una cantidad específica e cuando el tamaño de la muestra es:

$$n = \left(\frac{\left(Z_{1-\frac{\alpha}{2}}\right)\sigma}{e}\right)^2$$

Ejemplo: ¿Qué tan grande se requiere que sea la muestra del ejemplo (1) si se desea una confianza del 95% de que la estimación de μ difiera de ésta por menos de $0,05$?

$e = 0,05$ $\qquad Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$ $\qquad \sigma = 15$

$\alpha = 0,05$

$$n = \left(\frac{1,96(15)}{0,05}\right)^2 = 345.744$$

Luego, se puede tener una confianza del 95% de que la muestra aleatoria de tamaño 345.744 proporcionará una estimación de $\overline{x}$ que difiere de μ por una cantidad menor que $0,05$.

Observación: Todo lo anterior también es aplicable a poblaciones no normales con varianza conocida cuando $n > 30$

Ejercicios

1) Las medidas de los diámetros de los rodamientos tiene una desviación estándar de 0,042 cm. Se selecciona una muestra aleatoria de 200 bolas de rodamientos producidas por una máquina en una semana, los diámetros dieron una media de 0,824 cm. Hallar un intervalo de confianza del 95 % y 99 % para el diámetro de todos los rodamientos.

2) Suponga que la duración de un componente tiene distribución normal con $(\mu, 9)$. Se prueban 20 componentes y se anotan sus tipos de fallas $x_1, x_2, x_3, \ldots, x_{20}$. Suponga además que la media de la muestra es 100,9 horas. Obtener un intervalo de confianza del 99 % para la duración media de todos los componentes.

3) Se administra un test estándar a una numerosa clase de estudiantes. La puntuación media de una muestra de 100 estudiantes es de 75 puntos. Supóngase que la varianza admitida de las puntuaciones para este test sea de 2.500 puntos. Hallar :

a) Intervalo de confianza del 98 % para μ .

b) Límite superior del intervalo de confianza del 95 % para μ .

c) Límite inferior del intervalo de confianza del 90 % para μ .

4) Al medir el tiempo de reacción de una persona, un psicólogo estima que la desviación estándar es de $0,05$ segundos. ¿De qué tamaño ha de tomarse una muestra de medidas para tener una confianza del 95% y 99% de que el error de la estimación no supera los $0,01$ segundos?

Solución

1) $95\% \Rightarrow (0,8182; 0,8298)$ $99\% \Rightarrow (0,816; 0,8316)$

2) $(99,17; 102,63)$

3)
a) $(63,35; 86,65)$

b) $84,8$

c) $66,775$

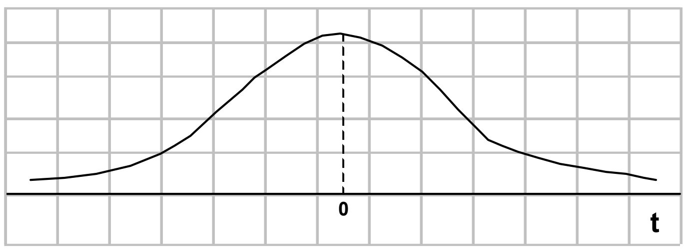
4) $95\% \Rightarrow n = 96,04 \approx 97$ $99\% \Rightarrow n = 166,4 \approx 167$

A_2) Varianza desconocida

Sea $x_1, x_2, \ldots\ldots, x_n$ una muestra aleatoria de $X \sim N(\mu\,, \sigma^2)$ con σ^2 desconocida.

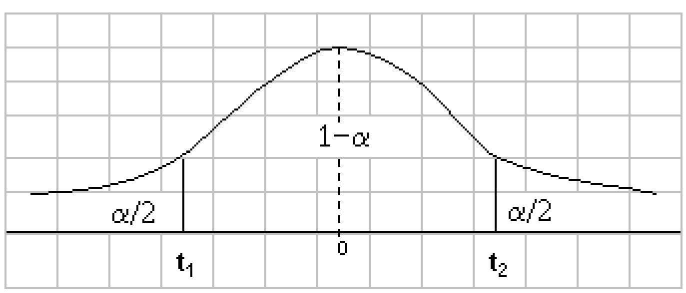
$T = \dfrac{\overline{x} - \mu}{\dfrac{s}{\sqrt{n}}}$ tiene distribución $t - student$ con $(n-1)$ grados de libertad (s es la desviación estándar de la muestra).

La función de densidad de la $t - student$ gráficamente es similar a la función de densidad de la normal.



Su función de distribución acumulada se encuentra tabulada.

El parámetro que caracteriza a la $t - student$ se conoce como ***grados de libertad.***



$\mathrm{P}(t_1 < T < t_2) = 1 - \alpha$

$$\mathrm{P}\left(-\,t_{\dfrac{\alpha}{2}} < \dfrac{\overline{x} - \mu}{\dfrac{s}{\sqrt{n}}} < t_{\dfrac{\alpha}{2}} \right) = 1 - \alpha$$

$$\mathrm{P}\left(\overline{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right) = 1 - \alpha$$

Si $\overline{x}$ es la media de la muestra aleatoria (m.a.) de tamaño n de una población normal con varianza desconocida, el intervalo de confianza de $(1-\alpha)100\%$ para μ es:

$$\boldsymbol{\left(\overline{x} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\ ,\ \overline{x} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)}$$ **con $\boldsymbol{n-1}$ grados de libertad**

Ejemplo: Un fabricante de pintura quiere determinar el tiempo de secado promedio para una nueva pintura para pared interior. Si para una prueba de 12 áreas de igual tamaño obtiene un tiempo medio de secado de $66,3$ minutos y una desviación estándar de $8,4$ minutos. Construya un intervalo del 95% de confianza para μ si el tiempo de secado tiene distribución normal.

Solución: $n = 12$ $\qquad n-1 = 11$

$\overline{x} = 66,3$ $\qquad t_{\frac{\alpha}{2}} = t_{0,025} \Rightarrow t_{\frac{\alpha}{2}} = 2,201$

$s = 8,4$ $\alpha = 0,05$

$$\left(66,3 - (2,201)\frac{8,4}{\sqrt{12}}; 66,3 + (2,201)\frac{8,4}{\sqrt{12}}\right)$$

$(61; 71,6)$

Teorema: Si se usa $\overline{x}$ como una estimación de μ , se puede tener una confianza del $(1-\alpha)100\%$ de que el ***error*** no excederá de:

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

Para el ejemplo anterior: $t_{\frac{\alpha}{2}} = 2,201$ $\qquad s = 8,4$ $\sqrt{n} = \sqrt{12}$

$$e = (2,201)\frac{8,4}{\sqrt{12}}$$

$$e = 5,34$$

Como: $$e = t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$$

$$\sqrt{n} = \frac{t_{\frac{\alpha}{2}}\, s}{e}$$

$$n = \left(\frac{t_{\frac{\alpha}{2}}\, s}{e}\right)^2$$ (Tamaño de la muestra)

En el ejemplo del fabricante de pintura, determine el tamaño de la muestra si el error no debe exceder de $0,25$:

$$n = \left(\frac{2,201(8,4)}{0,25} \right)^2$$

$$n = 5469$$

Ejercicios

1) Se van a realizar durante un mes pruebas de mercado de un nuevo instrumento, en determinadas tiendas de una ciudad. Los resultados para una muestra de 16 tiendas señalaron ventas promedio de \$ 12.000 con una desviación estándar de \$ 180. Estime un intervalo de confianza del 99 % de las ventas promedio reales de este nuevo instrumento. Suponga distribución normal.

2) Suponga que se hacen 20 mediciones sobre la resistencia de cierto tipo de alambre. La media de la muestra es 10,48 ohms y la desviación estándar 1,36 ohms. Obtener un intervalo de confianza de un 99 % para la resistencia promedio real si ellas se distribuyen normalmente.

3) Una muestra aleatoria de 100 propietarios de automóviles indica que, en el estado XX, un automóvil recorre un promedio de 23.500 Km. por año con una desviación estándar de 3.900 Km. Determine un intervalo de confianza de 98 % para la cantidad promedio de Km. que un automóvil recorre anualmente en el estado XX. Suponga distribución normal.

4) Una muestra aleatoria de 8 cigarros de una marca determinada tiene un contenido promedio de nicotina de 2,6 milígramos y una desviación estándar de 0,9 milígramos. Determine un intervalo de confianza de 95 % para el contenido promedio real de nicotina en esta marca de cigarros en particular, si se sabe que la distribución de los contenidos de nicotina son normales.

Solución

1) $(11867,385;12132,615)$

2) $(9,61;11,35)$

3) $(22578,04;24421,96)$

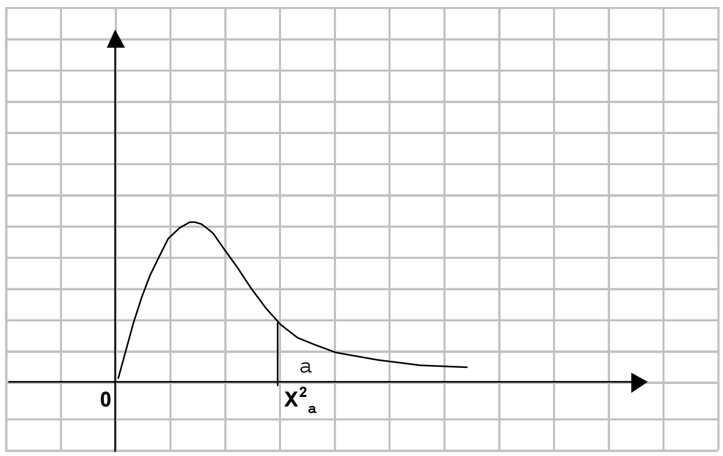
4) $(1,847;3,353)$

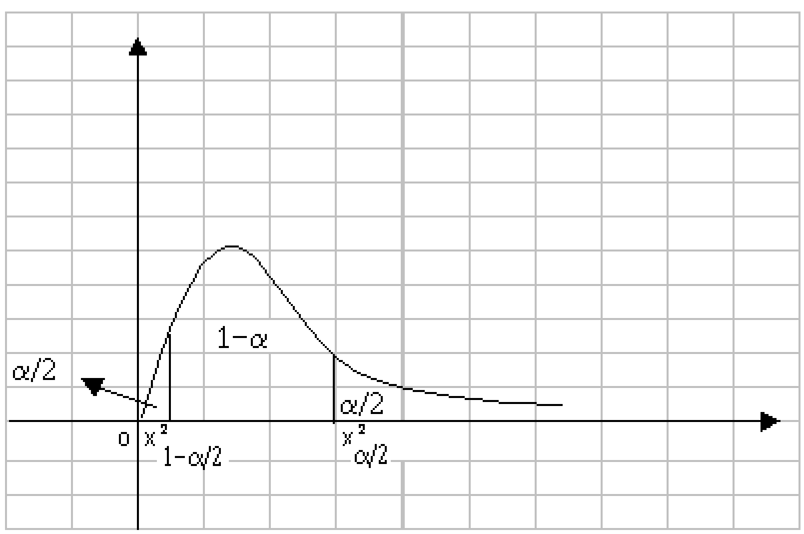
B) Intervalo de confianza para la varianza (σ^2) de una población normal

Sea $x_1, x_2, ..., x_n$ una muestra aleatoria de $X \sim N(\mu\,, \sigma^2)$ con σ^2 desconocida.

$X^2 = \dfrac{(n-1)s^2}{\sigma^2}$ tiene distribución $ji\,cuadrado$ con $(n-1)$ grados de libertad $(s^2$ es la varianza de la muestra).

La gráfica de la función de densidad de esta distribución es:





$$P\left(X^2_{1-\frac{\alpha}{2}} < X^2 < X^2_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(X^2_{1-\frac{\alpha}{2}} < \frac{(n-1)s^2}{\sigma^2} < X^2_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(\frac{1}{X^2_{\frac{\alpha}{2}}} < \frac{\sigma^2}{(n-1)s^2} < \frac{1}{X^2_{1-\frac{\alpha}{2}}}\right) = 1 - \alpha$$

$$P\left(\frac{(n-1)s^2}{X^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{X^2_{1-\frac{\alpha}{2}}}\right) = 1 - \alpha$$

Si s^2 es la varianza de una muestra aleatoria de tamaño n de una población normal, su intervalo de confianza es del $(1-\alpha)100\%$ para σ^2 es:

$$\boldsymbol{\left(\frac{(n-1)s^2}{X^2_{\frac{\alpha}{2}}}, \frac{(n-1)s^2}{X^2_{1-\frac{\alpha}{2}}}\right)}$$

donde $X^2_{\frac{\alpha}{2}}$ y $X^2_{1-\frac{\alpha}{2}}$ son valores X^2 con $n-1$ grados de libertad, con áreas de $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$, respectivamente, a la derecha.

Ejemplo:

1) Determine un intervalo de confianza del 95% para la varianza de una muestra de 10 paquetes de semilla, si la varianza de la muestra es $0,286$

$\alpha = 0,05$ $\qquad 1-\frac{\alpha}{2} = 0,975$ $\qquad \frac{\alpha}{2} = 0,025$

$n = 10$ $\qquad n-1 = 9$ $\qquad s^2 = 0,286$

$X^2_{1-\frac{\alpha}{2}} = 2,700 \quad X^2_{\frac{\alpha}{2}} = 19,023$

$$\left(\frac{9(0,286)}{19,023}, \frac{9(0,286)}{2,700}\right)$$

$(0,135; 0,953)$

2) Se obtiene una m. a. de 20 estudiantes con una media de $\overline{x} = 72$ y una varianza de $s^2 = 16$ en un examen de Estadística. Suponga que las calificaciones tienen distribución normal. Determine un intervalo de confianza del 98% para la varianza poblacional

$\alpha = 0,02$ $\qquad\qquad 1 - \dfrac{\alpha}{2} = 0,99$

$n = 20$ $\qquad\qquad n - 1 = 19$

$s^2 = 16$

$X^2_{1-\frac{\alpha}{2}} = X^2_{(0,99)(19)} = 7,633$

$X^2_{\frac{\alpha}{2}} = X^2_{(0,01)(19)} = 36,191$

$$\left(\frac{19(16)}{36,191}, \frac{19(16)}{7,633}\right)$$

$(8,39; 39,82)$

Ejercicios

1) Un fabricante de baterias para automóvil asegura que sus baterias duran en promedio, 3 años con una desviación estándar de un año. Si 5 de estas baterias tienen una desviación estándar de 0,9028 años. determine un intervalo de confianza de 95 % para la varianza real e indique si es válida la afirmación del fabricante. Suponga que la población de las duraciones de las baterias se distribuye aproximadamente en forma normal.

2) Suponga que se hacen 20 mediciones sobre la resistencia de cierto tipo de alambre. La media de la muestra es 10,48 ohms y la desviación estándar 1,36 ohms. Obtener un intervalo de confianza de un 95 % para la varianza real si las resistencias se distribuyen normalmente.

3) Una muestra aleatoria de 25 cigarros de una cierta marca tiene un contenido promedio de nicotina de 1,3 milígramos y una desviación estándar de 0,17 milígramos. Encuentre un intervalo de confianza del 90 % y 98 % para la varianza real de esta determinada marca de cigarros si se supone que las mediciones se distribuyen normalmente.

4) Una muestra aleatoria de 100 propietarios de automóviles indica que, en el estado XX, un automóvil recorre un promedio de 23.500 Km. por año con una desviación estándar de 3.900 Km. Determine un intervalo de confianza de 99 % para la varianza real de la cantidad de Km. por año que recorren los automóviles del estado XX.

Solución

1) $(0,29;6,79)$ La afirmación del fabricante es válida, porque la varianza poblacional está dentro del intervalo que se determinó con una confianza del 95%.

2) $(1,069;3,949)$

3) $90\ \% \Rightarrow (0,019;0,05)$

$98\ \% \Rightarrow (0,016;0,064)$

4) $(10741065,69;22374294,2)$

Autoevaluación

1) Suponga que una tienda de pinturas quisiera estimar la cantidad correcta de pintura que hay en latas de un galón, compradas a un conocido fabricante. Por las especificaciones del productor se sabe que la desviación estándar de la cantidad de pintura es igual a 0,02 galones. Se selecciona una muestra aleatoria de 50 galones y la cantidad promedio de pintura es 0,975 galones. Establezca un intervalo de confianza del 95 % y 99 % de la cantidad promedio real de la población de pintura incluida en una lata de un galón.

2) La vida útil promedio de una muestra aleatoria de 10 focos es de 4000 horas con una desviación estándar de 200 horas. Se supone que la vida útil de los focos tiene una distribución normal.Estime un intervalo de confianza del 90 y 95 % para la vida útil promedio

3) El departamento de servicios a clientes de una empresa local de servicios públicos de gas querría estimar la variación en el tiempo entre la llegada de la solicitud de servicio y la conexión del mismo. De los registros disponibles del año anterior se seleccionó una muestra aleatoria de 15 casas que dieron una desviación estándar de 20,03. Estime un intervalo de confianza del 92 y 96 % para la varianza poblacional.

Solución

1) Usar Tabla Normal

Para 95% : $(0,969\ ;\ 0,981)$

Para 99% : $(0,968\ ;\ 0,982)$

2) Usar Tabla t-student

Para 90% : $(3884,26\ ;\ 4115,74)$

Para 95% : $(3857,07\ ;\ 4142,93)$

3) Usar Tabla ji cuadrado

Para 95% : $(215,05\ ;\ 997,83)$

Para 99% : $(163,66\ ;\ 1378,36)$

Unidad N°4: Pruebas de Hipótesis

Son procedimientos de decisión basado en datos que puedan producir una conclusión acerca de algún sistema científico.

Una ***hipótesis estadística*** es una afirmación o conjetura acerca de una o más poblaciones.

No es posible saber con absoluta certeza la verdad o falsedad de una hipótesis estadística, pues para ello habría que trabajar con toda la población. En la práctica se toma una muestra aleatoria de la población de interés y se utilizan los datos que contiene tal muestra para proporcionar evidencias que confirmen o no la hipótesis. Si la evidencia de la muestra es inconsistente con la hipótesis planteada, entonces ésta se rechaza y si la evidencia apoya a la hipótesis planteada, entonces se acepta ésta.

La aceptación de una hipótesis implica tan sólo que los datos no proporcionan evidencia suficiente para refutarla. Por otro lado, el rechazo implica que la evidencia de la muestra la refuta.

La estructura de una prueba de hipótesis consiste en la formulación de una ***hipótesis nula*** , es decir, cualquier hipótesis que se desee probar, se denota por $\boldsymbol{H_0}$. El rechazo de H_0, genera la aceptación de una ***hipótesis alternativa*** , que se denota por $\boldsymbol{H_1}$.

Una hipótesis nula referente a un parámetro poblacional siempre debe establecerse de manera que especifique un valor exacto del parámetro, mientras que la hipótesis alternativa admite la posibilidad de varios valores.

Por ejemplo:

1) $H_0 : \mu = 20$ $\quad H_1 : \mu > 20$

2) $H_0 : \mu = 20$ $\quad H_1 : \mu < 20$

3) $H_0 : \mu = 20$ $\quad H_1 : \mu \neq 20$

En la hipótesis alternativa se plantea usualmente lo que se cree verdadero y en la hipótesis nula lo que se desea rechazar.

Para tomar una decisión acerca de un parámetro es necesario una ***prueba estadística*** para cuantificar esta decisión. Esto se logra al establecer primero la distribución muestral que sigue la muestra estadística (es decir, la media) y después calcular la prueba estadística apropiada. Esta prueba estadística mide qué tan cerca de la hipótesis nula se encuentra el valor de la muestra. La prueba estadística suele seguir una distribución estadística conocida (normal, t-student, ji cuadrado)

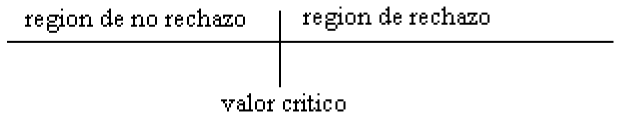
La distribución apropiada de la prueba estadística se divide en dos regiones

a) región de rechazo (región crítica)

b) región de no rechazo

Si la prueba estadística cae en la región de no rechazo no se puede rechazar la hipótesis nula y si cae en la región de rechazo, se rechaza la hipótesis nula.

Para decidir con relación a la hipótesis nula, primero se tiene que determinar el ***valor crítico*** para la distribución estadística de interés. El valor crítico separa la región de no rechazo de la de rechazo.



Errores al realizar una prueba de hipótesis

Al utilizar una muestra para obtener conclusiones sobre una población existe el riesgo de llegar a una conclusión incorrecta. Pueden ocurrir dos errores diferentes:

1) ***Error tipo I*** consiste en rechazar H_0 cuando ésta es verdadera.

2) ***Error tipo II*** consiste en aceptar H_0 cuando ésta es falsa.

Al probar cualquier hipótesis estadística, existen cuatro posibles situaciones que determinan si la decisión es correcta o equivocada.

	H_0es verdadera	H_0es falsa
se acepta H_0	decisión correcta	error tipo II
se rechaza H_0	error tipo I	decisión correcta

La probabilidad de cometer error tipo I, es decir, rechazar H_0cuando es verdadera, se denomina ***nivel de significación*** y se denota por α .P(error tipo I) $= \alpha$

La probabilidad de no cometer error tipo I, es decir, aceptar H_0cuando es verdadera, se denota por $1 - \alpha$ .P(error tipo I)$^c = 1 - \alpha$

La probabilidad de cometer error tipo II, es decir, aceptar H_0cuando es falsa, se representa por β .P(error tipo II) $= \beta$

La probabilidad de no cometer error tipo II, es decir, rechazar H_0cuando es falsa, se denomina ***potencia de la prueba*** y se denota por $1 - \beta$ .P(error tipo I)$^c = 1 - \beta$

El ideal al rechazar una prueba de hipótesis es determinar los procedimientos o reglas que conduzcan a maximizar la potencia de una prueba, para un α fijo. α se suele especificar antes de tomar una muestra, es frecuente que $\alpha = 0,05$ o $\alpha = 0,01$

Esquema para realizar una prueba de hipótesis acerca de un parámetro $\boldsymbol{\theta}$

1) Plantear la hipótesis nula y la hipótesis alternativa.

$a)$ $H_0 : \theta \leq \theta_1$
$H_1 : \theta > \theta_1$

$b)$ $H_0 : \theta \geq \theta_1$
$H_1 : \theta < \theta_1$

$c)$ $H_0 : \theta = \theta_1$
$H_1 : \theta \neq \theta_1$

2) Seleccionar el test estadístico o estadístico de prueba.

3) Fijar α $(0,05\,;0,01\,;0,10)$

4) Construir la regla de decisión o ***región crítica*** con el valor elegido de α .

5) Extraer una muestra aleatoria de tamaño n y calcular el valor del test estadístico.

6) Si el valor calculado del test estadístico cae en la región crítica rechazar H_0, en caso contrario no rechazar H_0 y concluir que la muestra aleatoria no proporciona evidencia para rechazarla.

Pruebas unilaterales y bilaterales

Una prueba de hipótesis será unilateral $($de una cola$)$ en los siguientes casos

$a)$ $H_0 : \theta = \theta_1$
$H_1 : \theta > \theta_1$

$b)$ $H_0 : \theta = \theta_1$
$H_1 : \theta < \theta_1$

$c)$ $H_0 : \theta \leq \theta_1$
$H_1 : \theta > \theta_1$

$d)$ $H_0 : \theta \geq \theta_1$
$H_1 : \theta < \theta_1$

Una prueba de hipótesis será bilateral $($de dos colas$)$ si

$H_0 : \theta = \theta_1$
$H_1 : \theta \neq \theta_1$ $(\theta < \theta_1 \vee \theta > \theta_1)$

Pruebas de hipótesis

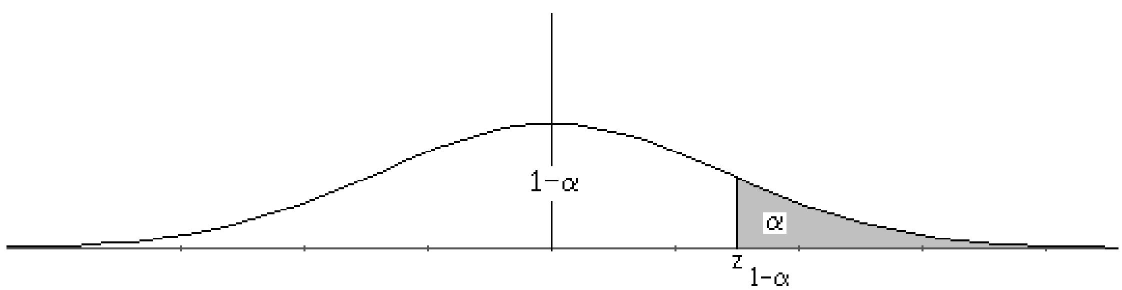
a) Para la media (μ) si la varianza (σ^2)es conocida

Recuerde que si $X \sim N(\mu\,, \sigma^2)$, entonces $\overline{X} \sim N\left(\mu\,, \frac{\sigma^2}{n}\right)$. Luego la prueba estadísta adecuada debe ser

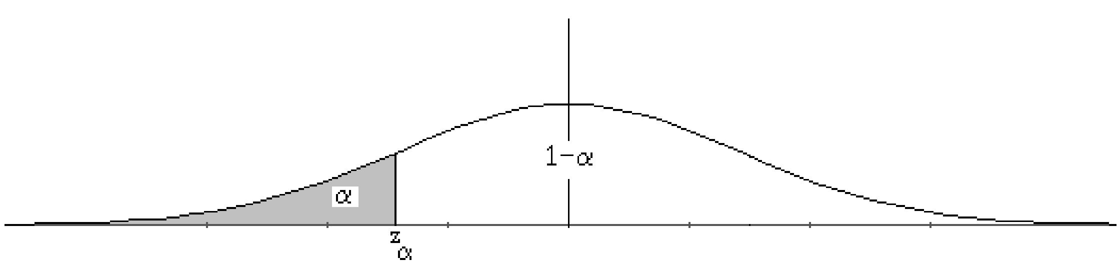
$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

1) Para pruebas de hipótesis unilaterales

$a)$ $\begin{array}{l} H_0 : \theta = \theta_1 \\ H_1 : \theta > \theta_1 \end{array}$ o $\begin{array}{l} H_0 : \theta \leq \theta_1 \\ H_1 : \theta > \theta_1 \end{array}$

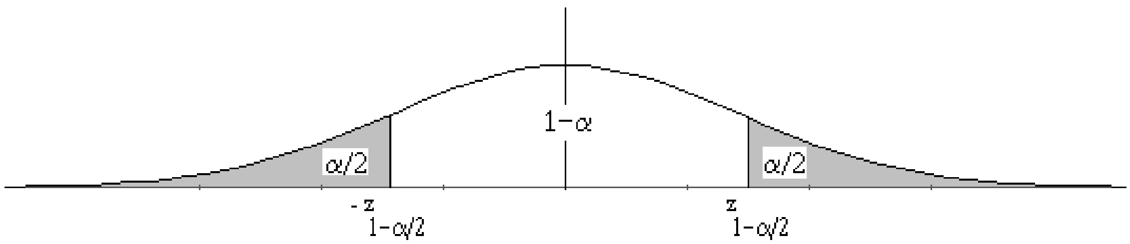


$b)$ $\begin{array}{l} H_0 : \theta = \theta_1 \\ H_1 : \theta < \theta_1 \end{array}$ o $\begin{array}{l} H_0 : \theta \geq \theta_1 \\ H_1 : \theta < \theta_1 \end{array}$



2) Para pruebas bilaterales

$H_0 : \theta = \theta_1$
$H_1 : \theta \neq \theta_1$



Ejemplos

1) Considere la hipótesis nula de que el peso promedio de estudiantes hombres de un cierto instituto es 68 kilos contra la hipótesis alternativa de que es diferente de 68 kilos. Suponga que los pesos se distribuyen normalmente con una desviación estándar de 3,6 kilos. Se elige una muestra aleatoria de 36 estudiantes y se obtiene un peso promedio de 67,5 kilos. Utilice un nivel de significación del 5 %.

$H_0 : \mu = 68$
$H_1 : \mu \neq 68$

$\alpha = 0,05$

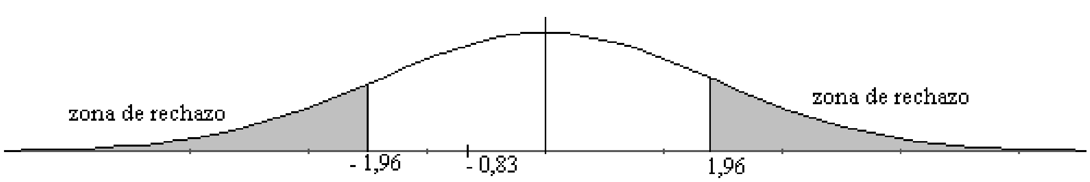
Región crítica $= RC$

$RC : -z_{1-\alpha/2} < z < z_{1-\alpha/2} \qquad \Rightarrow RC : -z_{0,975} < z < z_{0,975}$

$\Rightarrow RC : -1,96 < z < 1,96$

$\overline{x} = 67,5 \qquad n = 36 \qquad \sigma = 3,6$

$$z = \frac{67,5 - 68}{\frac{3,6}{6}} = -0,83$$



Se acepta H_0, es decir, no es posible decidir si el peso promedio de los estudiantes de un cierto instituto es distinto de 68 kilos.

2) Una muestra aleatoria de 100 muertos registrados en Chile durante el año pasado mostró una vida promedio de 71,8 años. Suponiendo una desviación estándar poblacional de 8,9 años. ¿Parecería esto indicar que la vida promedio hoy día es mayor que 70 años? Utilice un nivel de significación de 0,05

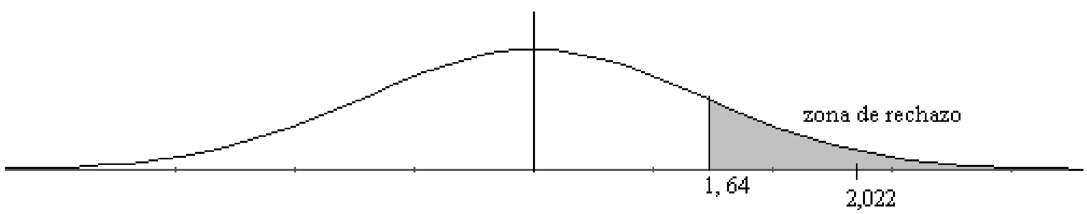
$H_0 : \mu \leq 70$
$H_1 : \mu > 70$

$\alpha = 0,05$

$RC : z_{1-\alpha} \qquad \Rightarrow RC : z_{0,95} \qquad \Rightarrow RC : z = 1,64$

$\overline{x} = 71,8 \qquad n = 100 \, \sigma = 8,9$

$$z = \frac{71,8 - 70}{\frac{8,9}{10}} = 2,022$$



Se rechaza H_0, es decir, es verdad que la vida promedio hoy en día supera los 70 años.

3) Un fabricante de equipo deportivo ha desarrollado un nuevo sedal sintético para pesca que se considera tiene una resistencia a la ruptura de 8 kilógramos con una desviación stándar de 0,5 kilógramos. Pruébese la hipótesis de que $\mu = 8$ Kg. en contraposición a la alternativa de que $\mu \neq 8$Kg. si se prueba una muestra aleatoria de 50 sedales y se encuentra que tiene una resistencia promedio a la ruptura de 7,8 Kg. Utilice un nivel de significación de 0,01

$H_0 : \mu = 8$
$H_1 : \mu \neq 8$

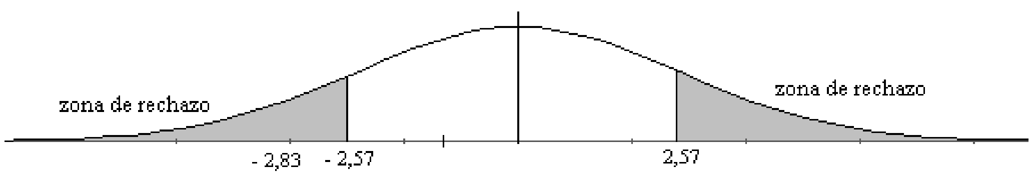
$\alpha = 0,01$

$RC: \; -z_{1-\alpha/2} < z < z_{1-\alpha/2} \qquad \Rightarrow RC: \; -z_{0,995} < z < z_{0,995}$

$\Rightarrow RC: \; -2,57 < z < 2,57$

$\overline{x} = 7,8 \qquad n = 50 \qquad \sigma = 0,5$

$$z = \frac{7,8-8}{\frac{0,5}{\sqrt{50}}} = -2,83$$



Se rechaza H_0, por tanto la resistencia a la ruptura es distinta de 8 Kg

Ejercicios

1) Una empresa eléctrica fabrica focos que tienen una duración que está distribuida aproximadamente en forma normal con una media de 800 horas y una desviación estándar de 40 horas. Pruebe la hipótesis de que $\mu = 800$ horas en contraposición de la alternativa de que $\mu \neq 800$ horas si una muestra aleatoria de 30 focos tiene una duración promedio de 788 horas. Utilice un nivel de significación de 0,04.

2) Un fabricante de cigarros afirma que el contenido promedio de nicotina no excede de 3,5 milígramos , con una desviación stándar de 1,4 milímetros. Para una muestra de 8 cigarros se tiene un contenido promedio de nicotina de 4,2 milígramos .¿Está esto de acuerdo con la afirmación del fabricante ?. Use nivel de significación de 0,05.

3) Las tensiones de ruptura de los cables fabricados por una empresa tienen media de 1800 lb y una desviación estándar de 100 lb. Se desea comprobar si un nuevo proceso de fabricación aumenta dicha tensión media. Para ello se toma una muestra de 50 cables y se encuentra que su tensión media de ruptura es 1850 lb. ¿Se puede afirmar la mejoría del nuevo proceso al nivel de significación del 1%?

4) Se requiere que la tensión de ruptura promedio de un hilo utilizado en la fabricación de material de tapicería sea al menos de 100 psi. La experiencia ha indicado que la desviación estándar de la tensión de ruptura es 2 psi. Se prueba una muestra aleatoria de nueve especímenes, y la tensión de ruptura promedio observada en ella es de 98 psi.

¿Debe aceptarse la fibra como aceptable con $\alpha = 0{,}05$?

Solución

1) Se acepta H_0, es decir, los focos tienen una duración promedio de 800 horas .

2) Se acepta H_0, es decir, es correcta la afirmación del fabricante .

3) Se rechaza H_0, es decir, el nuevo proceso de fabricación aumenta la tensión de ruptura.

4) Se rechaza H_0, es decir, la tensión de ruptura promedio es menor que 100 psi.

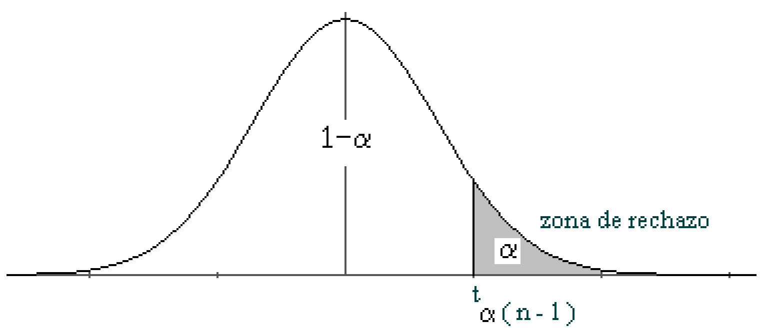
b) Para la media (μ) si la varianza (σ^2)es desconocida

Recuerde que cuando σ^2 es desconocida se usa s^2 y por lo tanto la prueba estadística adecuada es t

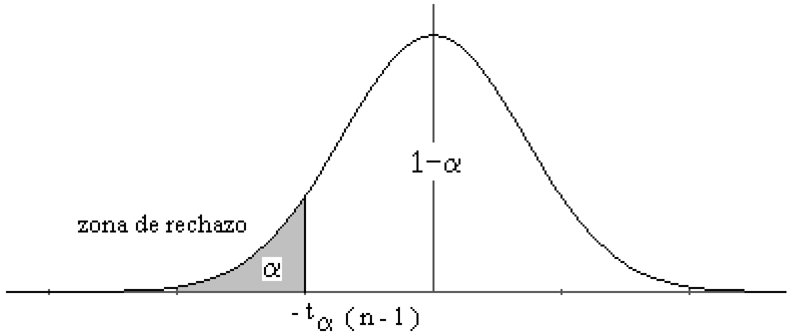
$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

1) Para pruebas de hipótesis unilaterales

$a)$ $\quad H_0 : \theta = \theta_1 \quad$ o $\quad H_0 : \theta \leq \theta_1$
$\quad\;\; H_1 : \theta > \theta_1 \quad\quad\;\; H_1 : \theta > \theta_1$

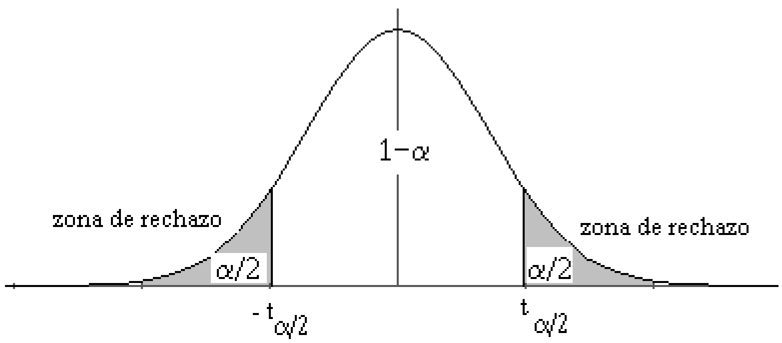


$b)$ $\quad H_0 : \theta = \theta_1 \quad$ o $\quad H_0 : \theta \geq \theta_1$
$\quad\;\; H_1 : \theta < \theta_1 \quad\quad\;\; H_1 : \theta < \theta_1$



2) Para pruebas bilaterales

$H_0: \theta = \theta_1$
$H_1: \theta \neq \theta_1$



Ejemplos

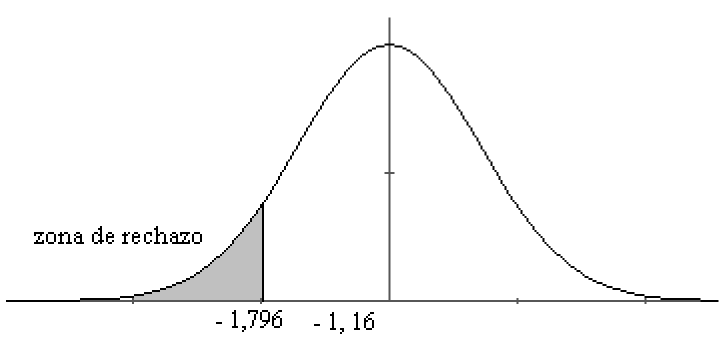
1) Una compañía de electricidad ha publicado cifras acerca de la cantidad anual de kilowtts-hora consumida por varios aparatos para el hogar. Se afirma que la aspiradora consume un promedio de 46 kilowtts-hora al año. Si una muestra aleatoria de 12 hogares incluidos en un estudio planeado indica que las aspiradoras consumen un promedio de 42 kilowtts-hora al año con una desviación stándar de 11,9 kilowtts-hora. ¿Sugiere esto, con un nivel de significación de 0,05 , que las aspiradoras consumen, en promedio, menos de 46 kilowtts-hora al año?. Suponga que la población de kilowtts-hora es normal.

$H_0: \mu = 46$
$H_1: \mu < 46$

$\alpha = 0,05 \qquad \overline{x} = 42 \qquad n = 12 \qquad s = 11,9$

$RC: \ -t_{0,05(11)} = -1,796$

$$t = \frac{42 - 46}{\frac{11,9}{\sqrt{12}}} = -1,16$$



Se acepta H_0, es decir, la muestra elegida no da pruebas que el consumo de kilowtts-hora al año de la aspiradora sea menor que 46.

2) El gerente de producción de una empresa cuyo proceso consiste en llenar cajas de cereal desea saber si efectivamente en cada caja se está depositando, en promedio, los 368 gramos que se supone es lo que la empresa asegura a sus vendedores. Para ello, se selecciona una muestra aleatoria de 25 de estas cajas obteniéndose una media de 364,1 gramos y una desviación stándar de 17,3 gramos. Considere que la distribución de los pesos de las cajas de cereales es normal y trabaje con un nivel de significación de 0,05. ¿Qué decide el gerente de producción?.

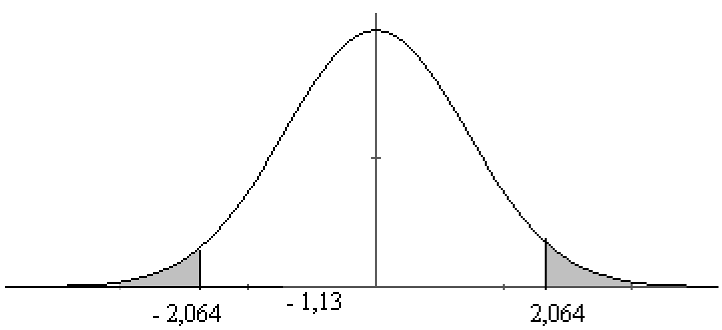
$H_0 : \mu = 368$
$H_1 : \mu \neq 368$

$\alpha = 0,05$

$\overline{x} = 364,1 \qquad n = 25 \qquad s = 17,3$

$RC: -t_{0,025(24)} < t < t_{0,025(24)} \Rightarrow RC: -2,064 < t < 2,064$

$$t = \frac{364,1 - 368}{\frac{17,3}{5}} = -1,13$$



Se acepta H_0, es decir, el gerente de producción puede estar seguro que, en promedio, cada caja contiene 368 gramos de cereal.

3) Suponga que en el mismo ejemplo anterior, del proceso de llenado de cajas de cereal, que la empresa es visitada por un representante de la oficina de protección al consumidor y que le interesa averiguar si las cajas, en promedio, están faltas de peso, es decir, si el peso promedio es inferior a 368 gramos. Considere un nivel de significación de 0,01.

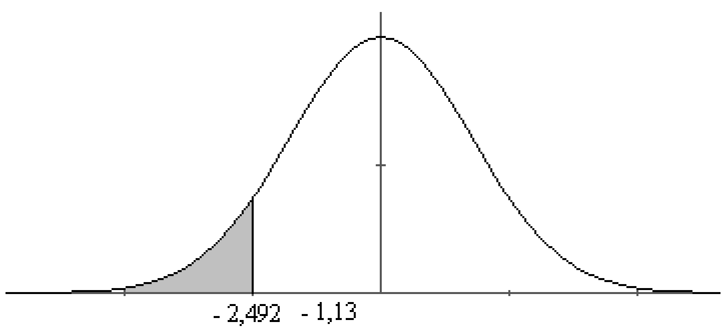
$H_0 : \mu \geq 368$
$H_1 : \mu < 368$

$\alpha = 0,01$

$\overline{x} = 364,1 \qquad n = 25 \qquad s = 17,3$

$RC : -t_{0,01(24)} = -2,492$

$$t = \frac{364,1 - 368}{\frac{17,3}{5}} = -1,13$$



Se acepta H_0, es decir, el representante de la oficina de protección al consumidor puede estar seguro que, en promedio, el peso de cada caja de cereal no es inferior a 368 gr.

Ejercicios

1) Una muestra aleatoria de 36 refrescos de una máquina despachadora automática tiene un contenido promedio de 21,9 decílitros con una desviación estándar de 1,42 decílitros. Pruebe la hipótesis de que $\mu = 22{,}2$ decílitros en contraposición a la hipótesis alternativa, $\mu < 22,2$ decílitros, en el nivel de significación 0,05.

2) Se afirma que un automóvil recorre un promedio anual de más de 20.000 kilómetros. Para probar esta afirmación, se le solicita a una muestra aleatoria de 100 propietarios de automóvil que lleve un registro de los kilómetros que recorren. ¿ Estaría usted de acuerdo con esta afirmación si en la muestra aleatoria resulta un promedio de 23.500 kilómetros y una desviación estándar de 3.900 kilómetros?. Use un nivel de significación de 0,01.

3) En un informe de una investigación de J.M.N. se afirma que los ratones con una vida promedio de 32 meses llegarán hasta casi 40 cuando 40 % de las calorías en su alimentación se reemplacen con vitaminas y proteínas. ¿Hay alguna razón para creer que la vida promedio será inferior a 40 meses si 64 ratones que se han sujetado a esta dieta tienen una vida promedio de 38 meses con una desviación estándar de 5,8 meses?. Utilice un nivel de significación de 0,025.

4) Una empresa eléctrica afirma que un compactador de basura se usa un promedio de 125 horas al año. Si una muestra aleatoria de 49 hogares equipados con compactadores de basura indica un uso promedio de anual de 126,9 horas con una desviación estándar de 8,4 horas, ¿sugiere esto con un nivel de significación de 0,05 , que estos aparatos se usan en promedio más de 125 horas?.

5) En el pasado una máquina ha producido arandelas con un grosor promedio de 0,050 pulgadas. Para determinar si la máquina sigue en buenas condiciones de producción, se toma una muestra de 10 arandelas, que resulta tener un grosor medio de 0,053 pulgadas y una desviación estándar de 0,003 pulgadas. Ensayar la hipótesis de que la máquina está en buenas condiciones de producción al nivel de significación del

a) 0,05
b) 0,01

6) La duración media de una muestra de 100 tubos fluorescentes producidos por una compañía resulta ser 1570 horas, con una desviación estándar de 120 horas. Si μ es la duración media de todos los tubos producidos por la compañía, comprobar la hipótesis $\mu = 1600$ horas contra la hipótesis alternativa $\mu \neq 1600$ horas con un nivel de significación de

a) 0,05
b) 0,01

Solución

1) Se acepta H_0, es decir, $\mu = 22,2$ decílitros.

2) Se rechaza H_0, es decir, un automóvil recorre un promedio anual superior a 20000 Km.

3) Se rechaza H_0, es decir, la vida promedio es inferior a 40 meses.

4) Se acepta H_0, es decir, un compactador de basura se usa en promedio menos de 125 horas al año.

5) a) Se rechaza H_0, es decir, la máquina no está en buenas condiciones de producción.

b) Se acepta H_0, es decir, la máquina está en buenas condiciones de producción.

6) a) Se rechaza H_0, es decir, $\mu \neq 1600$ horas .

b) Se acepta H_0, es decir, $\mu = 1600$ horas .

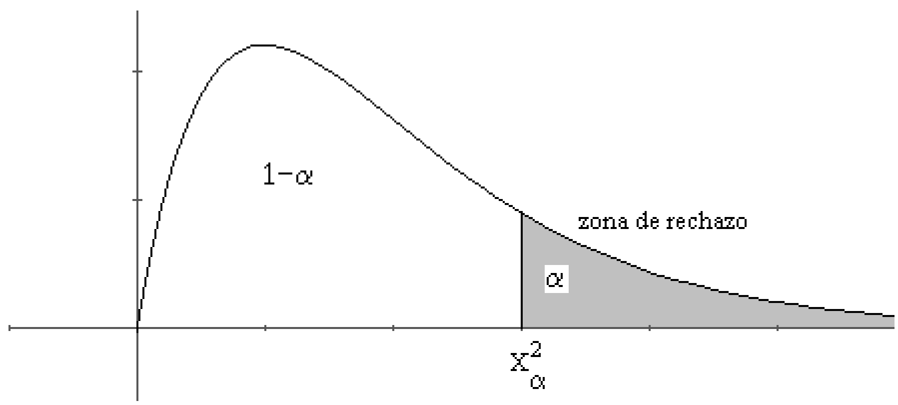
c) Pruebas de hipótesis relacionadas con varianzas

Se utilizan para probar uniformidad de una población. Para ello se usa como prueba estadística la distribución ji cuadrada

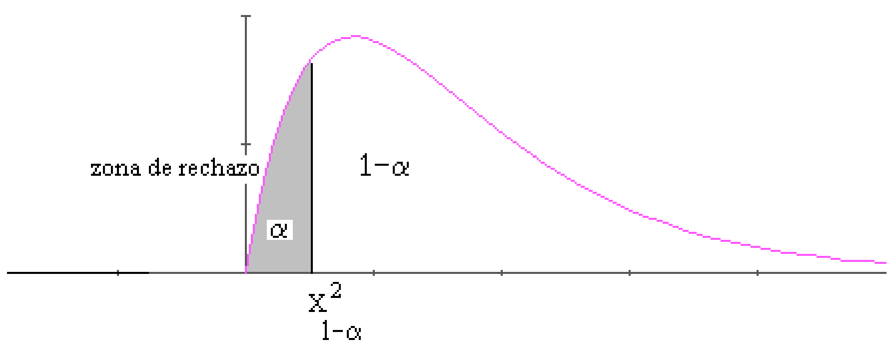
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

1) Para pruebas de hipótesis unilaterales

$a)$ $\quad H_0 : \theta = \theta_1 \quad$ o $\quad H_0 : \theta \leq \theta_1$
$\quad\;\; H_1 : \theta > \theta_1 \quad\;\;\; H_1 : \theta > \theta_1$

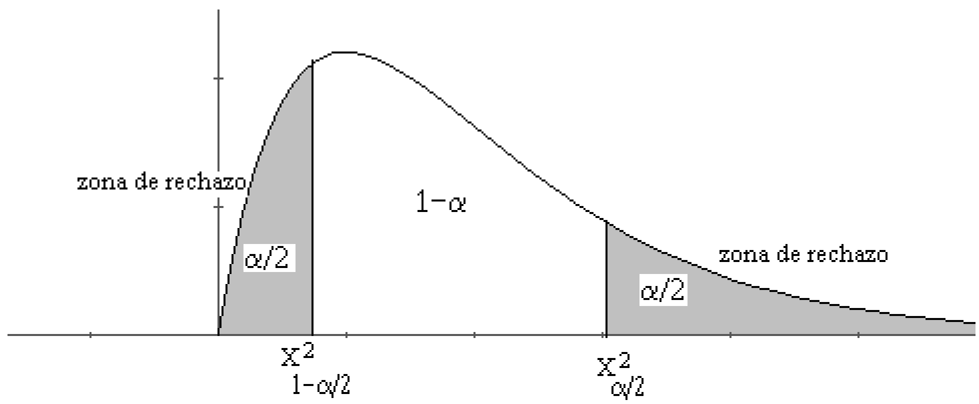


$b)$ $\quad H_0 : \theta = \theta_1 \quad$ o $\quad H_0 : \theta \geq \theta_1$
$\quad\;\; H_1 : \theta < \theta_1 \quad\;\;\; H_1 : \theta < \theta_1$



2) Para pruebas bilaterales

$H_0 : \theta = \theta_1$
$H_1 : \theta \neq \theta_1$



Ejemplos

1) Un fabricante de baterías para automóvil asegura que la duración de sus baterías tiene distribución aproximadamente normal con una desviación stándar de 0,9 años. Si una muestra aleatoria de 10 baterias tiene una desviación stándar de 1,2 años ¿Piensa usted que $\sigma > 0{,}9$ años? Utilice un nivel de significación de 0,05

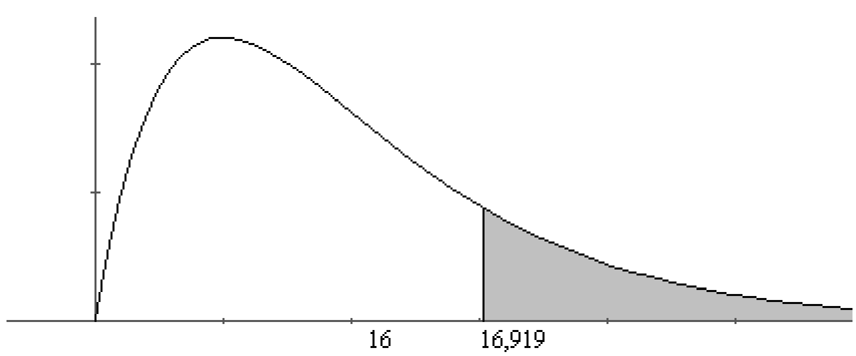
$H_0 : \sigma^2 = 0,81$
$H_1 : \sigma^2 > 0,81$

$\alpha = 0,05$

$s^2 = 1,44 \qquad\qquad n = 10$

$RC : \chi^2_{0,05(9)} = 16,919$

$$\chi = \frac{9 \cdot 1,44}{0,81} = 16$$



No es posible rechazar H_0

2) Se sabe que el contenido de nicotina de una marca de cigarros tiene distribución aproximadamente normal con una varianza de 1,3 milímetros. Pruebe la hipótesis de que $\sigma^2 = 1,3$ en contraposición a la alternativa de que $\sigma^2 \neq 1,3$ si una muestra aleatoria de 8 de estos cigarros tiene una desviación stándar de 1,8 milímetros. Use un nivel de significación de 0,05.

$H_0 : \sigma^2 = 1,3$
$H_1 : \sigma^2 \neq 1,3$

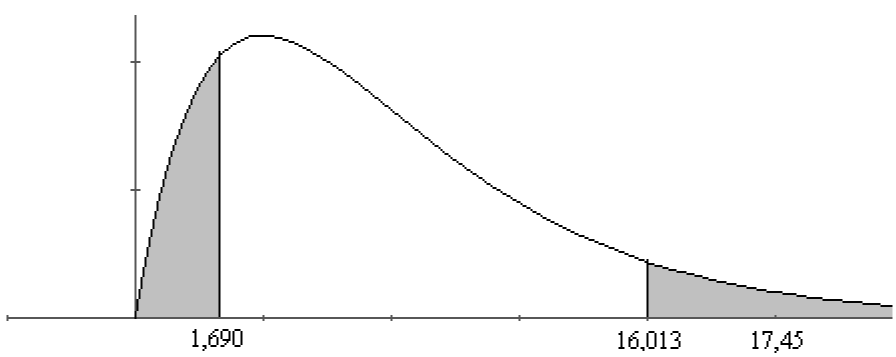
$\alpha = 0,05$

$s^2 = 3,24$ $\qquad\qquad\qquad n = 8$

$RC : \quad \chi^2_{0,975(7)} = 1,690$

$\qquad\quad \chi^2_{0,025(7)} = 16,013$

$$\chi = \frac{7 \cdot 3,24}{0,13} = 17,45$$



Se rechaza H_0, es decir, $\sigma^2 \neq 1,3$

3) Experiencias pasadas indican que el tiempo para que alumnos del último año realicen un examen estandarizado es una v.a normal con una desviación stándar de 6 minutos. Pruebe la hipótesis de que $\sigma = 6$ en contraposición a la alternativa de que $\sigma < 6$ si una muestra aleatoria de 20 studiantes tiene una desviación stándar de 4,51 minutos al realizar este examen. Utilice un nivel de significación de 0,01

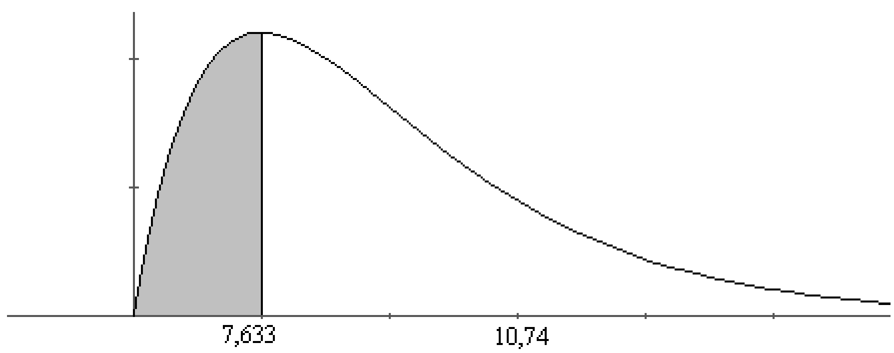
$H_0 : \sigma^2 = 36$
$H_1 : \sigma^2 < 36$

$\alpha = 0,01$

$s^2 = 20,3401$ $\qquad n = 20$

$RC : \chi^2_{0,99(19)} = 7,633$

$$\chi = \frac{19 \cdot 20,3401}{36} = 10,74$$



Con la información de la muestra, no es posible rechazar H_0

__Ejercicios__

1) Se sabe que la capacidad de los recipientes de un determinado lubricante tiene distribución normal con una varianza de 0,03 litros^2. Pruebe la hipótesis de que $\sigma^2 = 0,03$ en contraposición a la alternativa de que $\sigma^2 \neq 0{,}03$ para la muestra aleatoria de 10 recipientes que tiene una desviación estándar de 0,25. Use nivel de significación de 0,01.

2) Se sabe que el contenido de nicotina de una marca de cigarros tiene una distribución aproximadamente normal con una varianza de 1,3 milígramos. Pruebe la hipótesis de que $\sigma^2 = 1,3$ en contraposición a la alternativa de que $\sigma^2 \neq 1,3$ si una muestra aleatoria de 8 de estos cigarros tiene una desviación estándar de 1,8. Use nivel de significación de 0,05.

3) En el pasado la desviación estándar de los pesos de ciertos paquetes de 40 onzas, llenados por una máquina era de 0,25 onzas. Una muestra aleatoria de 20 paquetes dio una desviación estándar de 0,32 onzas. ¿Es el aparente incremento de variabilidad significativa al nivel de significación del

a) 0,05
b) 0,01

4) Se formula la hipótesis de que la desviación estándar del ingreso doméstico anual de cierta comunidad es de 3.000 dólares. En una muestra de 15 hogares aleatoriamente seleccionados, la desviación estándar es 2.000 dólares. Se supone que las cifras de ingreso doméstico de la población siguen una distribución normal. Con base en este resultado muestral, ¿puede rechazarse la hipótesis nula con un nivel de significación del

a) 0,05 ?
b) 0,01 ?

__Solución__

1) Se acepta H_0, es decir, $\sigma^2 = 0,03$

2) Se rechaza H_0, es decir, $\sigma^2 \neq 1,3$

3) a) Se rechaza H_0, es decir, existe un aumento de variabilidad

b) Se acepta H_0, es decir, no existe un aumento de variabilidad

4) a) Se acepta H_0, es decir, $\sigma = 3.000$ dólares

b) Se acepta H_0, es decir, $\sigma = 3.000$ dólares

Autoevaluación

1) Los sistemas de escape de emergencia para tripulaciones de aeronaves 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 50 cm/seg. Se sabe que la desviación estándar de esta rapidez es 2 cm/seg. El experimentador decide especificar un nivel de significación de 0,05. Selecciona una muestra aleatoria de 25 y obtiene una rapidez promedio de combustión de 51,3 cm/seg. ¿A qué conclusiones debe llegar?

2) Se inserta un remache en un agujero. Si la desviación estándar del diámetro del agujero es mayor que 0,01 mm, entonces existe una probabilidad inaceptablemente grande de que el remache no entre en el agujero. Se toma una muestra aleatoria de 15 piezas, y se mide el diámetro del agujero, la desviación estándar es de 0.008 mm. ¿Existe evidencia fuerte que indique que la desviación estándar del diámetro del agujero es mayor que 0,01 mm? Utilice un nivel de significación de 0,01

3) La brillantez de un cinescopio de televisión puede evaluarse midiendo la corriente necesaria para alcanzar un nivel de brillantez particular. Un ingeniero ha diseñado un cinescopio para el que cree que requiere , en promedio, 300 microamperes de corriente para producir el nivel deseado de brillantez. Se toma una muestra de 10 cinescopios y se obtiene una media de 317,2 microamperes con una desviación estándar de 15,7 microamperes. Utilice un nivel de significación de 0,05

4) El contenido de azúcar del almibar de los duraznos enlatados tiene una distribución normal, donde se cree que la varianza es 18 mg^2 . Pruebe la hipótesis $\sigma^2 = 18$ contra la alternativa $\sigma^2 \neq 18$ si al tomar una muestra de 10 latas la desviación estándar es 4,8 mg. Use un nivel de significación de 0,01

5) Un ingeniero civil hace pruebas con la resistencia a la compresión del concreto. Para ello examina 12 especímenes obteniendo una media de 2260 psi y una desviación estándar de 36 psi.Pruebe la hipótesis $\mu = 2270$ psi contra la alternativa $\mu > 2270$ psi . Use un nivel de significación de 0,05

Solución

1) $H_0 : \mu \leq 50$
$H_1 : \mu > 50$

Rechazar H_0, es decir, la rapidez promedio es superior a 50 cm/seg.

2) $H_0 : \sigma^2 \geq (0,01)^2$
$H_1 : \sigma^2 < (0,01)^2$

Aceptar H_0, es decir, no existe evidencia fuerte que indique que la desviación estándar del diámetro del agujero es menor que 0,01 mm

3) $H_0 : \mu = 300$
$H_1 : \mu \neq 300$

Rechazar H_0, es decir, el cinescopio requiere sobre 300 microamperes de corriente para producir el nivel deseado de brillantez.

4) $H_0 : \sigma^2 = 18$
$H_1 : \sigma^2 \neq 18$

Aceptar H_0, es decir, la varianza es de 18 mg^2

5) $H_0 : \mu = 2250$
$H_1 : \mu > 2250$

Aceptar H_0, es decir, la resistencia promedio a la compresión del concreto es de 2250 psi.

Unidad N°5 : Regresión Lineal

El análisis de ***Regresión*** se utiliza para fines de predicción.

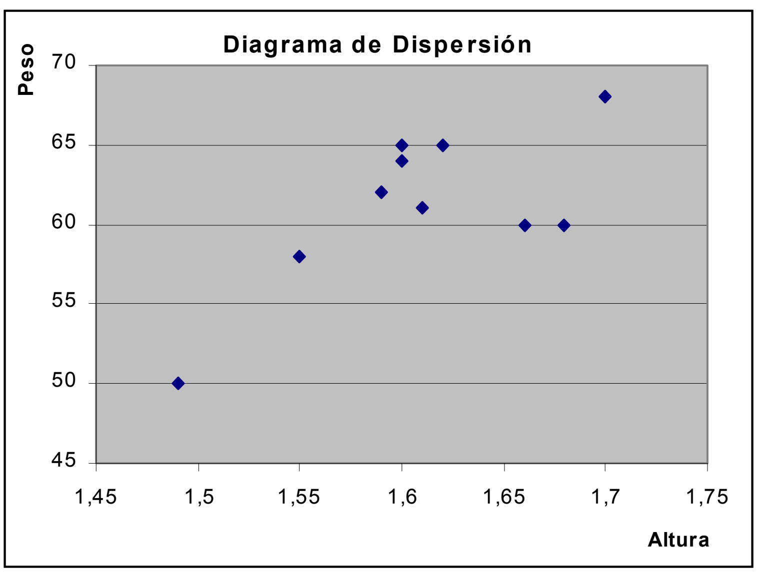
A menudo existen relaciones entre 2 ó más variables, por ejemplo, entre el peso y la estatura de una persona, las horas de estudio y la calificación obtenida, etc. Suele ser deseable expresar tales relaciones en forma matemática determinando una ecuación que conecte a las variables.

Para hallar una ecuación que relacione las variables, el primer paso es recoger datos que muestran valores correspondientes de las variables bajo consideración.

Así por ejemplo, la siguiente tabla muestra las alturas y peso de una muestra de 10 personas:

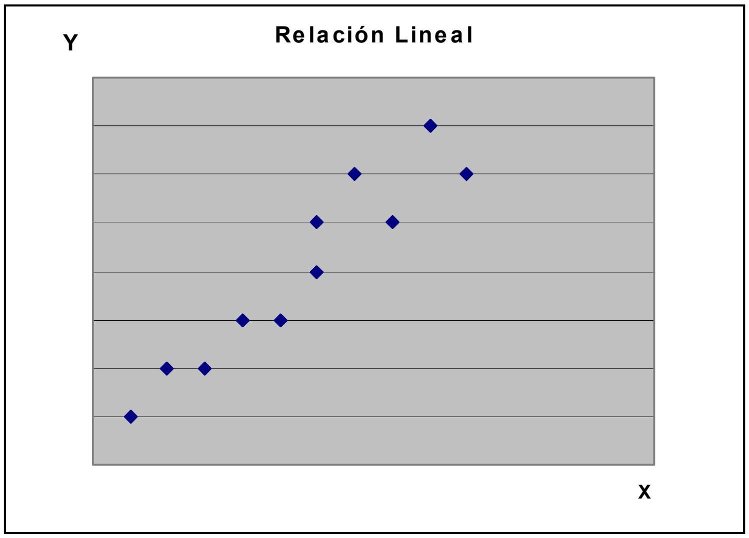
Altura (x)	1.66	1.59	1.62	1.60	1.61	1.49	1.70	1.68	1.55	1.60
Peso (y)	60	62	65	65	61	50	68	60	58	64

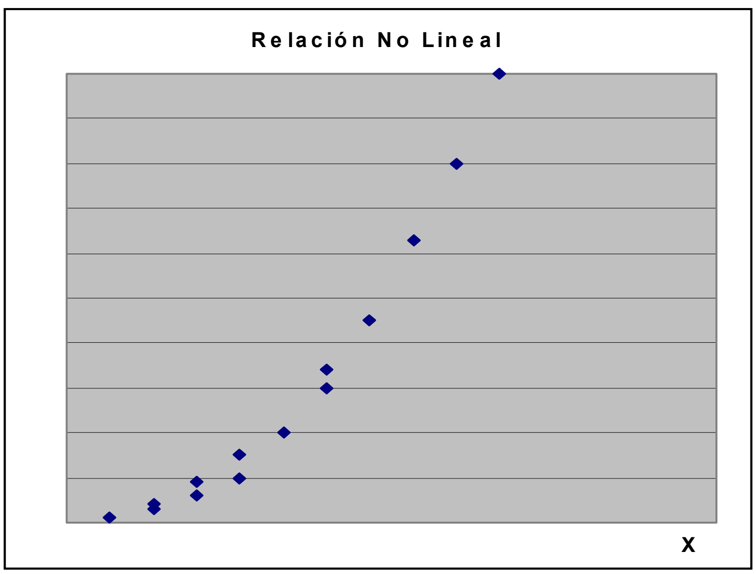
El próximo paso es marcar los puntos (x, y) en un sistema de coordenadas rectangulares, el conjunto de puntos resultantes se denomina ***Diagrama de dispersión***



A partir del Diagrama de Dispersión es posible (a veces), visualizar una curva que aproxima los datos. Tal curva se denomina ***Curva Apoximante.***

Los siguientes diagramas de dispersión:





muestran una relación lineal en el primer caso y una relación no lineal en el segundo.

El problema general de hallar ecuaciones de curvas aproximantes que se ajusten a un conjunto de datos se denomina ***Ajuste de curvas***

Uno de los propósitos principales de la curva de ajuste es estimar una de las variables (la variable dependiente) conocida otra (la variable independiente). El proceso de estimación se conoce como ***Regresión.***

Los tipos más comunes de curvas aproximantes y sus ecuaciones se representan en la siguiente lista:

Línea Recta $\Rightarrow y = a_0 + a_1 x$

Parábola $\Rightarrow y = a_0 + a_1 x + a_2 x^2$

Curva Cúbica $\Rightarrow y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$

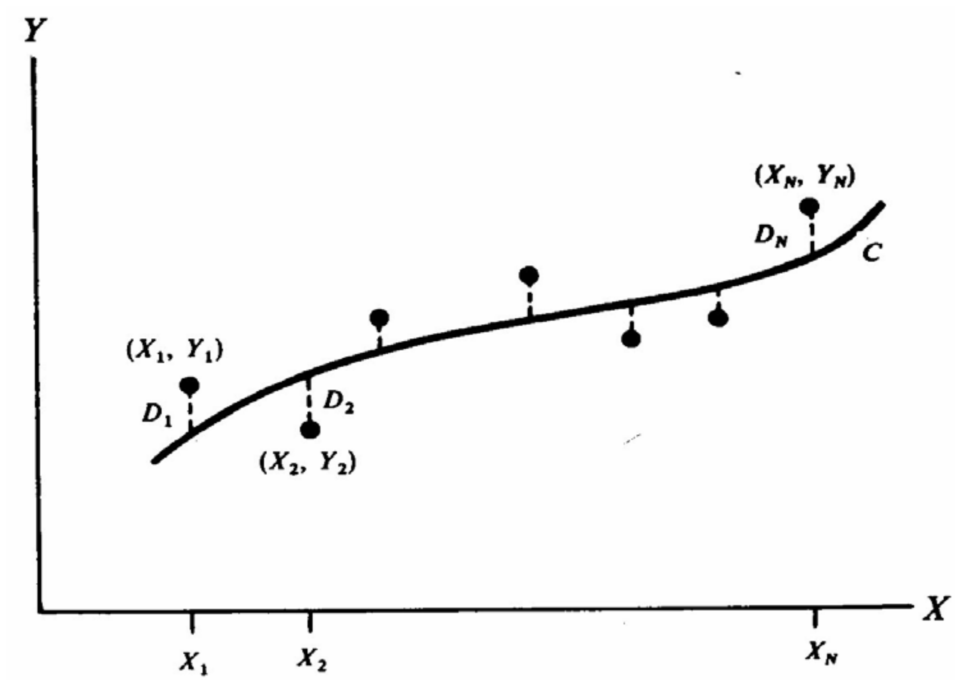
Todas las letras excepto x e y representan constantes. La variable x es la variable independiente y la variable y es la variable dependiente. Aunque esto se puede cambiar, es decir, en algunos casos la variable x será la dependiente y la variable y la independiente.

Para decidir que curva usar es útil observar el diagrama de dispersión. Con el diagrama de dispersión se puede tener una idea aproximada de la relación entre las variables. La relación más sencilla es la ***lineal.***

A menudo se recurre a la intuición personal para dibujar una curva que se ajuste a un conjunto de datos. Este método tiene la desventaja de que diferentes observadores obtendrán distintas curvas y ecuaciones.

Para evitar juicios subjetivos al construir rectas, parábolas u otras curvas aproximantes de ajuste de datos se utiliza el ***Método de Mínimos Cuadrados.***

Dado el siguiente Diagrama de Dispersión:



Una medida de la bondad del ajuste de la curva a los datos dados está proporcionada por la cantidad:

$D_1^2 + D_2^2 + D_3^2 + ... + D_n^2$

Si esta cantidad es pequeña, el ajuste es bueno. Si la cantidad es grande, el ajuste es malo.

Definición: De todas las curvas que aproximan un conjunto de datos, la que tiene la propiedad de que $D_1^2 + D_2^2 + D_3^2 + ... + D_n^2$ es mínimo se llama una ***Curva de Ajuste Optimo.***

Estas diferencias D_i con $i = 1, 2, 3, ..., n$ pueden ser positivas, negativas o iguales a cero.

Una curva que cumpla con la condición de que $D_1^2 + D_2^2 + D_3^2 + ... + D_n^2$ sea mínimo se denomina ***Curva de Mínimos Cuadrados.*** Esta curva puede ser: una recta, una parábola, una parábola cúbica, etc.

La Recta de los Mínimos Cuadrados

El análisis de regresión lineal simple tiene por objeto encontrar la línea recta que mejor se ajuste a los datos, esto significa que se desea encontrar la línea recta para la cual las diferencias entre los valores reales de y y los valores estimados $\hat{y}$ sean lo más pequeñas posible.

La recta de mínimos cuadrados que aproxima el conjunto de puntos: (x_1, y_1), (x_2, y_2), $(x_3, y_3), ..., (x_n, y_n)$ tiene por ecuación la recta $\boldsymbol{y = a_0 + a_1 x}$, donde las constantes a_0 y a_1 se determinan al resolver el siguiente sistema de ecuaciones:

$$\sum y = a_0 n + a_1 \sum x$$

$$\sum x\, y = a_0 \sum x + a_1 \sum x^2$$

que se denominan las ***Ecuaciones Normales*** para la recta de mínimos cuadrados.

Otra forma de determinar estas constantes a_0 y a_1, es a través de las siguientes fórmulas que se deducen de las Ecuaciones Normales:

$$a_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \left(\sum x\right)^2} \qquad a_0 = \overline{y} - a_1 \overline{x}$$

donde: $\overline{x}$ e $\overline{y}$ corresponden al promedio de los datos dados para x e y , respectivamente.

Lo anterior se utiliza cuando x es la variable independiente e y es la variable dependiente.

Si se toma x como la variable dependiente, la recta toma la forma $x = b_0 + b_1 y$, y las ecuaciones normales serían:

$$\sum x = b_0 n + b_1 \sum y$$

$$\sum xy = b_0 \sum y + b_1 \sum y^2$$

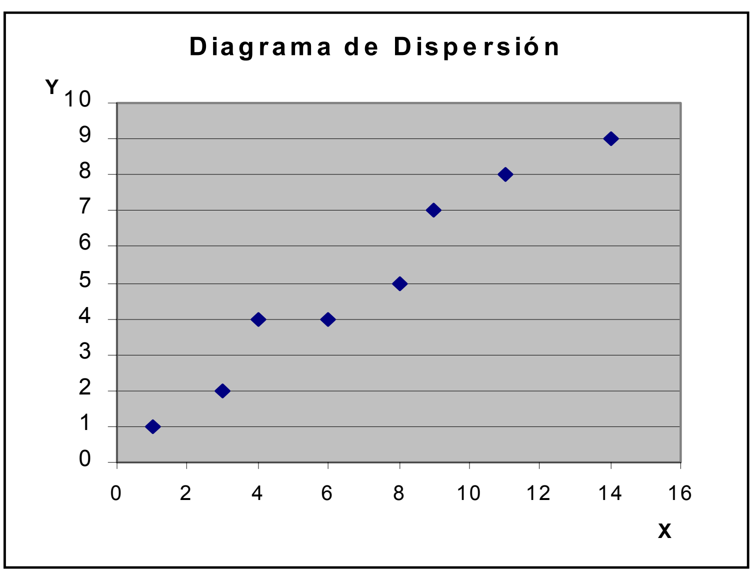
$$b_1 = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - \left(\sum y\right)^2} \qquad b_0 = \overline{x} - b_1 \overline{y}$$

La recta de mínimos cuadrados resultante no es, generalmente, la misma que la obtenida antes.

Ejemplo:

1) Determine la recta de mínimos cuadrados considerando:
a) x como la variable independiente
b) x como la variable dependiente
para la siguiente tabla:

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9



a) x como la variable independiente $\Rightarrow y = a_0 + a_1 x$

Para determinar las constantes usamos las ecuaciones normales:

$$\sum y = a_0 n + a_1 \sum x$$
$$\sum x\,y = a_0 \sum x + a_1 \sum x^2$$

De la tabla se tiene que: $n = 8$ $\quad \sum x = 56$ $\quad \sum y = 40$

$\sum x^2 = 524$ $\quad \sum xy = 364$

Luego, las ecuaciones normales que se deben resolver son:

$$\begin{aligned} 40 &= 8a_0 + 56a_1 / (-7) \\ 364 &= 56a_0 + 524a_1 \end{aligned}$$

$$\begin{aligned} -280 &= -56a_0 - 392a_1 \\ 364 &= 56a_0 + 524a_1 \end{aligned}$$

$$84 = 132a_1$$

$$a_1 = \frac{84}{132} \qquad \Rightarrow a_1 = \frac{21}{33} \qquad \Rightarrow \boldsymbol{a_1 = \frac{7}{11}}$$

Reemplazando $a_1 = \dfrac{7}{11}$ en: $40 = 8a_0 + 56a_1$ se tiene:

$$40 = 8a_0 + 56\left(\frac{7}{11}\right) \qquad / \div 8$$

$$5 = a_0 + \frac{49}{11} \qquad \Rightarrow a_0 = 5 - \frac{49}{11} \quad \Rightarrow \boldsymbol{a_0 = \frac{6}{11}}$$

Luego, la recta de mínimos cuadrados es: $\boxed{\boldsymbol{y = \frac{6}{11} + \frac{7}{11}x}}$

La ecuación determinada se puede graficar sobre el diagrama de dispersión de los datos.

b) x como la variable dependiente $\Rightarrow x = b_0 + b_1 y$

Para determinar las constantes usamos las ecuaciones normales:

$$\sum x = b_0 n + b_1 \sum y$$
$$\sum x\, y = b_0 \sum y + b_1 \sum y^2$$

De la tabla se tiene que: $n = 8$ $\quad \sum x = 56 \quad \sum y = 40$

$\sum y^2 = 256 \quad \sum xy = 364$

Las ecuaciones normales son:

$$\begin{aligned} 56 &= 8b_0 + 40b_1 \; /(-5) \\ 364 &= 40b_0 + 256b_1 \end{aligned}$$

$$\begin{aligned} -280 &= -40b_0 - 200b_1 \\ 364 &= 40b_0 + 256b_1 \end{aligned}$$

$$84 = 56b_1$$

$$b_1 = \frac{84}{56} \Rightarrow \boldsymbol{b_1 = \frac{3}{2}}$$

Reemplazando $b_1 = \frac{3}{2}$ en: $56 = 8b_0 + 40b_1$ se tiene:

$$56 = 8b_0 + 40\left(\frac{3}{2}\right) \qquad / \div 8$$

$$7 = b_0 + \frac{15}{2} \qquad \Rightarrow b_0 = 7 - \frac{15}{2} \Rightarrow \boldsymbol{b_0 = -\frac{1}{2}}$$

Luego, la recta de mínimos cuadrados es: $\boxed{\boldsymbol{x = -\frac{1}{2} + \frac{3}{2}y}}$

La ecuación determinada se puede graficar sobre el diagrama de dispersión de los datos.

Las rectas de mínimos cuadrados que hemos determinado nos sirven para estimar, basados en datos de una muestra, el valor de una variable y correspondiente a un valor dado de la variable x. La curva resultante se denomina ***Curva de Regresión de y sobre x***, ya que y se estima a partir de x.

En el ejemplo anterior, la ecuación de la curva de regresión de y sobre x es:

$$\boxed{\boldsymbol{y = \frac{6}{11} + \frac{7}{11}x}}$$

Podemos estimar el valor y para $x = 5$ de la siguiente forma:

$$y = \frac{6}{11} + \frac{7}{11}(5) \Rightarrow y = \frac{6}{11} + \frac{35}{11} \Rightarrow \hat{y} = \frac{41}{11} \approx 3.73$$

Para $x = 4$ $\qquad \Rightarrow y = \frac{6}{11} + \frac{7}{11}(4) \qquad \Rightarrow y = \frac{6}{11} + \frac{28}{11} \Rightarrow \hat{y} = \frac{34}{11} \approx 3.09$

La ecuación $x = -\frac{1}{2} + \frac{3}{2}y$ permite estimar el valor de x a partir de un valor de y. Esta ecuación se denomina ***Ecuación de Regresión de x sobre y***.

Los valores estimados a través de las ecuaciones encontradas no necesariamente corresponden a los valores dados en la tabla.

Las ecuaciones de las rectas de regresión $y = a_0 + a_1 x$ y $x = b_0 + b_1 y$ se intersectan en un punto llamado ***Centroide*** que se denota por $(\overline{x}, \overline{y})$, donde:

$$\overline{x} = \frac{\sum x}{n} \qquad\qquad \overline{y} = \frac{\sum y}{n}$$

Para el ejemplo anterior, se tiene que el centroide es:

$$\overline{x} = \frac{56}{8} = 7 \qquad\qquad \overline{y} = \frac{40}{8} = 5 \qquad\qquad \therefore \text{Centroide} = (\overline{x}, \overline{y}) = (7, 5)$$

El método de regresión responde a tres tipos de objetivos:

1) Estudiar si ambas variables están relacionadas

2) Determinar que tipo de relación, si existe, las une

3) Predecir los valores de una variable a partir de valores conocidos de la otra.

Conocer el grado de relación existente entre ambas variables, permitirá saber si la predicción realizada con el modelo matemático establecido, es buena o mala.

Para medir el grado de relación existente entre la variable independiente y la variable dependiente, lo que más se utiliza es el ***Coeficiente de Correlación Lineal*** (r de Pearson), cuyo método abreviado de cálculo está dado por la siguiente fórmula:

$$r=\frac{n\sum xy-\left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2-\left(\sum x\right)^2\right]\left[n\sum y^2-\left(\sum y\right)^2\right]}}$$

El valor de r se encuentra en el intervalo $[-1,1]$

Si $r=0$, entonces no existe correlación entre las variables

Si $r=-1$, entonces la correlación es perfecta y negativa

Si $r=1$, entonces la correlación es perfecta y positiva

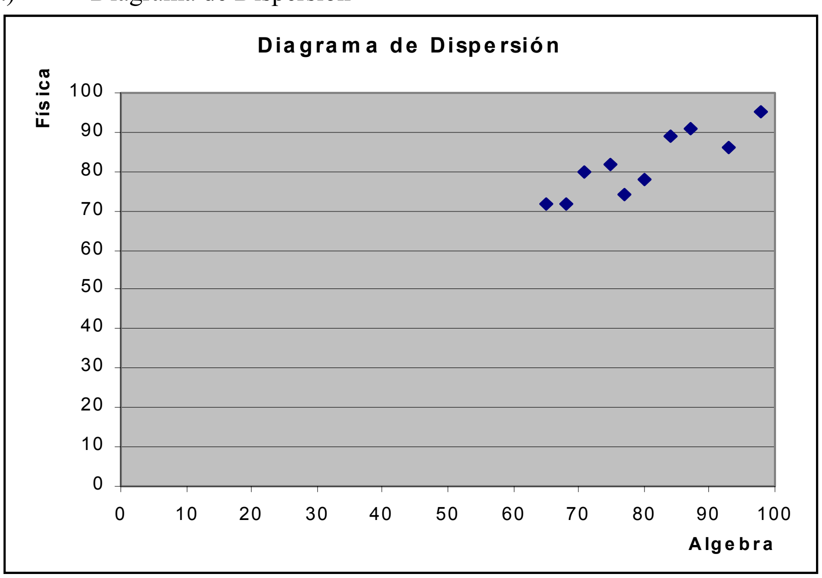
Si $r\in\left]-0.5,0.5\right[$ entonces la correlación es mala

Si $r\notin\left]-0.5,0.5\right[$ entonces la correlación es buena

Ejemplo: La siguiente tabla representa las notas en Algebra y Física de 10 estudiantes elegidos al azar:

Algebra(x)	75	80	93	65	87	71	98	68	84	77
Física(y)	82	78	86	72	91	80	95	72	89	74

a) Diagrama de Dispersión



De los datos dados se tiene que: $\sum x = 798$ $\sum y = 819$

$n = 10$ $\sum xy = 66045$ $\sum x^2 = 64722$ $\sum y^2 = 67675$

b) Determine la recta de regresión de y sobre x $\Rightarrow y = a_0 + a_1 x$

Para determinar las constantes a_0 y a_1, tenemos:

$$\sum y = a_0 n + a_1 \sum x$$
$$\sum x\, y = a_0 \sum x + a_1 \sum x^2$$

Luego, el sistema de ecuaciones que debemos resolver es el siguiente:

$$\begin{aligned} 819 &= 10a_0 + 798a_1 \quad /(399) \\ 66045 &= 798a_0 + 64722a_1 \quad /(-5) \end{aligned}$$

$$\begin{aligned} 326781 &= 3990a_0 + 318402a_1 \\ -330225 &= -3990a_0 - 323610a_1 \end{aligned}$$

$$-3444 = -5208a_1 \qquad \Rightarrow a_1 = 0.66$$

Luego, reemplazando a_1 en: $819 = 10a_0 + 798a_1$ se tiene que:

$$a_0 = \frac{819 - 798(0.66)}{10} \qquad \Rightarrow a_0 = 29.23$$

Por lo tanto, la ecuación de la recta de regresión de y sobre x es:

$y = 29.23 + 0.66x$

c) Determine el centroide $(\overline{x}, \overline{y})$

$$\overline{x} = \frac{798}{10} = 79.8 \;\; \overline{y} = \frac{819}{10} = 81.9$$

Luego, el centroide es $(\overline{x}, \overline{y}) = (79.8, 81.9)$

d) Halle el coeficiente de correlación lineal

$$r = \frac{10(66045) - 798(819)}{\sqrt{[10(64722) - (798)^2][10(67675) - (819)^2]}}$$

$r = \dfrac{6888}{\sqrt{62381424}} \approx 0.87$ Por lo tanto, la correlación entre las variables es buena.

e) Estimar el valor de y para $x = 70$ y $x = 65$

$x = 70$

$y = 29.23 + 0.66(70) = 75.43$

El valor estimado para $x = 70$ es $\hat{y} = 75$

$x = 65$

$y = 29.23 + 0.66(65) = 72.13$

El valor estimado para $x = 65$ es $\hat{y} = 72$

f) Si un estudiante tiene 75 puntos en Algebra. ¿Cuál es su nota esperada en Física?

La nota esperada en Física es:

$\hat{y} = 29.23 + 0.66(75)$
$\hat{y} = 78.73$
$\hat{y} \approx 79$ puntos

Ejercicios

1) Determine para los datos del ejemplo anterior la ecuación de regresión de x sobre y

2) Dados los siguientes datos en forma de pares (x, y): $(2,2)$, $(4,3)$, $(5,3)$, $(6,4)$, $(7,6)$, $(8,5)$, $(9,7)$, $(10,6)$

a) Dibujar el diagrama de dispersión
b) Hallar la ecuación de la recta de regresión de y sobre x
c) Estimar los valores de y para $x = 5,\ x = 8,\ x = 14$

3) En la siguiente tabla se presentan datos que relacionan el número de semanas de experiencia (x) de un trabajador y el número de artículos defectuosos (y) elaborado por cada uno de ellos:

x	7	9	9	14	8	12	10	4	2	11	8	5	4	6
y	26	20	28	16	23	18	24	26	38	22	32	25	35	30

a) Trace el diagrama de dispersión
b) Determine la ecuación de la recta de regresión de y sobre x
c) Grafique la recta de regresión de y sobre x
d) Estime el número de artículos defectuosos para empleados que tienen:
- Tres semanas de experiencia laboral
- Doce semanas de experiencia laboral
e) Calcule el coeficiente de correlación lineal e interprete
f) Determine el centroide de los datos dados

4) El gerente de personal de una empresa intuye que quizás exista relación entre el ausentismo laboral y la edad de los trabajadores. Desea tomar la edad de los trabajadores para desarrollar un modelo de predicción de días de ausencia durante un año laboral. Se seleccionó una muestra aleatoria de 10 trabajadores y se obtuvo los siguientes datos:

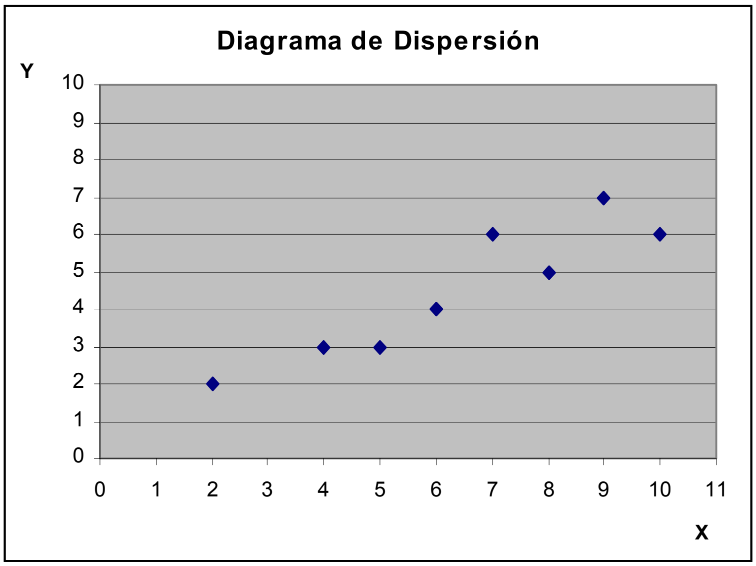
Trabajador	Edad(x)	Días Ausentes(y)
1	27	15
2	61	6
3	37	10
4	23	18
5	46	9
6	58	7
7	29	14
8	36	11
9	64	5
10	40	8

a) Diagrama de Dispersión
b) Determine la recta de regresión de y sobre x
c) Calcule el coeficiente de correlación lineal e interprete
d) Estime los días de ausentismo laboral para trabajadores que tienen 25 años, 34 años y 50 años de edad

Solución

1) $x = -14,39 + 1,15y$

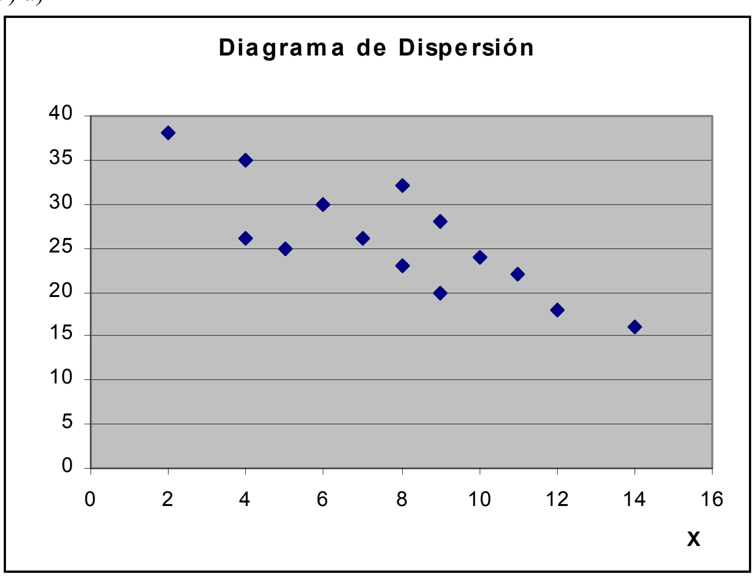
2) a)



b) $y = 0,63 + 0,61x$

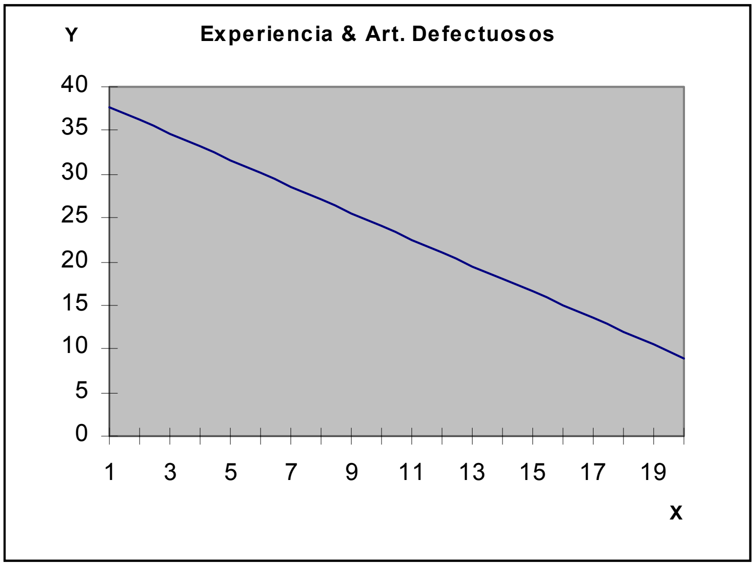
c) $\hat{y} = 3,7$ $\hat{y} = 5,5$ $\hat{y} = 9,2$ (Respectivamente)

3) a)



b) $y = 37,748 - 1,518x$

c) Gráfico de la recta de regresión: $y = 37,748 - 1,518x$



d) $x = 3 \Rightarrow \hat{y} = 33,194 \approx 33$

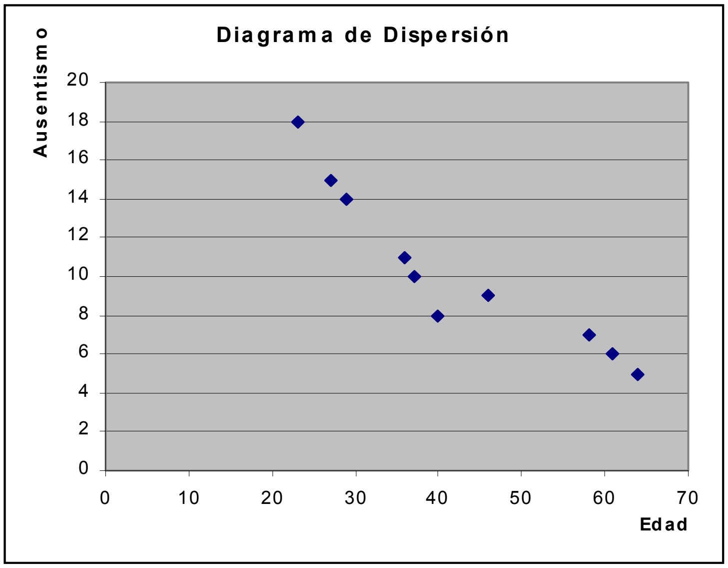
$x = 12 \Rightarrow \hat{y} = 19,532 \approx 20$

e) $r = -0,818$

El coeficiente de correlación lineal $r = -0,818$ nos indica que la correlación lineal entre las variables es buena y negativa, es decir, a mayor experiencia laboral menos artículos defectuosos elabora un trabajador.

f) Centroide $(\overline{x}, \overline{y}) = (7,786; 25,929)$

4) a)



b) $y = 21,583 - 0,268\,x$

c) $r = -0,93$ La correlación entre la edad y el ausentismo laboral es muy buena y negativa, es decir, a mayor edad menos días de ausencia laboral.

d) $x = 25 \text{ años} \Rightarrow \hat{y} = 14,883 \approx 15 \text{ días}$

$x = 34 \text{ años} \Rightarrow \hat{y} = 12,471 \approx 12 \text{ días}$

$x = 50 \text{ años} \Rightarrow \hat{y} = 8,183 \approx 8 \text{ días}$

Análisis de Residuos

El análisis de residuos sirve para verificar si el modelo lineal es el que mejor se ajusta a los datos dados.

Se define un ***residuo*** (e_i) como la diferencia entre el valor observado y y el valor estimado $\hat{y}$, es decir,

$\boxed{e_i = y_i - y'_i}$ donde y_i = valor observado

y'_i = valor estimado

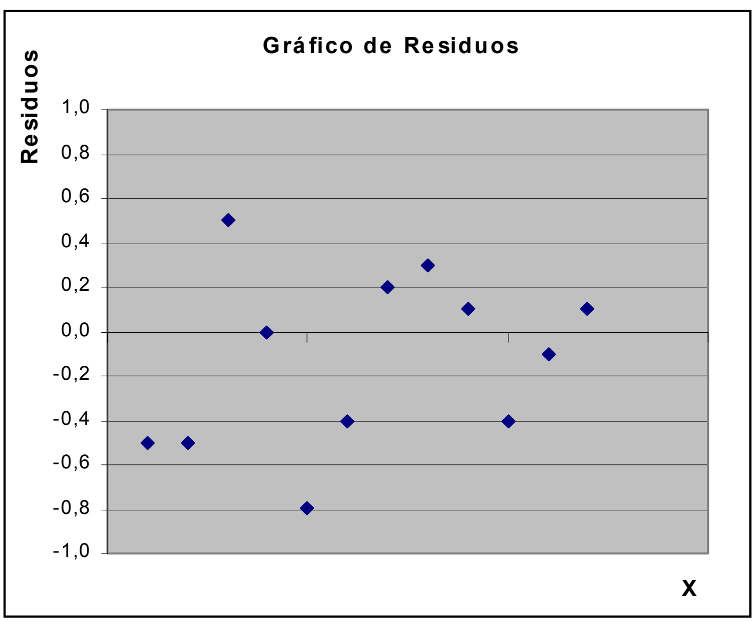
El análisis de residuos nos permite llegar a conclusiones tales como:

a) La función de regresión es lineal

b) La función de regresión no es lineal

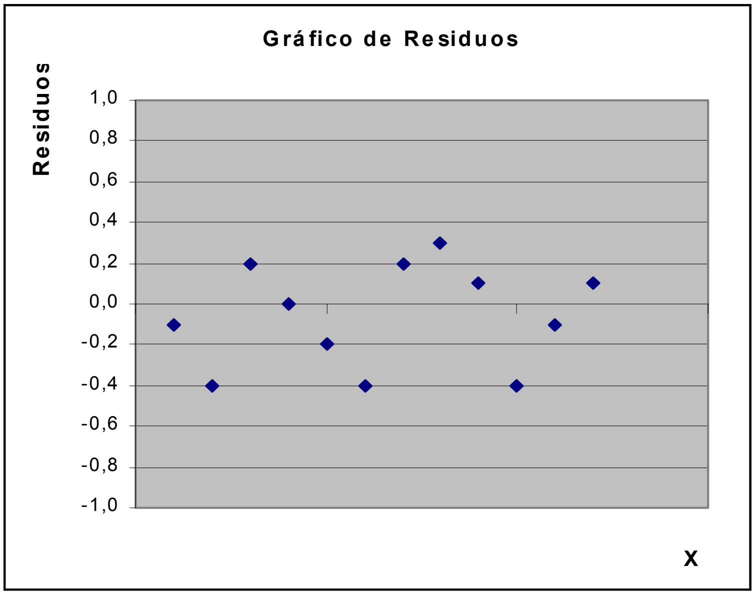
c) El modelo de regresión lineal se ajusta a todas excepto una o varias observaciones atípicas. Estas observaciones atípicas pueden no considerarse si el número de datos es grande (mayor que 30).

La forma más común de enfrentar el problema del análisis de residuos, es mediante un estudio gráfico de ellos. Para graficar los residuos se considera el siguiente gráfico:



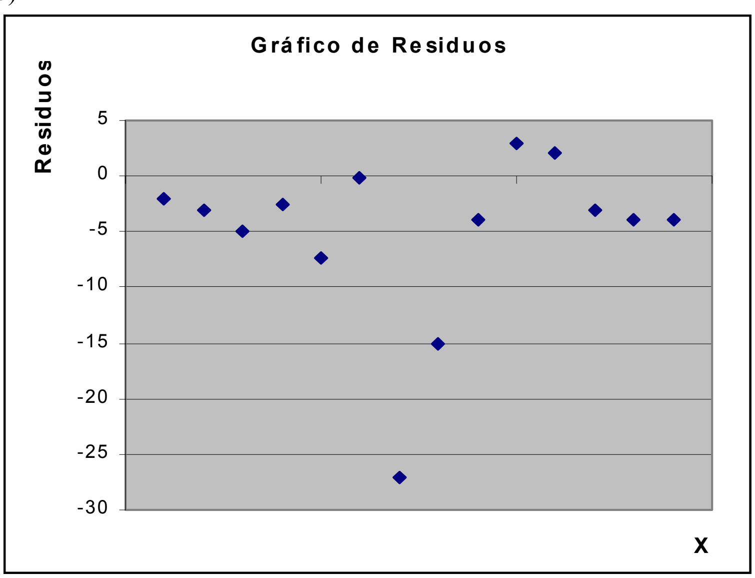
Las siguientes figuras, muestran diferentes situaciones que se presentan con cierta frecuencia:

a)



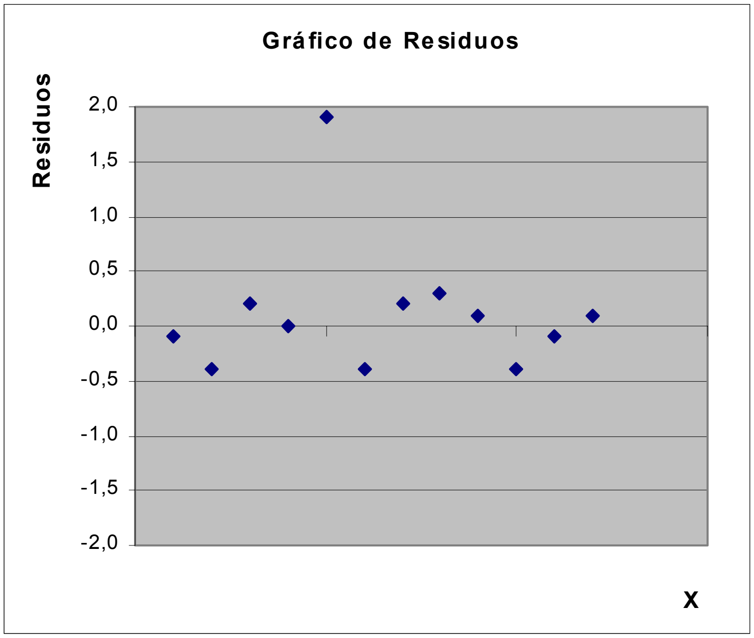
La figura anterior muestra un caso típico de residuos cuando el modelo lineal es adecuado. Todos los residuos tienden a caer en una banda horizontal centrada alrededor del cero.

b)



La figura anterior indica una desviación clara de la linealidad, sugiriendo la necesidad de ajustar una función de regresión no lineal.

c)



La figura anterior presenta una observación atípica, es decir, se escapa del modelo lineal que tienen los otros datos. La influencia de estos puntos atípicos, será mayor si el número de datos es pequeño (menor o igual a 30).

Ejemplo: Dada la siguiente tabla y la recta de regresión de y sobre x:

x	65	63	67	64	68	62	70	66	68	67	69	71
y	68	66	68	65	69	66	68	65	71	67	68	70

$$y = 35,82 + 0,476x$$

Determine:

a) Los valores estimados de y
b) Los residuos e_i para cada caso
c) Represente gráficamente los residuos
d) ¿Qué puede concluir de este gráfico?

Solución:

a) Los valores estimados de y, que aparecen en la tabla, se determinan reemplazando x en la recta dada:

$$y = 35,82 + 0,476x$$

Por ejemplo, para $x = 65$ se tiene que $y = 35,82 + 0,476(65) \Rightarrow y' = 66,8$

El mismo procedimiento se debe realizar para los demás valores de x

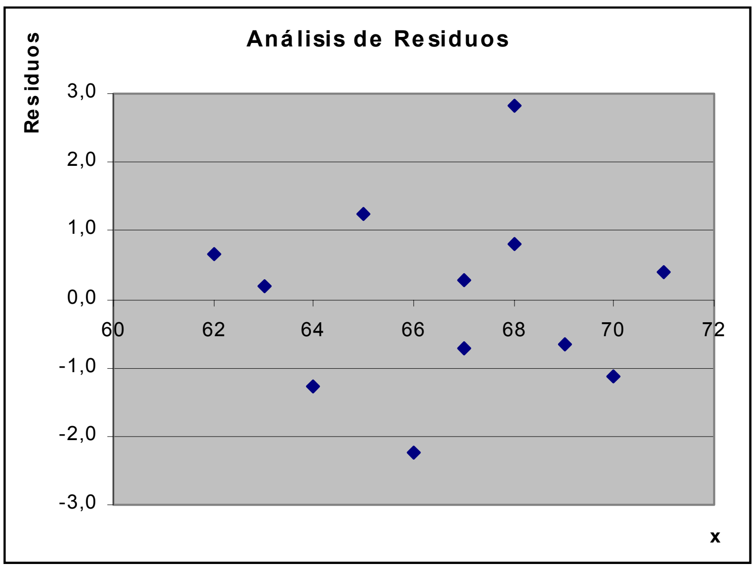
b) Los residuos e_i , que aparecen en la tabla, se determinan de la siguiente forma:

Para $y = 68$ se tiene: $e_i = y - y'$
$e_i = 68 - 66,8$
$e_i = 1,2$

El mismo procedimiento se debe realizar para los demás valores de y

x	65	63	67	64	68	62	70	66	68	67	69	71
y	68	66	68	65	69	66	68	65	71	67	68	70
y'	$66,8$	$65,8$	$67,7$	$66,3$	$68,2$	$65,3$	$69,1$	$67,2$	$68,2$	$67,7$	$68,7$	$69,6$
e_i	$1,2$	$0,2$	$0,3$	$-1,3$	$0,8$	$0,7$	$-1,1$	$-2,2$	$2,8$	$-0,7$	$-0,7$	$0,4$

c)



d) Los residuos nos indican que la recta de regresión dada en algunos casos no es la mejor estimadora para y. Existen 5 puntos que se escapan del intervalo $[-1,1]$

Ejercicio

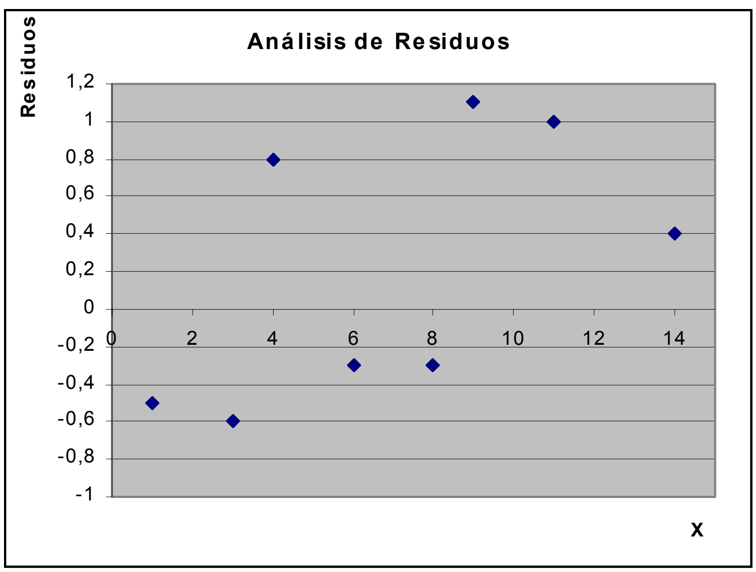
1) Dada la siguiente información, ¿qué puede concluir a través del análisis de residuos?

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9
y'	$1,5$	$2,6$	$3,2$	$4,3$	$5,3$	$5,9$	$7,0$	$8,6$

Solución:

1)

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9
y'	$1,5$	$2,6$	$3,2$	$4,3$	$5,3$	$5,9$	$7,0$	$8,6$
e_i	$-0,5$	$-0,6$	$0,8$	$-0,3$	$-0,3$	$1,1$	$1,0$	$0,4$



Los residuos son muy grandes para los datos dados. Por lo tanto, no existe una relación lineal entre los datos dados.

Autoevaluación

1) Dada la siguiente tabla:

x	3	5	6	8	9	11
y	2	3	4	6	5	8

Determine:

a) Diagrama de Dispersión

b) La ecuación de la recta de regresión de y sobre x. Grafíquela

c) El coeficiente de correlación lineal. Interprete.

d) Estime los valores de y, para $x = 4$ y $x = 6$

2) Dada la siguiente tabla:

Hrs. Estudio (x)	20	16	34	23	27	32	18	22	20
Nota Examen (y)	64	61	84	70	88	92	72	77	66

a) Elabore el Diagrama de Dispersión

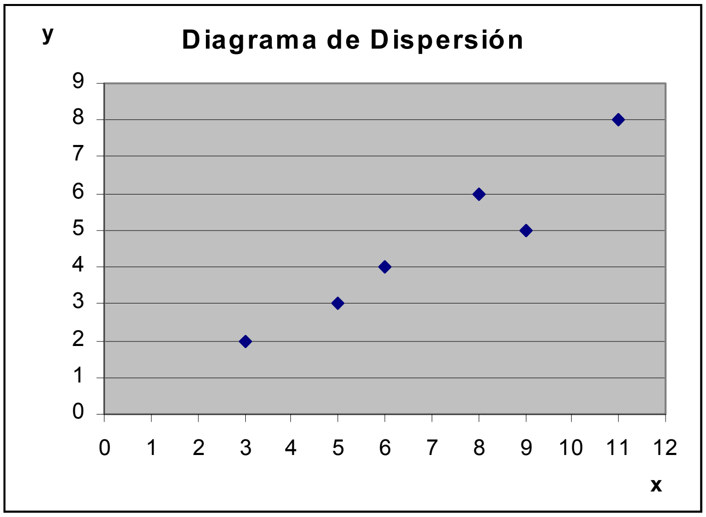
b) Determine el Coeficiente de Correlación Lineal. Concluya.

c) Determine la recta de regresión de y sobre x. Grafique.

d) Estime la nota de Examen para un alumno que estudió 25 horas

Solución

1) a)

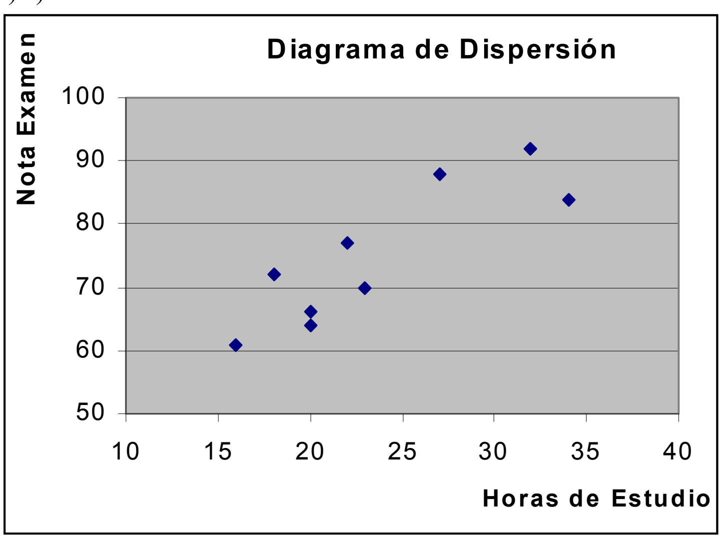


b) $y = -\dfrac{1}{3} + \dfrac{5}{7}x$

c) $r = 0,958$ La correlación entre las variables es buena.

d) $x = 4 \Rightarrow y = 2,52$ $x = 6 \Rightarrow y = 3,95$

2) a)



b) $r = 0,8675$ La correlación entre las variables es buena.
c) $y = 38,61 + 1,54\,x$
d) $x = 25$ $y = 77$