PRÁCTICO 1

EJERCICIO 1

A) NO B) 20% C) 40% D) 70%

EJERCICIO 2: Obtengo las frecuencias relativas y absolutas para las distintas categorías

	F. A	F. R
J	10	10/60 = 1/6
F	9	9/60
B	7	7/60
M	4	4/60
T	3	3/60
N	6	6/60 = 1/10
O	21	21/60

EJERCICIO 3: (I) 0, 20, 40, 50, 60, 80, 100 (II) 0, 48, 49, 50, 51, 52, 100
(III) 0, 1, 2, 50, 98, 99, 100

A) Promedio de cada uno de ellos

(I) $\bar{x} = 50$ (II) $\bar{x} = 50$ (III) $\bar{x} = 50$

B) ¿Cuál tiene mayor DE y cuál menor?

(I) $S = \sqrt{\frac{24500}{6} - n.\bar{x}^2} = \sqrt{\frac{24500}{6} - \frac{7.2500}{6}} = \sqrt{\frac{7000}{6}} = \sqrt{1166,66}$
$= 34,157$

(II) $S = \sqrt{\frac{22510 - 7.2500}{6}} = \sqrt{\frac{5010}{6}} = \sqrt{835} = 28,89$

(III) $S = \sqrt{\frac{31910 - 7.2500}{6}} = \sqrt{2401,66} = 49,006$

MAYOR D.E. = (III) MENOR D.E = (II)

EJERCICIO 4: Un profesor toma un examen con tres preguntas y asigna un punto a cada una de ellos. Un 30% de la clase consigue 3 puntos, un 50% 2 puntos, un 10% 1 punto, y el 10% restante 0 puntos.

A) Si la clase se compone de 10 alumnos ¿Cuál es la nota promedio?

$$\bar{x} = \frac{3.3 + 5.2 + 1 + 0}{10} = 2$$

B) Si la clase se compone 20 alumnos ¿Cuál es la nota promedio

$$\bar{x} = \frac{6.3 + 10.2 + 2}{20} = 2$$

C) ¿Podría decir el valor de la nota promedio sin saber la cantidad de alumnos?

$$\bar{x} = 0,3.3 + 0,5.2 + 0,10.1 = 2;$$

Scanned by CamScanner

EJERCICIO 5 Tres profesores están comparando notas de finales. Cada profesor tiene 99 alumnos.
En la clase A un estudiante obtuvo 1 punto, otro obtuvo 99 y el resto 50 puntos.
En la clase B, 49 estudiaron sacaron 1, 1 obtuvo 50 y 49 sacaron 99.
En la clase C, 1 estudiante sacó 1 punto, otro sacó 2, otro 3 y así sucesivamente hasta 99

PROMEDIOS

$$\bar{X}_A = \frac{(1).1 + (1).99 + (97).50}{99} = 50$$

$$\bar{X}_B = \frac{49 + 50 + 49.99}{99} = 50$$

$$\bar{X}_C = \frac{\sum_{i=1}^{n} i}{99} = \frac{n(n+1)}{2} \cdot \frac{1}{99} = \frac{n(n+1)}{2.99} = \frac{99.100}{2.99}$$

$$= \frac{100}{2} = 50$$

DESVIACIÓN ESTÁNDAR

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

$$S_A = \sqrt{\frac{1 + 99^2 + 97.50^2 - 99.50^2}{98}} = \sqrt{49} = 7$$

$$S_B = \sqrt{\frac{49 + 50^2 + 49.99^2 - 99.50^2}{98}} = \sqrt{2401} = 49$$

$$S_C = \sqrt{\frac{\sum_{i=1}^{n} i^2 - n\bar{x}}{n-1}} = \sqrt{\frac{\frac{n(n+1)(2n+1)}{6} - n.\bar{x}}{n-1}} = \sqrt{\frac{n(n+1)(2n+1)}{6(n-1)} - \frac{n.\bar{x}}{n-1}} =$$

$$= \sqrt{\frac{99.100.199}{588} - 50,5} = \sqrt{3300,01} = 57,446;$$

EJERCICIO 6: Datos sobre fuerza en una sola pierna para carga pesada de trabajo. (...)

A) Calcule e interpreta la media y mediana muestrales
Tamaño muestral $n = 20$; $\sum x_i = 3856$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{3856}{20} = 192,8; \quad \tilde{X} = \frac{X_{n/2} + X_{n/2+1}}{2} = 189;$$

B) Suponga que la última observación fue 211 en vez de 254 ¿Como cambiarían la media y la mediana?

$$\bar{X}: \bar{X} = \frac{\sum x_i - 254}{19} = \frac{3856 - 254}{19} = 189,58;$$

$$\tilde{X} = X_{n\,div\,2+1} = X_{10} = 187;$$

C) La MEDIA MUESTRAL α-RECORTADA consiste en calcular el promedio muestral excluyendo de la muestra el α% de las observaciones más pequeñas y el α% de las más grandes.
Calcule una media sin el primer dato ni el último ¿Qué porcentaje de recorte es?

$$\bar{X} = \frac{\sum x_i - 254 - 160}{18} = \frac{3856 - 254 - 160}{18} = 191,\hat{2} = \bar{X}_{5\%}$$

20 — 100%
1 — X = 5%

D) Calcule la media muestral 20% recortada:

100% — 20
20% — X = 4

$$\bar{X}_{20\%} = \frac{179 + 180 + 180 + 181 + 183 + 187 + 191 + 194 + 200 + 202 + 204 + 205}{12} = 190,5$$

Scanned by CamScanner

PRÁCTICO 2

EJERCICIO 1: La biblioteca de una universidad tiene 5 ejemplares de un libro. Dos ejemplares (1 y 2) son primeras impresiones y el resto son segundas impresiones. Un estudiante examina estos libros en orden aleatorio, deteniéndose sólo cuando selecciona una segunda impresión.

a) Hacer una lista de todos los resultados posibles:

$S = \{123, 124, 125, 1-3, 1-4, 1-5, 2-1-3, 2-1-4, 2-1-5, 2-3, 2-4, 2-5, 3, 4, 5\}$

$\#S = 15$

b) Resultados en el evento A: "Sólo un libro es examinado?

$A = \{3, 4, 5\}$

c) Resultados en el evento B: "el libro 5 es seleccionado?

$B = \{125, 215, 15, 25, 5\}$

d) Resultados en el evento C: "el libro 1 no se examina"?

$C = \{23, 24, 25, 3, 4, 5\}$

e) Expresar $A \cap B$, $A \cup B$ y $\overline{(B \cap C)}$

$A \cap B = \{3, 4, 5\} \cap \{125, 215, 15, 25, 5\} = \{5\}$

$A \cup B = \{125, 215, 15, 25, 3, 4, 5\}$

$\overline{(B \cap C)} = S - (B \cap C) = S - \{25, 5\}$

EJERCICIO 3: Una empresa ha licitado 3 proyectos. Sea $A_i = \{$proyecto i otorgado$\}$ para $i = 1, 2, 3$ y supongamos que $P(A_1) = 0{,}22$, $P(A_2) = 0{,}25$, $P(A_3) = 0{,}28$, $P(A_1 \cap A_2) = 0{,}11$; $P(A_1 \cap A_3) = 0{,}05$; $P(A_2 \cap A_3) = 0{,}07$ y $P(A_1 \cap A_2 \cap A_3) = 0{,}01$.

Calcule: P(...

a) $A_1 \cup A_2$; $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0{,}22 + 0{,}25 - 0{,}11 = 0{,}36$

b) $\overline{A_1} \cap \overline{A_2}$; $P(\bar{A}_1 \cap \bar{A}_2) = P(\overline{A_1 \cup A_2}) = 1 - P(A_1 \cup A_2) = 1 - 0{,}36 = 0{,}64$

c) $A_1 \cup A_2 \cup A_3$; $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2 \cap A_3)$

$- P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) = 0{,}22 + 0{,}25 + 0{,}28 - 0{,}01 - 0{,}11 - 0{,}05 - 0{,}07$

$=$

d) $P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = P(\overline{A_1 \cup A_2 \cup A_3}) = 1 - P(A_1 \cup A_2 \cup A_3)$

EJERCICIO 4: Cinco empresas $F_1, F_2, \ldots F_5$ hacen propuestas con respecto a tres contratos separados C_1, C_2 y C_3. Una empresa sólo puede obtener a lo sumo 1 contrato. Los contratos son diferentes, de tal forma que la asignación de C_1 a F_1 se debe diferenciar de C_2 a F_1.

a) ¿Cuántos puntos muestrales hay en total en este experimento?

$$\underbrace{5}_{C_1} \times \underbrace{4}_{C_2} \times \underbrace{3}_{C_3} = 60$$

b) Encuentre la probabilidad de que se conceda un contrato a la empresa F_3, bajo el supuesto de que todos los puntos son equiprobables.

$$P = \frac{1 \times 4 \times 3}{60} \times 3 \quad (\text{ya que } F_3 \text{ puede recibir 3 contratos distintos})$$

$$= \frac{3}{5} = 0{,}6.$$

Scanned by CamScanner

EJERCICIO 5: Al poco tiempo de ser puestos en servicios, algunos colectivos fabricados por cierta compañía presentan grietas en su parte inferior. Una ciudad tiene 25 de estos colectivos y han apareudo grietas en 8 de ellos. Se seleccionan aleatoriamente 5 colectivos para hacer una inspección.

a) ¿Cuál es la probabilidad de que 4 de los 5 seleccionados tengan grietas?

$p = \frac{8}{25}$

$$\boxed{p(r) = \binom{n}{n-r} p^{r} (1-p)^{n-r}} \Rightarrow p(4) = \binom{5}{5-4}\left(\frac{8}{25}\right)^{4}\left(1-\frac{8}{25}\right)^{1}$$

$$= \binom{5}{1}(0{,}01)(0{,}68) \doteq 0{,}0034$$

b) ¿Cuál es la probabilidad de que al menos 4 de los seleccionados tengan grietas?

$$p(x \geq 4) = p(4) + p(5) = p(4) + \binom{5}{0}\left(\frac{8}{25}\right)^{5} = p(4) + 0{,}00335 = 0{,}037;$$

EJERCICIO 6: Una profesora desea programar una reunión con cada uno de sus 8 ayudantes, 4 hombres y 4 mujeres; Se supone que todos los ordenamientos posibles de las reuniones son equiprobables.

a) ¿Cuál es la probabilidad de que por lo menos 1 asistente mujer esté entre los 3 primeros?

Espacio muestral (cantidad de combinaciones posibles):

$8.7.6.5.4.3.2.1 = 8! = \#S$

Calculamos lo opuesto, que los primeros 3 días sean sólo hombres $= P(H)$

$\underbrace{4.3.2}_{\text{sólo entre hombres}}.5.4.3.2.1 = 2880$ $\qquad P(M) = 1 - P(H) = 1 - \frac{2880}{8!} = 1 - 0{,}07$

$= 0{,}928$

b) ¿Cuál es la probabilidad de que después de las cinco primeras reuniones, la profesora se haya reunido con todas las asistentes?

Esto implica que las últimas 3 reuniones son sólo hombres $\binom{4}{3}.3!$ (importa el orden), y el resto de las reuniones en todas sus combinaciones posibles es $5!$ $\Rightarrow$

$$P = \frac{\binom{4}{3}.3!.5!}{8!} = 4.3.2.1.5.4.3.2$$

EJERCICIO 7: Un departamento académico con cinco miembros de la facultad: Anderson, Box, Cox, Cramer y Fisher, debe seleccionar a dos de sus miembros para prestar servicio en una comisión. El representante será seleccionado colocando 5 boletos en una caja y seleccionando 2.

a) ¿Cuál es la probabilidad de que Anderson y Box sean seleccionados?

Es la misma probabilidad que la de cualquier par de personas:

$\binom{5}{2} = 10$ posibles formas de seleccionar dos personas,

entonces $P(\text{Anderson} \,\&\, \text{Box}) = 1/10$

b) ¿Cuál es la probabilidad de que se seleccione por lo menos a uno de los dos miembros cuyos apellidos empiecen con C? = B

$$P(B) = \frac{\binom{2}{1}\binom{3}{1}}{10} + \frac{\binom{2}{2}}{10} = \frac{6}{10} + \frac{1}{10} = \frac{7}{10} = 0{,}7$$

Scanned by CamScanner

PRÁCTICO 2 - hoja 2

c) Si los 5 miembros seleccionados han dado clases durante 3, 6, 7, 10 y 14 años respectivamente, ¿Cuál es la probabilidad de que entre los elegidos sumen 15 años por lo menos de enseñanza? $= C$

$S' = \{ \cdot 3+14 \cdot 6+10 \cdot 6+14 \cdot 7+10 \cdot 7+14 \cdot 10+14 \}$

$\# S' = 6$; $P(C) = \frac{6}{10} = 0,6$;

EJERCICIO 8: Una gran tienda vende camisas de distintos talles y estilos. La probabilidad (proporción) de cada una se representa en la tabla (...).

a) ¿Cuál es la probabilidad de que la camisa sea mediana, de manga larga y estampada?

0,05

b) mediana y estampada? $0,05 + 0,07 = 0,12$

c) manga corta? $0,04 + 0,02 + 0,05 + 0,08 + 0,07 + 0,12 + 0,03 + 0,07 + 0,08 = 0,56$
manga larga? $0,03 + 0,02 + 0,03 + 0,1 + 0,05 + 0,07 + 0,04 + 0,02 + 0,08 = 0,44$
$(1 - 0,56)$

d) mediano? $(0,08 + 0,07 + 0,12) + (0,10 + 0,05 + 0,07) = 0,49$
estampado? $(0,02 + 0,07 + 0,07) + (0,02 + 0,05 + 0,02) = 0,25$

e) Dado que la camisa vendida era a cuadros y manga corta, ¿Cuál es la probabilidad de que su talle sea mediano?

$$\left\{ P(A_i | B) = \frac{P(A_i) \cdot P(B|A_i)}{\sum_{j=1}^{n} P(A_j) \cdot P(B|A_j)} \right\}$$

$P(\text{cuadros y corta}) = 0,04 + 0,08 + 0,03 = 0,15 = P(CC)$

$P(\text{mediano}) = 0,49 = P(M)$

$$\Rightarrow P(M/CC) = \frac{P(M \cap CC)}{P(CC)} = \frac{0,08}{0,15} = 0,54$$

f) Dado que la camisa era a cuadros y mediana ¿Cuál es la P de que sea de manga corta?

$P(C \cup M) = 0,08 + 0,10 = 0,18$
$P(MC) = 0,56$

$$P(MC/C \cup M) = \frac{P(MC \cap C \cup M)}{P(C \cup M)} = \frac{0,08}{0,18} = 0,44$$

¿y P de manga larga?

$$P(C \cup M) = 0,18 \Rightarrow P(ML/C \cup M) = \frac{P(ML \cap C \cup M)}{P(C \cup M)} = \frac{0,10}{0,18} = 0,56 = 1 - 0,44$$

Scanned by CamScanner

EJERCICIO 9: Una caja contiene 6 cubos rojos y 4 verdes, y una segunda caja contiene 7 cubos rojos y 3 verdes. Se elige al azar un cubo de la primera y se pone en la segunda caja. Luego se selecciona al azar un cubo de la segunda caja y se pone en la primera caja.

a) ¿Cuál es la probabilidad de que se seleccione un cubo rojo de la primera caja y un cubo rojo de la segunda caja?

Se selecciona rojo en la primera $= P(R1) = \binom{6}{1}/10 = 6/10 = 0{,}6$

Se pone en la segunda caja. Ahora allí hay 8 rojos y 11 en total:

Rojo en la segunda $P(R_2/R_1) = \binom{8}{1}/11 = 8/11 = 0{,}73$

que sucedan estas cosas al mismo tiempo: $P(R_1 \cap \underset{R1}{R_2}) = P(R_1) \cdot P(R_2) = \frac{6}{10} \cdot \frac{8}{11} = 0{,}44$

b) Al finalizar el proceso de selección ¿cuál es la probabilidad de que los números de cubos rojos y verdes de la primera caja sean idénticos a los que habían al comienzo?

$P((R_1 \cap R_2|R_1) \cup (V_1 \cap V_2|V_1) = 0{,}44 + P(V_1 \cap V_2|V_1)$

$P(V_1 \cap V_2|V_1) = P(V_1) \cdot P(V_2|V_1) = \binom{4}{1}/10 \cdot \binom{4}{1}/11 = 0{,}145$

$\Rightarrow P[(R_1 \cap (R_2|R_1)) \cup (V_1 \cap (V_2|V_1))] = 0{,}44 + 0{,}145 = 0{,}585$

EJERCICIO 10:

a) Dados los eventos A y B con $P(B) > 0$, demuestre que $P(A|B) + P(\bar{A}|B) = 1$

$P(A|B) = \frac{P(A \cap B)}{P(B)}$; $P(\bar{A}|B) = \frac{P(\bar{A} \cap B)}{P(B)}$

$\frac{P(A \cap B)}{P(B)} + \frac{P(\bar{A} \cap B)}{P(B)} = \frac{1}{P(B)}[P(A \cap B) + P(\bar{A} \cap B)] = \frac{1}{P(B)}[P(A \cap B) + (P(B) - P(A \cap B))]$

$= \frac{1}{P(B)}[P(A \cap B) + P(B) - P(A \cap B)] = P(B)/P(B) = 1$;

b) Si $P(B|A) > P(B)$ demuestre $P(\bar{B}|A) < P(\bar{B})$

$P(B|A) > P(B)$; $(1 - P(\bar{B}|A)) > 1 - P(\bar{B})$;

$-P(\bar{B}|A) > \cancel{1} - \cancel{1} - P(\bar{B})$; $-P(\bar{B}|A) > -P(\bar{B})$; $P(\bar{B}|A) < P(\bar{B})$ ✓

c) Dados los eventos A, B y C con $P(C) > 0$, demuestre que $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$

$P(A \cup B|C) = \frac{P[(A \cup B) \cap C]}{P(C)} = \frac{P[(A \cap C) \cup (B \cap C)]}{P(C)} = \frac{P(A \cap C) + P(B \cap C) - P(A \cap C \cap B \cap C)}{P(C)}$

$= \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C)}{P(C)} - \frac{P((A \cap B) \cap C)}{P(C)} = P(A|C) + P(B|C) - P((A \cap B)|C)$

Scanned by CamScanner

PRÁCTICO 2 - hoja 3

EJERCICIO 11: Uno de cada 25 adultos está afectado de cierta enfermedad para la que se ha desarrollado una prueba de diagnóstico. La prueba es tal que, cuando un individuo padece la enfermedad, el resultado de la prueba es positivo un 99% de las veces, mientras que si no está afectado el resultado positivo aparecerá sólo el 2% de las veces.

a) ¿Cuál es la probabilidad de que un resultado de la prueba sea positivo?

+: resultado positivo ; A: afectado ; NA: No afectado

$$P(+) = P(+\cap A) + P(+\cap NA) = 0{,}0588$$

$$P(+\cap A) = 0{,}99 \cdot \frac{1}{25} = 0{,}0396$$

$$P(+\cap NA) = 0{,}02 \cdot \frac{24}{25} = 0{,}0192$$

b) Dado que el resultado de la prueba es positivo, ¿Cuál es la probabilidad de que el individuo tenga la enfermedad?

$$P(A|+) = \frac{P(A\cap +)}{P(+)} = \frac{0{,}0396}{0{,}0588} = 0{,}067$$

c) Dado que el resultado de la prueba es negativo, ¿cuál es la probabilidad de que el individuo no tenga enfermedad?

$$P(NA|-) = \frac{P(NA\cap -)}{P(-)} = \frac{P(\bar{A}\cap \bar{+})}{P(-)} = \frac{P(\overline{A\cup +})}{1-P(+)} = \frac{1-P(A\cup +)}{1-P(+)}$$

$$= \frac{1-[P(A)+P(+)-P(A\cap +)]}{1-P(+)} = \frac{1-[1/25+0{,}0588-0{,}0396]}{1-0{,}0588}$$

$$= \frac{1-[0{,}0592]}{0{,}9412} = 0{,}9996 \;;$$

EJERCICIO 13: Una costura hecha en un avión necesita 25 remaches. La costura tendrá que volver a realizarse si al menos uno de los remaches está defectuoso. Suponga que los remaches están defectuosos independientemente y de manera equiprobable.

a) Si el 14% de todas las costuras necesitan rehacerse ¿Cuál es la probabilidad de que un remache esté defectuoso? RD: Remaches defectuosos

$$P(\text{rehacer costura}) = P(RD \geq 1) = 1 - P(RD < 1) = 1 - P(RD = 0)$$

Scanned by CamScanner

PRÁCTICO 3:

EJERCICIO 1: Sea X = número de neumáticos de un automóvil, seleccionado al azar, que tenga baja la presión
a) ¿Cuál de las siguientes tres funciones p(·) es una función de distribución de probabilidad para X? Justifique.
El sé La segunda es la única cuya sumatoria de todos los posibles valores de la función da 1.
b) Con la función de distribución de probabilidad obtenida en (a), calcular: $P(2 \leq X \leq 4)$, $P(X \leq 2)$ y $P(X \neq 0)$
$P(2 \leq X \leq 4) = p(2) + p(3) + p(4) = 0{,}1 + 0{,}1 + 0{,}3 = 0{,}5$.
$P(X \leq 2) = p(0) + p(1) + p(2) = 0{,}4 + 0{,}1 + 0{,}1 = 0{,}6$
$P(X \neq 0) = p(1) + p(2) + p(3) + p(4) = 0{,}1 + 0{,}1 + 0{,}1 + 0{,}3 = 0{,}6$
c) Obtener la función de distribución acumulada de X.

d) Si $p(x) = K(5 - x)$ para $x = 0, 1, 2, 3, 4$, ¿cuál debe ser el valor de la constante K para que p sea una función de distribución de probabilidad?
$K(5-0) + K(5-1) + K(5-2) + K(5-3) + K(5-4) = 5K + 4K + 3K + 2K + K = 15K = 1$;
$= \left(K \sum_{x=0} x\right)$ $K = 1/15 = 0{,}06\hat{}$

EJERCICIO 2: Un negocio de computadoras que atiende pedidos por correo tiene seis líneas telefónicas. Denotemos por X el número de líneas en uso en un momento específico. Supongamos que la función de distribución de probabilidad de X está dada en la siguiente tabla: (...)
a) A = A lo sumo tres líneas están en uso
• $P(X \leq 3) = p(0) + p(1) + p(2) + p(3) = 0{,}10 + 0{,}15 + 0{,}20 + 0{,}25 = 0{,}7$
b) B = Menos de tres líneas están en uso: $P(X < 3) = p(0) + p(1) + p(2) = 0{,}10 + 0{,}15 + 0{,}20 = 0{,}45$
c) C = Por lo menos tres líneas están en uso: $P(X \geq 3) = p(3) + p(4) + p(5) + p(6) = 0{,}25 + 0{,}20 + 0{,}06 + 0{,}04 = 0{,}55$
d) D = Entre 2 y 5 líneas están en uso: $P(2 \leq X \leq 5) = p(2) + p(3) + p(4) + p(5) = 0{,}20 + 0{,}25 + 0{,}20 + 0{,}06 = 0{,}71$
e) E = Entre 2 y 4 líneas no están en uso: $P(X \leq 1 \,||\, X \geq 5) = 1 - P(2 \leq X \leq 5) = 1 - 0{,}71 = 0{,}29$
f) F = Por lo menos 4 líneas no están en uso: $1 - P(X \geq 4) = P(X \leq 3) = p(0) + p(1) + p(2) + p(3) = 0{,}70$
g) Obtener la función de distribución acumulada de X

EJERCICIO 3:

Scanned by CamScanner

EJERCICIO 4: Una compañía de seguros ofrece a sus tenedores de pólizas varias opciones diferentes para el pago de primas. Para un tenedor seleccionado al azar, sea x = número de meses entre pagos sucesivos. La función de distribución acumulada de x es como sigue:

$$F(x) = \begin{cases} 0 & x<1 \\ 0{,}30 & 1 \le x < 3 \\ 0{,}40 & 3 \le x < 4 \\ 0{,}45 & 4 \le x < 6 \\ 0{,}60 & 6 \le x < 12 \\ 1 & 12 \le x \end{cases}$$

a) ¿Cuál es la función de distribución de probabilidad de x?

[0] [1 2] [3] [4 5] [6 7 8 9 10 11] [12 ...)

0 0,30 0,10 0,15 0,15,... 0,40

b) Sólo con el uso de la función de distribución acumulada, calcule $P(3 \le X \le 6)$ y $P(4 \le X)$

$P(3 \le x \le 6) = F(6) - F(2$

$P(4 \le x) = 1 - F(3) =$

EJERCICIO 5: Sea x = resultado de un dado no cargado se hace rodar una vez.

a) Calcular la esperanza de X y de $1/X$.

$p(x) = 1/6$;

$E(X) = \sum_{x \in Im} x \cdot p(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{1}{6}(1+2+3+4+5+6) = 21/6 = 7/2$

$E(1/X) = \sum_{x \in Im} 1/x \, p(x) = \frac{1}{6}(1/1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6) = 1/6\left(\frac{120+60+40+30+24+20}{120}\right)$

$= \frac{294}{720} = 0{,}408$

b) Si antes de arrojar el dado se ofrece al tirador retirarse con $1/3{,}5$ dólares o jugar obteniendo $h(x) = 1/x$ dólares ¿Qué le conviene?

$E(1/x) = 0{,}408$ le conviene seguir jugando pues el valor esperado

$1/3{,}5 = 0{,}286$ para $1/X$ es mayor.

EJERCICIO 6: Un distribuidor de aparatos electrodomésticos vende tres modelos de diferentes congeladores verticales con capacidad de 13,5, 15,9 y 19,1 pies cúbicos de espacio de almacenaje, respectivamente. Sea X la cantidad de espacio de almacenaje comprado por un cliente que va a comprar un congelador. Supongamos que X tiene la siguiente función de distribución de probabilidad: (...)

a) Calcular $E(X)$, $E(X^2)$ y $V(X)$

$E(X) = \sum_{x \in Im} x \, p(x) = 13{,}5 \cdot 0{,}2 + 15{,}9 \cdot 0{,}5 + 19{,}1 \cdot 0{,}3 = 2{,}7 + 7{,}95 + 5{,}73 = 16{,}38$

$E(X^2) = \sum_{x \in Im} x^2 \cdot p(x) = 182{,}25 \cdot 0{,}2 + 252{,}81 \cdot 0{,}5 + 364{,}81 \cdot 0{,}3 = 36{,}45 + 126{,}40 + 109{,}44$
$= 272{,}3$

$V(X) = E(X^2) - [E(X)]^2 = 272{,}3 - 268{,}3 = 4$

husai

Scanned by CamScanner

(PRÁCTICO 3 - hoja 2)

b) Si el precio de un ~~cong~~ congelador que tiene una capacidad de x pies cúbicos es de $25X - 8,5$ ¿cuál es la esperanza y varianza del precio pagado por un cliente que va a comprar un congelador?

$h(x) = 25X - 8,5$; $E(h(X)) = \sum_{x \in Im} h(x).p(x) = (25.13,5 - 8,5).0,2 + (25.15,9 - 8,5).0,5 + (25.19,1 - 8,5).0,3 = 400,7$

$E((h(x))^2) =$

$V(x) = E((h(x))^2) - [E(h(x))]$

c) Suponga que mientras la capacidad nominal de un congelador es X, la capacidad real es $h(x) = X - 0,01X^2$ ¿Cuál es la capacidad real esperada del congelador?

EJERCICIO 7: Suponga que sólo el 20% de los automovilistas se detienen por completo en un cruce donde hay un semáforo con luz roja intermitente en todas las direcciones, cuando no haya otros automóviles visibles.

a) ¿Cuál es la probabilidad de que de 20 automovilísticos, seleccionados al azar, que lleguen al cruce en estas condiciones:

$X \sim bin(x; 20; 0.20)$

I) a lo sumo 5 se detengan por completo?

$\binom{20}{x} 0,20^x . (1-0,20)^{20-x}$ y $B(X) = \sum_{x=0}^{5} b(x) = \binom{20}{0} 0,20^0 . (1-0,20)^{20} + \binom{20}{1} 0,20 . (1-0,20)^{19} + \binom{20}{2} 0,20^2 . (1-0,20)^{18} + \binom{20}{3} 0,20^3 . (1-0,20)^{17} + \binom{20}{4} 0,20^4 . (1-0,20)^{16} + \binom{20}{5} 0,20^5 . (1-0,20)^{15}$

$= \binom{20}{0}(0,80)^{20} + \binom{20}{1} 0,20 . (0,80)^{19} + \binom{20}{2} 0,20^2 . (0,80)^{18} + \binom{20}{3} 0,20^3 . (0,80)^{17} + \binom{20}{4} 0,20^4 . (0,80)^{16} + \binom{20}{5} 0,20^5 . (0,80)^{15}$

II) exactamente 5 se detengan por completo?

$b(5) = \binom{20}{5} 0,20^5 . 0,80^{15}$

húsares

Scanned by CamScanner

III) por lo menos 5 se detengan por completo?

$1 - B(5-1) = 1 - B(4)$

b) ¿Cuántos, de los siguientes 20 automovilistas, espera el lector que se detengan por completo?

$E(X) = n \cdot p = 20 \cdot 0{,}20 = 4$

EJERCICIO 8: Un tipo particular de raqueta de tenis se fabrica en tamaños mediano y extragrande. El 60% de todos los clientes, de cierta tienda, buscan el tamaño extragrande.

a) Entre 10 clientes seleccionados al azar que desean ese tipo de raqueta, ¿cuál es la probabilidad de que por lo menos 6 busquen el tamaño extragrande?

$X \sim bin(10; 0{,}60)$ = Cantidad de clientes que quieren la extragrande.

$P(X \geq 6) = 1 - B(6-1) = 1 - B(5)$; $b(x) = \binom{10}{x} 0{,}60^{x} \cdot (1-0{,}60)^{10-x}$;

$B(5) = \sum_{i=0}^{5} b(i) = \binom{10}{0} 0{,}60^{0} \cdot 0{,}40^{10} + \binom{10}{1} 0{,}60 \cdot 0{,}40^{9} + \binom{10}{2} 0{,}60^{2} \cdot 0{,}40^{8} + \binom{10}{3} 0{,}60^{3} \cdot 0{,}40^{7}$

$+ \binom{10}{4} 0{,}60^{4} \cdot 0{,}40^{6} + \binom{10}{5} 0{,}60^{5} \cdot 0{,}40^{5} =$

b) Entre 10 clientes seleccionados al azar que desean ese tipo de raqueta, ¿Cuál es la probabilidad de que el número de clientes que buscan la extragrande esté dentro de una desviación estándar del valor medio?

X = nro clientes que buscan extra.

$E(X) = 10 \cdot 0{,}6 = 6$; $V(X) = 10 \cdot 0{,}6 \cdot 0{,}4 = 2{,}4$; $\sigma_X = \sqrt{V(X)} = \sqrt{2{,}4} = 1{,}55 \approx 2$

$P(E(X) - \sigma_X \leq X \leq E(X) + \sigma_X) = P(6 - 2 \leq X \leq 6 + 2) = P(4 \leq X \leq 8)$

$= \sum_{i=4}^{8} b(i) = \binom{8}{k} (0{,}6)^{i} (0{,}4)^{8-i}$

Scanned by CamScanner

PRÁCTICO 3:
hoja3

c) La tienda tiene actualmente 6 raquetas de cada modelo, ¿cuál es la probabilidad de que los siguientes diez clientes que busquen esta raqueta puedan comprar el modelo que buscan, de entre la existencia actual?
12 raquetas en total

$P(4 \leq x \leq 6)$

no entiendo la solución

EJERCICIO 9: De todas las reparaciones hechas en aparatos de TV en cierta tienda, el 80% se hace en aparatos que ya no tienen garantía.
a) Entre 20 aparatos llevados a reparación, ¿cuál es el número esperado de aparatos con garantía?
X = "aparatos con garantía"
$E(X) = 20 \cdot 0{,}20 = 4$
b) ¿Y sin garantía? $E(X) = 20 \cdot 0{,}8 = 16$ (X = "aparatos sin garantía")
c) ¿Cual es la prob de que por lo menos el 75% de los 20 ya no tenga garantía?
$20 \cdot 0{,}75 = 15$ X = "aparatos sin garantía"

$P(X \geq 15) = \sum_{i=15}^{20} b(i) = \sum \binom{20}{i} 0{,}8^i (0{,}2)^{20-i}$ $(P(X \geq 15) = 0{,}78)$

d) Suponga que hay 12 aparatos en la tienda, de los cuales 4 tienen garantía. Si 5 de los 12 son llevados a reparación en orden aleatorio y se reparan en el mismo orden, ¿cuál es la función de distribución de probabilidad de X = número de aparatos con garantía de entre los 5 reparados?
¿Cuanto vale E(X) y V(X)?
~~P(x) = 1/3~~. HIPERGEOMÉTRICA $X \sim h(x; n, M, N) = h(x; 5, 4, 12)$

$$P(x) = P(X = x) = h(x; 5, 4, 12) = \frac{\binom{4}{x}\binom{8}{5-x}}{\binom{12}{5}} = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = h(x; n, M, N)$$

$$\boxed{E(X) = n \cdot \frac{M}{N}}$$

$$E(X) = 5 \cdot \frac{4}{12} = \frac{20}{12} = 1{,}7 \approx 2 \; ; \; V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

$$V(X) = \left(\frac{12-5}{11}\right) \cdot 5 \cdot \frac{4}{12} \cdot \left(1 - \frac{4}{12}\right) = \frac{7}{11} \cdot \frac{20}{12} \cdot \frac{8}{12} = \frac{7}{11} \cdot \frac{\cancel{20}^{5}}{\cancel{6}_{3}} \cdot \frac{\cancel{4}^{2}}{\cancel{6}_{3}}$$
$$= 70/99 = 0{,}70$$

EJERCICIO 10: Cada uno de 10 refrigeradores de cierto tipo ha sido devuelto debido a un ruido. Supongamos que 7 de esos 10 tienen compresores defectuosos. Si se examinan 6 refigeradores al azar, sea X = el nro de refrigeradores con compresor defectuoso de los 6.
a) Calcular: i) $P(X=1)$ $X \sim h(x; 6, 7, 10)$ $P(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$; $P(X=1) = p(1) = \frac{\binom{7}{1}\binom{3}{6-1}}{\binom{10}{7}}$

Es imposible calcular esta posibilidad ya que no hay forma de elegir 5 no defectuosos ya que sólo hay 3.

Scanned by CamScanner

ii) $P(X \geq 4) = p(4) + p(5) + p(6) = \dfrac{\binom{7}{4}\binom{3}{6-4}}{\binom{10}{7}} + \dfrac{\binom{7}{5}\binom{3}{6-5}}{\binom{10}{7}} + \dfrac{\binom{7}{6}\binom{3}{0}}{\binom{10}{7}}$

$= \dfrac{\binom{7}{4}\binom{3}{2} + \binom{7}{5}\binom{3}{1} + \binom{7}{6}\binom{3}{0}}{\binom{10}{7}} =$ (calcular)

iii) $P(3 \leq X \leq 5) = p(3) + p(4) + p(5) = \dfrac{\binom{7}{3}\binom{3}{3} + \binom{7}{4}\binom{3}{2} + \binom{7}{5}\binom{3}{1}}{\binom{7}{10}}$

b) ¿Cuánto vale E(X) y V(X)? $E(X) = n \cdot \frac{M}{n}$ $V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$

$E(X) = 6 \cdot \frac{7}{10} = 4{,}2 \simeq 4$ $V(X) = \frac{4}{9} \cdot 6 \cdot \frac{7}{10} \cdot \frac{3}{10} = 2 \cdot \frac{7}{25} = 0{,}56$

EJERCICIO 11: Suponga que $p = P(\text{Nace niño}) = 0{,}5$. Una pareja desea exactamente dos niñas. Tendrán hijos hasta satisfacer esa condición.

X = "Cantidad de niños hasta llegar a las 2 niñas"

$X \sim nb(x; r, p) = nb(x; 2, 0{,}5)$

a) $nb(x) = \binom{x+r-1}{r-1} p^r (1-p)^x = \binom{x+1}{1} 0{,}5^2 (0{,}5)^x = (x+1) 0{,}5^{x+2}$

b) ¿Cuál es la prob. de que la familia tenga 4 hijes?

$P(X=2) = p(2) = nb(2) = (2+1) 0{,}5^4 = 3 \cdot 0{,}5^4 = 0{,}1875$

c) ¿Cuál es la probabilidad de que tengan a lo sumo 4 hijes?

$P(X \leq 2) = p(0) + p(1) + p(2) = 0{,}5^2 + 2 \cdot 0{,}5^3 + (0{,}1875) = 0{,}25 + 0{,}25 + 0{,}1875 = 0{,}6875 = 0{,}69$

d) ¿Cuántos hijos varones se esperaría que tenga la familia? ¿Cuántos hijos se esperaría que tenga la familia?

$E(X) = \frac{r(1-p)}{p}$ $E(X) = \frac{2(1-0{,}5)}{0{,}5} = \frac{2 \cdot 0{,}5}{0{,}5} = 2$

Y = hijes totales

$E(Y) = E(X) + 2 = 2 + 2 = 4$

Scanned by CamScanner

PRÁCTICO 3; hoja 4.

EJERCICIO 12: Suponga que X= número de tornados observados en una región en particular, durante 1 año, tiene una distribución de Poisson con $\lambda = 8$.

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

a) Calcular $P(X \leq 5)$

$P(X \leq 5) = p(1) + p(2) + p(3) + p(4) + p(5)$

$= \frac{e^{-8} \cdot 8^1}{1!} + \frac{e^{-8} 8^2}{2!} + \frac{e^{-8} \cdot 8^3}{3!} + \frac{e^{-8} \cdot 8^4}{4!} + \frac{e^{-8} 8^5}{5!}$

$= e^{-8}\left(8 + \frac{8^2}{2!} + \frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!}\right) = e^{-8}(8 + 32 + 85,34 + 170,67 + 273,07) = 0,191$

$P(6 \leq x \leq 9) = p(6) + p(7) + p(8) + p(9)$

$= e^{-8}\left(\frac{8^6}{6!} + \frac{8^7}{7!} + \frac{8^8}{8!} + \frac{8^9}{9!}\right) = 0,525$

$P(X \geq 10) = 1 - P(X < 10) = 1 - (p(X \leq 5) + p(6 \leq X \leq 9)) = 1 - (0,191 + 0,525) =$

$1 - 0,191 - 0,525 = 0,284$;

$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X=0) = 1 - p(0) = 1 - e^{-8} \cdot \frac{8^0}{0!} = 1 - e^{-8} = 0,999$

b) ¿Cuántos tornados se puede esperar que se observen durante un período de un año?

$E(X) = V(X) = \lambda = 8$

¿Cuál es la desviación estándar de X?

$\sigma_X = \sqrt{V(X)} = \sqrt{8} = 2,83$;

EJERCICIO 13: Un estacionamiento tiene dos entradas. Los coches llegan por hora a la entrada I de acuerdo con una distribución de Poisson de parámetro $\lambda = 3$ y a la entrada II de acuerdo con una distribución de Poisson de $\lambda = 4$. ¿Cuál es la probabilidad de que 3 coches lleguen al estacionamiento durante una hora dada? (Se supone que los números de coches que llegan a las dos entradas son independientes).

X = nro de autos que llegan al estacionamiento; Y = nro de autos por entrada I;
Z = nro de autos por entrada II;

$P(X=3) = P(Y=0) \cdot P(Z=3) + P(Y=1) \cdot P(Z=2) + P(Y=2) \cdot P(Z=1) + P(Y=3) \cdot P(Z=0)$

$= (e^{-3} 3^0/0!) \cdot (e^{-4} 4^3/3!) + (e^{-3} 3/1!) \cdot (e^{-4} 4^2/2!) + (e^{-3} \cdot 3^2/2!)(e^{-4} 4/1!) + (e^{-3} \cdot 3^3/3!)(e^{-4} \cdot 4^0/0!)$

$= 0,0097 + 0,022 + 0,0164 + 0,004$

$= \underline{0,052}$;

EJERCICIO 14: Se supone que el nro de defectos Y (por cm) de la producción diaria de cierto tipo de soga tiene una distribución de Poisson con una media de 2. Cuando se vende la soga, la ganancia por cm está dada por $X = 50 - 2Y - Y^2$. Dar la ganancia esperada por cm.

se sabe que

$E(Y) = 2$; $E(X) = E(50 - 2Y - Y^2) = 50 - 2E(Y) - E(Y^2) = 50 - 2 \cdot 2 - 6 = \underline{40}$;

y que $V(X) = E(X)$ en esta distribución, y $V(X) = E(X^2) - [E(X)]^2$

$V(Y) = 2 = E(Y^2) - 4$; $E(Y^2) = 2 + 4 = 6$;

Scanned by CamScanner

EJERCICIO 15: DEMOSTRACIONES

Scanned by CamScanner

PRÁCTICO 4:

EJERCICIO 1: Un libro está disponible 2 horas para la sala de lectura de la biblioteca de una universidad. Denotemos con x el tiempo de préstamo solicitado por un estudiante seleccionado al azar. Supongamos que X tiene la sig. función de densidad.

$$f(x) = \begin{cases} \frac{x}{2} & 0 \le x \le 2 \\ 0 & c.c \end{cases}$$

a) Calcular las siguientes probabilidades:

- $P(X \le 1) = F(1) = \int_0^1 f(y)\,dy$

$$= \int_0^1 \frac{y}{2} \cdot dy = \frac{1}{2}\int_0^1 y \cdot dy = \frac{1}{2}\left[\left(\frac{y^2}{2}\right)\right]\Big|_0^1$$

$$= \frac{1}{2}\left(\frac{1}{2} - 0\right) = \frac{1}{4};$$

- $P(1{,}5 < X) = 1 - F(1{,}5) = 1 - \int_0^{1,5} f(y)\,dy$

$$= 1 - \frac{1}{2}\left(\frac{y^2}{2}\right)\Big|_0^{1,5} = 1 - \frac{1}{2}\left(\frac{2{,}25}{2}\right) = 1 - 0{,}5625 = 0{,}4375$$

b) • $P(0{,}5 \le X \le 1{,}5)$

$$= F(1{,}5) - F(0{,}5)$$

$$= \int_0^{1,5} f(y)\,dy - \int_0^{0,5} f(y)\,dy$$

$$= \frac{1}{2}\left(\frac{y^2}{2}\right)\Big|_0^{1,5} - \frac{1}{2}\left(\frac{y^2}{2}\right)\Big|_0^{0,5}$$

$$= \frac{1}{2}\left(\frac{2{,}25}{2}\right) - \frac{1}{2}\left(\frac{0{,}25}{2}\right) =$$

$$= 0{,}5625 - 0{,}0625 = 0{,}5$$

b) Obtener la función de distribución acumulada de X:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy = \frac{1}{2} \cdot \frac{x^2}{2} = \frac{x^2}{4} \Rightarrow \boxed{F(X) = \begin{cases} x^2/4 & 0 \le x \le 2 \\ 0 & x < 0 \\ 1 & x > 2 \end{cases}}$$

c) Calcular $E(X)$, $V(X)$ y σ_X

$$E(X) = \int_0^2 x \cdot f(x) \cdot dx = \int_0^2 x \cdot \frac{x}{2} \cdot dx = \int_0^2 x^2/2 \cdot dx = \frac{1}{2}\int_0^2 x^2\,dx = \frac{1}{2}\left(\frac{x^3}{3}\right)\Big|_0^2 = \frac{x^3}{6}\Big|_0^2$$

$$= \left(2^3/6 - 0\right) = 8/6 = 4/3 = \underline{\underline{1{,}3}};$$

$$V(X) = E(X^2) - [E(X)]^2 = \int_0^2 x^2 f(x^2) - 1{,}78 = \int_0^2 x^2 \cdot x/2\,dx - 1{,}78 = \int_0^2 x^3/2\,dx - 1{,}78 =$$

$$= \frac{1}{2}\frac{x^4}{4}\Big|_0^2 - 1{,}78 = \frac{x^4}{8}\Big|_0^2 - 1{,}78 = 2 - 1{,}78 = \underline{\underline{0{,}22}};$$

$$\sigma_X = \sqrt{V(X)} = 0{,}47$$

d) Si a la persona que solicita el libro se le cobra $h(x) = x^2$ cuando la duración del prestamo es X, calcule el cobro esperado $E[h(x)]$;

$$E[h(x)] = E(X^2) = 2;$$

Scanned by CamScanner

EJERCICIO 2: Un profesor siempre termina su clase después de que suene el timbre y a menos de 2 min desde que suena. Sea X= "tiempo en minutos que transcurre ~~desde~~ entre que suena el timbre y el término de la clase. La función de densidad de X es:

$$f(x)=\begin{cases} k.x^2 & 0\le x\le 2 \\ 0 & c.c \end{cases}$$

a) Encuentre el valor de k;

• $f(x)\ge 0$; • $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^2 k.x^2.dx = k\int_0^2 x^2dx = k.x^3/3\Big|_0^2 = k.\frac{8}{3} = 1;\ \boxed{k=3/8}$$

b) ¿Cuál es la probabilidad de que la clase termine antes de 1 minuto después de que suene el timbre?

$$P(X\le 1)=\int_0^1 f(x)dx = \int_0^1 3/8.x^2.dx = \frac{3}{8}\int_0^1 x^2dx = \frac{3}{8}\left(\frac{x^3}{3}\right)\Big|_0^1 = \frac{1}{8} = \underline{0,125};$$

c) ¿Cuál es la probabilidad de que la clase continue entre 60 y 90 segundos?

$$P(1\le x\le 1,5)=\int_1^{1,5} f(x)dx = \int_1^{1,5} 3/8.x^2dx = \frac{3}{8}\int_1^{1,5} x^2dx = \frac{3}{8}\left(\frac{x^3}{3}\right)\Big|_1^{1,5} = \frac{3}{8}\left(\frac{1,5^3}{3}-\frac{1}{3}\right) = \frac{3}{8}\left(\frac{2,375}{3}\right)$$

$$= \underline{0,297};$$

d) Obtener la función de distribución acumulada de X;

$$F(x)=\begin{cases} 0 & x<0 \\ x^3/8 & 0\le x\le 2 \\ 1 & x>2 \end{cases}$$

e) ¿Cuál es el percentil 75 de la distribución?

$$F(\eta(p))=\int_0^{\eta(p)} (3/8).x^2.dx = 0,75;\quad \frac{3}{8}.\frac{(\eta(p))^3}{3} = 0,75;\quad \frac{(\eta(p))^3}{8} = \underline{0,75};$$

$$\eta(p) = \sqrt[3]{0,75.8} \simeq \underline{1,82};$$

f) ~~¿Cal~~ Calcular E(X) y σ_X

$$E(X)=\int_{-\infty}^{\infty} x.f(x).dx = \int_0^2 3/8.x^2.x.dx = \frac{3}{8}\int_0^2 x^3dx = \frac{3}{8}\left(\frac{x^4}{4}\right)\Big|_0^2 = \frac{3}{8}.4 = 3/2 = \underline{1,5};$$

$$V(X) = E(X^2)-[E(X)]^2;\quad E(X^2)=\int_{-\infty}^{\infty} x^2.f(x).dx = \int_0^2 x^2\,3/8.x^2dx = \frac{3}{8}\int_0^2 x^4dx = \frac{3}{8}\left(\frac{x^5}{5}\right)\Big|_0^2$$

$$= \frac{3}{8}\cdot\frac{32}{5} = \frac{12}{5} = 2,4;$$

$$V(X) = 2,4-2,25 = 0,15;\quad \sigma_X = \sqrt{V(X)} = \sqrt{0,15} \simeq 0,39;$$

e) ¿Cuál es la prob. de que X esté a menos de 1 desviación estándar de su valor medio?

$$P(1,5-0,39\le X\le 1,5+0,39) = P(1,11\le X\le 1,89)$$

$$= \int_{1,11}^{1,89} f(x)dx = \int_{1,11}^{1,89} 3/8\,x^2.dx = \frac{3}{8}\left(\frac{x^3}{3}\right)\Big|_{1,11}^{1,89} = \frac{1}{8}(1,89^3-1,11^3) \simeq \underline{0,67};$$

húsares

Scanned by CamScanner

PRÁCTICO 4 - hoja 2:

EJERCICIO 3: El tiempo X (en minutos) en que un asistente prepara el equipo para un experimento tiene una distribución uniforme en el intervalo [25, 35]:

a) Dar la función de densidad y la función de distribución acumulada de X.

$35-25=10$; Como es uniforme entonces: $f(x)=\frac{1}{B-A}$ y $F(x)=\frac{x-A}{B-A}$

$$f(x)=\begin{cases} 1/10 & 25\leq x\leq 35 \\ 0 & c.c \end{cases} \quad y \qquad A=25 \quad B=35$$

$$F(x)=\begin{cases} 0 & x<25 \\ \frac{x-25}{10} & 25\leq x\leq 35 \\ 1 & x>35 \end{cases};$$

b) ¿Cuál es la probabilidad de que el tiempo de preparación exceda 33 minutos?

$P(X>33)=1-P(X\leq 33)=1-F(33)=1-\frac{33-25}{10}=1-0,8=\underline{\underline{0,2}}$;

ó $P(X>33)=\int_{33}^{35} f(x)dx=\int_{33}^{35} 1/10\, dx=\frac{1}{10}\int_{33}^{35} 1\cdot dx=\frac{1}{10}x\Big|_{33}^{35}=\frac{35-33}{10}=\frac{2}{10}=\frac{1}{5}=\underline{0,2}$;

c) Para cualquier a tal que $25<a<a+2<35$ ¿Cual es la probabilidad de que el tiempo de preparación esté entre a y a+2?

$P(a\leq x\leq a+2)=\int_a^{a+2} f(x)dx=\frac{1}{10}\int_a^{a+2} 1\, dx=\frac{1}{10}(x)\Big|_a^{a+2}=\frac{1}{10}(a+2-a)=\frac{2}{10}=\frac{1}{10}=\underline{0,2}$;

d) Calcular $E(X)$ y σ_X;

$E(X)=\int_{-\infty}^{\infty} x f(x)dx=\int_{25}^{35} x\cdot 1/10\, dx=\frac{1}{10}\int_{25}^{35} x\,dx=\frac{1}{10}\left(\frac{x^2}{2}\right)\Big|_{25}^{35}=\frac{1}{10}(300)=\underline{\underline{30}}$ min.

$V(X)=E(x^2)-[E(x)]^2$; $E(x^2)=\int_{25}^{35} x^2 f(x)=\frac{1}{10}\left(\frac{x^3}{3}\right)\Big|_{25}^{35}\simeq 908,34$;

$V(X)=908,34-900\simeq 8,34$;

$\sigma_X=\sqrt{V(X)}\simeq \underline{\underline{2,89}}$;

e) ¿Cuál es la prob. de que x se encuentre a menos de 1 DE del tiempo medio? ¿Y a menos de 2 DE?

$P(30-2,89\leq x\leq 30+2,89)=P(27,11\leq x\leq 32,89)=$

$=\int_{27,11}^{32,89} f(x)dx=\frac{1}{10}(x)\Big|_{27,11}^{32,89}=\underline{\underline{0,578}}$;

$P(30-5,78\leq x\leq 30+5,78)=P(24,22\leq x\leq 35,78)=\underline{\underline{1}}$;

Scanned by CamScanner

EJERCICIO 4: Sea X una variable aleatoria con distribución normal con media 80 y varianza 100 ($\mathcal{N}(80,100)$).

$\mu = 80 \quad \sigma^2 = 100$

Calcular:

a) $P(X \leq 100) = P\left(Z \leq \frac{100-80}{10}\right) = P(Z \leq 2) = 0{,}9772 = \Phi(2)$

b) $P(65 \leq X \leq 100) = P\left(\frac{65-80}{10} \leq Z \leq \frac{100-80}{10}\right) = P(-1{,}5 \leq X \leq 2) = \Phi(2) - (\Phi(-1{,}5))$
$= \Phi(2) - (1 - \Phi(1{,}5)) = 0{,}9772 - 1 + 0{,}9332 = 0{,}9104$

c) $P(70 \leq X) = P\left(\frac{70-80}{10} \leq Z\right) = P(-1 \leq Z) = 1 - \Phi(-1) = 1 - (1 - \Phi(1))$
$= 1 - 1 + \Phi(1) = \Phi(1) = 0{,}8413$;

d) $P(85 \leq X \leq 95) = P\left(\frac{85-80}{10} \leq Z \leq \frac{95-80}{10}\right) = P(0{,}5 \leq Z \leq 1{,}5) = \Phi(1{,}5) - \Phi(0{,}5) =$
$= 0{,}9332 - 0{,}6915 = 0{,}2417$;

e) $P(|X-80| \leq 10) = P(-10 \leq X - 80 \leq 10) = P(70 \leq X \leq 90) = P\left(\frac{70-80}{10} \leq Z \leq \frac{90-80}{10}\right)$
$= P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = \Phi(1) - 1 + \Phi(1)$
$= 2.\Phi(1) - 1 = 2(0{,}8413) - 1 = 0{,}6826$

EJERCICIO 5: El diámetro de los árboles de determinado tipo, a cierta altura, se distribuye normalmente con $\mu = 8{,}8$ pulg. y $\sigma = 2{,}8$ pulg.

a) ¿Cuál es la probabilidad de que el diámetro de un árbol seleccionado al azar, sea a lo sumo 10 pulg.? ¿y que sea mayor que 10 pulg.?

X = "diámetro del árbol en pulgadas"; $P(X \leq 10) = P\left(Z \leq \frac{10-8{,}8}{2{,}8}\right) = P(Z \leq 0{,}428) =$

$= \Phi(0{,}43) = 0{,}6664$; $P(X > 10) = 1 - P(X \leq 10) = 0{,}3336$

b) ¿Cuál es la prob. de que X esté entre 5 y 10 pulg.?

$P(5 \leq X \leq 10) = P\left(\frac{5-8{,}8}{2{,}8} \leq Z \leq \frac{10-8{,}8}{2{,}8}\right) = P(-1{,}36 \leq Z \leq 0{,}43) = \Phi(0{,}43) - \Phi(-1{,}36)$
$= \Phi(0{,}43) - (1 - \Phi(1{,}36)) = \Phi(0{,}43) - 1 + \Phi(1{,}36) = 0{,}6664 - 1 + 0{,}9131 = 0{,}5795$;

c) ¿Qué valor de c es tal que el intervalo $(8{,}8 - c;\ 8{,}8 + c)$ incluye el 98% de todos los valores de diámetro?

$P(8{,}8 - c \leq X \leq 8{,}8 + c) = 0{,}98 = P\left(\frac{(8{,}8-c)-8{,}8}{2{,}8} \leq Z \leq \frac{(8{,}8+c)-8{,}8}{2{,}8}\right) = P\left(\frac{-c}{2{,}8} \leq Z \leq \frac{c}{2{,}8}\right)$

$= \Phi(c/2{,}8) - \Phi(-c/2{,}8) = \Phi(c/2{,}8) - 1 + \Phi(c/2{,}8) = 2.\Phi(c/2{,}8) - 1 = 0{,}98$;

$\Phi(c/2{,}8) = 0{,}99 = \Phi(2{,}33)$; $c/2{,}8 = 2{,}33$; $c = 2{,}33 . 2{,}8 = 6{,}524$

d) ¿Cuál es la probabilidad de que al menos 1 de 5 árboles elegidos al azar tenga diámetro entre 5 y 10 pulg.?

Y = "número de árboles con diámetro entre 5 y 10 pulg."

E = "diámetro entre 5 y 10" $= (5 \leq X \leq 10)$; $Y \sim b(n,p)$; $p = P(E) = P(5 \leq X \leq 10) = 0{,}58$

$p(x) = \binom{5}{x} 0{,}58^x (1 - 0{,}58)^{5-x} = \binom{5}{x} 0{,}58^x . 0{,}42^{5-x}$;

$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = 1 - p(0) = 1 - \binom{5}{0} 0{,}58^0 . 0{,}42^5 = 1 - 0{,}42^5$
$= 1 - 0{,}013 = 0{,}987 \simeq 0{,}99$;

húsares

Scanned by CamScanner

PRÁCTICO 4 - hoja 3:

EJERCICIO 6: La distribución de resistencia para resistores de cierto tipo es normal, el 10% de todos los resistores tiene una resistencia que excede los 10,256 ohms y el 5% una resistencia menor a 9,671 ohms. ¿Cuales son los valores de la media y la desviación estándar?

$P(X \geq 10{,}256) = 0{,}10$ & $P(X \leq 9{,}671) = 0{,}05$;

$P(Z \geq \frac{10{,}256-\mu}{\sigma_x}) = 1 - \Phi(\frac{10{,}256-\mu}{\sigma_x}) = 0{,}10$;

$\Phi(\frac{10{,}256-\mu}{\sigma_x}) = 0{,}9 \Rightarrow \frac{10{,}256-\mu}{\sigma_x} = \frac{1{,}28+1{,}29}{2} = 1{,}285$

$P(Z \leq \frac{9{,}671-\mu}{\sigma}) = 0{,}05 = \Phi(\frac{9{,}671-\mu}{\sigma_x}) = 1 - 0{,}95 = 1 - \Phi(1{,}645) = \Phi(-1{,}645)$

$\Rightarrow \frac{9{,}671-\mu}{\sigma_x} = -1{,}645$

$\frac{\mu}{\sigma_x} = -1{,}645 - \frac{9{,}671}{\sigma_x}$; $\frac{\mu}{\sigma_x} = 1{,}645 + \frac{9{,}671}{\sigma_x}$; $\mu = \sigma_x . 1{,}645 + 9{,}671$

(sustitución en la primera ecuación)

$\frac{10{,}256 - (\sigma_x . 1{,}645 + 9{,}671)}{\sigma_x} = 1{,}285$; $\frac{10{,}256 - \sigma_x . 1{,}645 - 9{,}671}{\sigma_x} = 1{,}285$

$\frac{1}{\sigma_x} . 10{,}256 - 1{,}645 - \frac{1}{\sigma_x} 9{,}671 = 1{,}285$; $\frac{1}{\sigma_x}(10{,}256 - 9{,}671) = 1{,}285 + 1{,}645 = 2{,}93$

$\frac{1}{\sigma_x}(0{,}585) = 2{,}93$; $\frac{1}{\sigma_x} = 2{,}93 / 0{,}585$; $\sigma_x = 0{,}585 / 2{,}93$; $\boxed{\sigma_x = 0{,}2}$;

$\mu = \sigma_x . 1{,}645 + 9{,}671 = (0{,}2) . 1{,}645 + 9{,}671$; $\boxed{\mu = 10}$

EJERCICIO 7:

a) Demuestre que si X tiene una distribución normal con parámetros μ y σ^2, entonces $Y = aX + b$ también tiene una distribución normal? ¿Cuáles son los parámetros de la distribución de Y ($E(Y)$ y $V(Y)$)?

buscar demostración en el libro

$E(Y) = E(aX + b) = aE(X) + b$; $E(Y^2) = E((aX+b)^2) = E(aX^2 + 2aXb + b^2)$

$= E(aE(X^2) + 2.aXb + b^2) =$

$E(aE(X^2)) + E(2aXb) + E(b^2) = aE(E(X^2)) + 2abE(X) + b^2$

$= aE(X^2) + 2abE(X) + b^2$;

$V(Y) = E(Y^2) - E(Y)^2 = aE(X^2) + 2abE(X) + b^2 - [aE(X) + b]^2$

$= aE(X^2) + 2abE(X) + b^2 - [(a.E(X))^2 + 2.aE(X).b + b^2]$

$= aE(X^2) + \cancel{2abE(X)} + \cancel{b^2} - a^2(E(X))^2 - \cancel{2abE(X)} - \cancel{b^2} = \underline{aE(X^2) - (aE(X))^2}$

Scanned by CamScanner

EJERCICIO 8: Sea X = "Distancia en metros que recorre un animal desde su lugar de nacimiento hasta el primer territorio vacante que encuentra". Se afirma que para los canguros, X tiene una distribución exponencial con parámetro $\lambda = 0{,}01386$.

a) ¿Cuál es la probabilidad de que la distancia sea a lo sumo 100m? ¿A lo sumo 200m? ¿Este entre 100 y 200?

$$\boxed{1 - e^{-\lambda x}}$$

$$P(X \leq 100) = F(100) = 1 - e^{-(0{,}01386).100} = \underline{0{,}75}$$

$$P(X \leq 200) = F(200) = 1 - e^{-0{,}01386.200} = \underline{0{,}94}$$

$$P(100 \leq X \leq 200) = F(200) - F(100) = 0{,}94 - 0{,}75 = \underline{0{,}19}$$

b) ¿Cuál es la probabilidad de que la distancia sea mayor que la distancia promedio en más de 2 desviaciones estándar?

$$E(X) = \mu = \frac{1}{\lambda} = \frac{1}{0{,}01386} = 72{,}15$$

$$\sigma^2 = \frac{1}{(0{,}01386)^2} = 5205.63 = \frac{1}{\lambda^2}$$

$$\sigma = \sqrt{\sigma^2} = 72{,}15 = \mu$$

$$P(\mu + 2\sigma \leq X) = P(\mu + 2\mu \leq X) = P(3\mu \leq X) = P(216{,}45 \leq X)$$

$$= 1 - P(X \leq 216{,}45) = 1 - F(216{,}45) = 1 - (1 - e^{-0{,}01386.216{,}45})$$

$$= e^{-0{,}01386.216{,}45} = \underline{0{,}05};$$

c) ¿Cuál es el valor de la mediana de la distancia?

$\tilde{X}$ = 50° percentil $\Rightarrow F(\tilde{X}) = 0{,}5$; $1 - e^{-0{,}01386.\tilde{X}} = 0{,}5$

$$e^{-0{,}01386.\tilde{X}} = 0{,}5;$$

$$-0{,}01386.\tilde{x} = \ln(0{,}5); \quad \tilde{x} = \ln(0{,}5)/-0{,}01386; \quad \boxed{\tilde{x} = 50{,}01}$$

EJERCICIO 9: Sea x = "tiempo (en 10^{-1} semanas) desde el envío de un producto defectuoso hasta que el cliente devuelve el producto". Suponga que el tiempo mínimo de devolución es 3,5 y que el exceso sobre el mínimo, $x - 3{,}5$, tiene una distribución WEIBULL con parámetros $\alpha = 2$ y $\beta = 1{,}5$

a) ¿Cuál es la función de distribución acumulada de X? $y = x - 3{,}5$

$$F(y + 3{,}5; 2, 1{,}5) = 1 - e^{-(x/1{,}5)^2} = F(x; 2; 1{,}5)$$

b) $P(X > 5) = 1 - P(X \leq 5) = 1 - F(5) = 1 - (1 - e^{-(5/1{,}5)^2}) = e^{-11{,}11} = 1{,}49 \cdot 10^{-5}$
$= 0{,}000015$

c) $P(5 \leq X \leq 8) = F(8) - F(5) = (1 - e^{-(8/1{,}5)^2}) - 0{,}000015 = 1 - 0{,}000015 = 0{,}99;$

$e^x = a$

$\ln(a) = x$

Scanned by CamScanner

PRÁCTICO 4 - hoja 4

EJERCICIO 10: En una fábrica se fabrican tornillos cuyo diámetro es una variable aleatoria normal. Si se cortan con la máquina A, el diámetro $\sim N(1;0,4)$; si se cortan con la B (cuyo costo de mantenimiento es mucho menor) el diámetro $\sim N(1,1;0,4)$. La máquina A se usa el 40% de las veces y la B el 60%. Para que el tornillo se considere aceptable, su diámetro debe estar entre 0,9 y 1,1 cm.

a) ¿Cuál es la probabilidad de que un tornillo elegido al azar cumpla el requerimiento deseado?

A = "que sea de la A" D = "diámetro en cm"

B = "que sea de la B" E = "que sea adecuado"

$$P(E) = P(A).P_A(0,9 \leq D \leq 1,1) + P(B).P_B(0,9 \leq D \leq 1,1)$$

$$= 0,4.\left(\frac{0,9-1}{0,4} \leq D' \leq \frac{1,1-1}{0,4}\right) + 0,6.\left(\frac{0,9-1,1}{0,4} \leq D' \leq \frac{1,1-1,1}{0,4}\right)$$

$$= 0,4(-0,25 \leq D' \leq 0,25) + 0,6(-0,5 \leq D' \leq 0)$$

$$= 0,4\left(\Phi(0,25) - (1-\Phi(0,25))\right) + 0,6\left(\Phi(0) - (1-\Phi(0,5))\right)$$

$$= 0,4(0,5987 - 1 + 0,5987) + 0,6(0,5 - 1 + 0,6915) = \underline{\underline{0,194}};$$

b) Se ha logrado estabilizar la máquina B para que $\mu_B = 1$ cm. ¿Qué valor debería tomar σ para cumplir los requerimientos de calidad con probabilidad mayor o igual que 0,90?

$$P(E_B) = P_B(0,9 \leq D \leq 1,1) = 0,90; \quad P\left(\frac{0,9-1}{\sigma} \leq D' \leq \frac{1,1-1}{\sigma}\right) = 0,90;$$

$$P(-0,10/\sigma \leq D' \leq 0,10/\sigma) = 0,90; \quad \Phi(0,10/\sigma) - \Phi(-0,10/\sigma) = \Phi(0,10/\sigma) - 1 + \Phi(0,10/\sigma)$$

$$= 2.\Phi(0,10/\sigma) - 1 = 0,90; \quad \Phi(0,10/\sigma) = \frac{0,9+1}{2} = 0,95$$

$$0,10/\sigma = 1,645; \quad \frac{1}{\sigma} = 1,645/0,10; \quad \sigma = 0,10/1,645 = 0,061;$$

EJERCICIO 11: Suponga que la dureza Rockwell de cierta aleación está normalmente distribuida con media de 70 y DE de 3.

A) Un especimen es aceptable si su dureza está entre 67 y 75, ¿Cual es la probabilidad de que un especimen seleccionado al azar tenga una dureza aceptable?

X = "dureza Rockwell"

$$P(67 \leq X \leq 75) =$$

$$P\left(\frac{67-70}{3} \leq Z \leq \frac{75-70}{3}\right) = P(-1 \leq Z \leq 1,67)$$

$$= \Phi(1,67) - \Phi(-1) = \Phi(1,67) - (1 - \Phi(1)) = \Phi(1,67) - 1 + \Phi(1) = 0,9525 - 1 + 0,8413$$

$$= \underline{\underline{0,79}};$$

b) Si la escala aceptable de dureza es $(70-c, 70+c)$, ¿Para qué valor de c tendría una dureza aceptable el 95% de todos los especímenes?

$$P(70-c \leq X \leq 70+c) = P\left(\frac{70-c-70}{3} \leq Z \leq \frac{70+c-70}{3}\right) = P(-c/3 \leq Z \leq c/3) = 0,95$$

$$= \Phi(c/3) - \Phi(-c/3) = \Phi(c/3) - 1 + \Phi(c/3) = 2\Phi(c/3) - 1 = 0,95;$$

$$\Phi(c/3) = 0,975 \Rightarrow c/3 = 1,96; \quad \boxed{c = 5,88};$$

Scanned by CamScanner

c) Si la escala aceptable es como en a), y la dureza de diez especímenes seleccionados al azar se determina independientemente, ¿Cuál es el número esperado de especímenes aceptables entre los 10?

EJERCICIO 14: El tiempo semanal, en horas, Y, durante el cual cierta máquina industrial no funciona, tiene aproximadamente una distribución Gamma con $\alpha = 1000$ y $\beta = 20$. La pérdida, en pesos, para la operación industrial debido a esta baja, está dada por $L = 30Y + 2Y^2$; Calcule $E(L)$ y $V(L)$.

$E(L) = E(30Y + 2Y^2) = 30E(Y) + E(Y^2) = 30(1000 \cdot 20) + E(Y^2) = 600.000 + E(Y^2)$

$V(Y) = E(Y^2) - [E(Y)]^2 = E(Y^2) - 360.000.000.000 = 400000;$

$E(Y^2)/100.000 = 4 - 3600000;\ E(Y^2) = -3599996 \cdot 100000$
$= -359999600000$

Scanned by CamScanner

PRÁCTICO 5

EJERCICIO 1: Sea X_1 el nro de clientes que están en espera en caja común, y X_2 el nro de clientes en caja rápida en ese mismo momento. fpc dada por la tabla (...)

a) ¿Cuál es la probabilidad de que haya exactamente 1 cliente en cada caja?

$P(X_1 = 1 \ \& \ X_2 = 1) = 0{,}15 = p(1,1)$

b) ¿Cuál es la probabilidad de que haya exactamente el mismo nro de clientes en las dos líneas?

$P(X_1 = X_2) = p(0,0) + p(1,1) + p(2,2) + p(3,3) = 0{,}08 + 0{,}15 + 0{,}10 + 0{,}07 = 0{,}4$

c) Sea A el evento de que haya por lo menos dos clientes más en una línea que la otra, ¿cuál es su prob.?

$A = \{(x,y) \mid |x-y| \geq 2\} = \{(0,2),(0,3),(1,3),\cancel{(1,4)},\cancel{(2,4)},(4,2),(4,1),(3,1),(3,0),(2,0)\}$

$P(A) = p(0,2) + p(2,0) + p(0,3) + p(3,0) + p(1,3) + p(3,1) + p(4,2) + p(4,1)$

$= 0{,}04 + 0{,}05 + 0{,}04 + 0{,}03 + 0{,}05 + 0{,}01 = \underline{\underline{0{,}22}}$

d) ¿Cuál es la probabilidad de que el nro total de clientes de las dos líneas sea exactamente 4?

$A = \{(x,y) \mid x + y = 4\} = \{(1,3),(2,2),(3,1),(4,0)\}$

$P(A) = p(1,3) + p(2,2) + p(3,1) + p(4,0) = 0{,}04 + 0{,}10 + 0{,}03 + 0 = \underline{\underline{0{,}17}}$

¿Y por lo menos 4?

$B = \{(x,y) \mid (x+y) \geq 4\}$; $P(B) = P(A) + p(2,3) + p(3,2) + p(3,3) + p(4,1) + p(4,2) + p(4,3)$

$= (0{,}17) + 0{,}06 + 0{,}04 + 0{,}07 + 0{,}01 + 0{,}05 + 0{,}06 = (0{,}17) + (0{,}29) = \underline{\underline{0{,}46}}$

e) Hallar las funciones de probabilidad marginales de X_1 y X_2, ¿Son independientes? Justificar.

$P_{X_1}(0) = 0{,}19$; $P_{X_1}(1) = 0{,}3$; $P_{X_1}(2) = 0{,}25$; $P_{X_1}(3) = 0{,}14$; $P_{X_1}(4) = 0{,}12$

$P_{X_2}(0) = 0{,}19$; $P_{X_2}(1) = 0{,}3$; $P_{X_2}(2) = 0{,}28$; $P_{X_2}(3) = 0{,}23$;

$p(0,0) = 0{,}08 \neq P_{X_1}(0) \cdot P_{X_2}(0) = 0{,}0361 \Rightarrow$ No son independientes;

EJERCICIO 2: Sean X, Y v.a con fdc:

$$f(x,y) = \begin{cases} k(x+y) & \text{si } 0 \leq x \leq 10 \text{ y } 0 \leq y \leq 10 \\ 0 & \text{c.c} \end{cases}$$

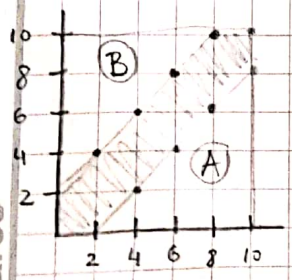
a) ¿Cuál es el valor de k? Sabemos que $\int_0^{10}\int_0^{10} k(x+y)\,dx\,dy = 1 \Rightarrow$

$$\int_0^{10}\int_0^{10} k(x+y)\,dx\,dy = \int_0^{10} k\left(\int_0^{10}(x+y)\,dx\right)dy = \int_0^{10} k\left[\left(\frac{x^2}{2} + xy\right)\Big|_0^{10}\right]dy = \int_0^{10} k\left[50 + 10y\right]dy$$

$$= k\int_0^{10} 50 + 10y\,dy = k\left(\int_0^{10} 50\,dy + \int_0^{10} 10y\,dy\right) = k\left(\left((50x)\right)\Big|_0^{10} + \cancel{10}^{5}\left(y^2/\cancel{2}\right)\Big|_0^{10}\right)$$

$= k(500 + 500) = k \cdot 1000 = 1$; $\boxed{k = 1/1000}$

b) ¿Cuál es la probabilidad de que el valor absoluto de la diferencia entre las variables aleatorias sea a lo sumo 2?



$A = B$

$P[1 - (A + B)] = P[1 - 2A]$

$= 1 - 2P(2 \leq x \leq 10 \ \& \ 0 \leq y \leq x-2)$

$$P(2 \leq x \leq 10 \ \& \ 0 \leq y \leq x-2) = \frac{1}{1000}\int_2^{10}\int_0^{x-2} x + y\,dy\,dx = \frac{1}{1000}\int_2^{10}\left[\left(xy + y^2/2\right)\Big|_0^{x-2}\right]dx$$

$$= \frac{1}{1000}\int_2^{10} x(x-2) + \frac{(x-2)^2}{2}\,dx = \frac{1}{1000}\int_2^{10}(x-2)\left(x + \frac{x-2}{2}\right)dx = \frac{1}{1000}\int_2^{10}(x-2)\left(\frac{3}{2}x - 1\right)dx =$$

Scanned by CamScanner

$$\cancel{= \frac{1}{1000}\left(\left(\frac{2}{3}\frac{x^3}{3} - 2x^2 + 2x\right)\Big|_2^{10}\right) = \frac{1}{1000}\left(\frac{2}{9}1000 - 200 + 20 - \frac{2}{9}8 + 8 - 4\right) =}$$

$$\cancel{= \frac{1}{1000}(42{,}\hat{2} - 5{,}778) = \frac{1}{1000}}$$

$$= \frac{1}{1000}\int_2^{10}\left(\frac{3}{2}x^2 - 3x - x + 2\right)dx = \frac{1}{1000}\left(\frac{3}{2}\frac{x^3}{3} - 2x^2 + 2x\right)\Big|_2^{10} = \frac{1}{1000}\left(\frac{1000}{2} - 200 + 20 - 4 + 8 - 4\right)$$

$$= \frac{320}{1000} = 0{,}32;\ P = 1 - 2.0{,}32 = \underline{\underline{0{,}36}};$$

c) Calcular $P(X+Y<5) = P(Y<5-X) = \int_0^5\int_0^{5-x}\frac{1}{1000}(x+y)\,dy\,dx$

$$= \int_0^5 \frac{1}{1000}\int_0^{5-x} x+y\,dy\,dx = \frac{1}{1000}\int_0^5\left(xy + y^2/2\right)\Big|_0^{5-x}dx = \frac{1}{1000}\int_0^5\left(x(5-x) + \frac{(5-x)^2}{2}\right.dx$$

$$= \frac{1}{1000}\int_0^5 (5-x)\left(x + \frac{(5-x)}{2}\right)dx = \frac{1}{1000}\int_0^5 (5-x)\left(\frac{2x+5-x}{2}\right)dx = \frac{1}{1000}\int_0^5 (5-x)\left(\frac{x}{2} + \frac{5}{2}\right)dx$$

$$= \frac{1}{1000}\int_0^5\left(\cancel{\frac{5}{2}x} + \frac{25}{2} - \frac{x^2}{2} - \cancel{\frac{5}{2}x}\right)dx = \frac{1}{1000}\int_0^5 \frac{25}{2} - \frac{x^2}{2}\,dx = \frac{1}{2000}\left(\frac{25}{2}x - \frac{x^3}{3}\right)\Big|_0^5$$

$$= 0{,}04167;$$

d) Hallar las funciones de densidad marginales de X e Y, ¿Son independientes? Justifique.

$$f_X(x) = \int_0^{10}\frac{1}{1000}(x+y)\,dy = \frac{1}{1000}\int_0^{10} x+y\,dy = \frac{1}{1000}\left(xy + y^2/2\right)\Big|_0^{10}$$

$$= \frac{1}{1000}(10x + 50) = \frac{1}{100}x + \frac{1}{20};$$

$$f_Y(y) = \int_0^{10}\frac{1}{1000}(x+y)\,dx = \frac{1}{1000}\int_0^{10}(x+y)\,dx = \frac{1}{1000}\left(\frac{x^2}{2} + yx\right)\Big|_0^{10} = \frac{1}{1000}(50 + 10y) = \frac{1}{100}y + \frac{1}{20};$$

$$f(1,3) = \frac{1}{1000}(4) = \frac{1}{250} = 4\times10^{-3} = 0{,}0040 \qquad f_X(1).f_Y(3) = \left(\frac{1}{100} + \frac{1}{20}\right)\left(\frac{3}{100} + \frac{1}{20}\right) = \left(\frac{1+2000}{100}\right)(200 = 4{,}8\cdot10^{-3} = 0{,}0048$$

No son independientes

e) Calcular $Cov(X,Y)$: $Cov(X,Y) = E(XY) - E(X).E(Y)$

$$E(X) = \int_0^{10} x f_X(x)\,dx = \int_0^{10} x\left(\frac{x+5}{100}\right)dx = \frac{1}{100}\int_0^{10} x^2 + 5x\,dx = \frac{1}{100}\left(\frac{x^3}{3} + \frac{5}{2}x^2\right)\Big|_0^{10}$$

$$= \frac{1000}{300} + \frac{500}{200} = \frac{10}{3} + \frac{5}{2} = \frac{20+15}{6} = \frac{35}{6} = E(Y)$$

$$\simeq 5{,}84;$$

husares

Scanned by CamScanner

PRÁCTICO 5 - hoja 2

$$E(XY) = \int_0^{10}\int_0^{10} xy\, f(x,y)\,dx\,dy = \frac{1}{1000}\int_0^{10}\int_0^{10} xy(x+y)\,dx\,dy = \frac{1}{1000}\int^{10}\int^{10} x^2y + y^2x\; dx\; dy$$

$$= \frac{1}{1000}\int_0^{10}\left(y\frac{x^3}{3} + y^2\frac{x^2}{2}\right)\Big|_0^{10} dy = \frac{1}{1000}\int_0^{10} \frac{1000}{3}y + 50y^2\, dy = \frac{1}{1000}\left(\frac{\overset{500}{\cancel{1000}}}{3}\frac{y^2}{\cancel{2}_1} + \frac{50}{3}y^3\right)\Big|_0^{10} =$$

$$= \frac{1}{1000}\left(\frac{50}{3}(10)^3 + \frac{50}{3}(10)^3\right) = \frac{1000}{1000}\cdot\frac{1}{3}10^3 = \frac{10^3}{3} = 33,\overline{33}$$

$$COV(X,Y) = 33,\overline{33} - (5,84)^2 = \underline{-0,6589}$$

EJERCICIO 3: Sea X el nro de errores de mecanografía de un primer artículo y e Y del 2º. Suponga que X e Y son independientes con distribución de Poisson de parámetros λ_1 y λ_2 respectivamente.

a) Dar la función de probabilidad conjunta de (X, Y)

Como son independientes $p(x,y) = p_x(x)\cdot p_y(y)$

$\Rightarrow p_x(x) = e^{-\lambda_1}\cdot\frac{\lambda_1^x}{x!}$ y $p_y(y) = e^{-\lambda_2}\cdot\frac{\lambda_2^y}{y!}$

$$p(x,y) = \begin{cases} e^{-\lambda_1-\lambda_2}\cdot\dfrac{\lambda_2^y\cdot\lambda_1^x}{x!\,y!} & \text{si } x,y>0 \\ 0 & \text{c.c} \end{cases}$$

b) ¿Cuál es la probabilidad de que a lo sumo se cometa un error entre los dos artículos?

$$P(X+Y\le 1) = P(Y\le 1-X) = \sum_{x=0}^{1}\sum_{y=0}^{1-x} p(x,y) = \sum_{y=0}^{1} p(0,y) + \sum_{y=0}^{0} p(1,y)$$

$$= p(0,0) + p(0,1) + p(1,0)$$

$$= e^{-(\lambda_1+\lambda_2)}\cdot\frac{\lambda_2^0\lambda_1^0}{0!\,0!} + e^{-(\lambda_1+\lambda_2)}\cdot\frac{\lambda_1^0\lambda_2^1}{0!\,1!} + e^{-(\lambda_1+\lambda_2)}\frac{\lambda_1^1\lambda_2^0}{1!\,0!} =$$

$$= e^{-(\lambda_1+\lambda_2)} + e^{-(\lambda_1+\lambda_2)}\lambda_2 + e^{-(\lambda_1+\lambda_2)}\lambda_1 = e^{-(\lambda_1+\lambda_2)}(1+\lambda_1+\lambda_2)$$

c) $P(X+Y=m) = P(Y=m-X) = \sum_{x=0}^{m}\sum_{y=0}^{m-x} p(x,y)$

(...)

EJERCICIO 4: Una persona tiene dos bombillas para una lámpara en particular. Sea X e Y el tiempo de duración, en miles de horas, para ambos bombillas respectivamente. Suponga que X e Y son independientes e idénticamente distribuidas, con distribución exponencial de parámetro $\lambda = 1$

a) Dar la función de densidad conjunta de (X, Y)

$$f_x(x) = \begin{cases} e^{-x} & x\ge 0 \\ 0 & \text{c.c} \end{cases} \qquad f_y(y) = \begin{cases} e^{-y} & y\ge 0 \\ 0 & \text{c.c} \end{cases}$$

$f(x,y) = f_x(x)\cdot f_y(y)$ porque son independientes $\Rightarrow f(x,y) = \begin{cases} e^{-(x+y)} & x,y\ge 0 \\ 0 & \text{c.c} \end{cases}$

húsares

Scanned by CamScanner

b) ¿Cuál es la probabilidad de que ambas bombillas duren a lo sumo 1000 hs?

$$P(X\leq 1;\ Y\leq 1)=\int_0^1\int_0^1 e^{-(x+y)}\,dx\,dy=\int_0^1\int_0^1 e^{-x}\cdot e^{-y}\,dx\,dy$$

$$=\int_0^1 e^{-y}\int_0^1 e^{-x}\,dx\,dy=\int_0^1 e^{-y}\left(-e^{-x}\Big|_0^1\right)dy=\int_0^1 e^{-y}\left(-\frac{1}{e}+1\right)dy=\left(1-\frac{1}{e}\right)\int_0^1 e^{-y}\,dy$$

$$=\left(1-\frac{1}{e}\right)\left(1-\frac{1}{e}\right)=\left(1-\frac{1}{e}\right)^2=0{,}399\simeq 0{,}4;$$

c) ¿Cuál es la probabilidad de que la duración total entre las dos bombillas sea a lo sumo 2000 hs?

$$P(X+Y\leq 2)=P(Y\leq 2-X)=\int_0^2\int_0^{2-x} e^{-(x)}\cdot e^{-y}\,dy\,dx=\int_0^2 e^{-x}\int_0^{2-x} e^{-y}\,dy\,dx=\int_0^2 e^{-x}\left(-e^{-y}\Big|_0^{2-x}\right)dx=$$

$$=\int_0^2 e^{-x}\left(-e^{x-2}+1\right)dx=\int_0^2\left(-e^{-2}+e^{-x}\right)dx=-\int_0^2 -e^{-2}\,dx+\int_0^2 e^{-x}\,dx=\left(-xe^{-2}\Big|_0^2\right)+\left(-e^{-x}\Big|_0^2\right)$$

$$=\left(-2e^{-2}\right)+\left(-e^{-2}+1\right)=-3e^{-2}+1\simeq 0{,}59$$

EJERCICIO 5:

a) Demuestre que si X e Y son variables independientes, entonces $E(XY)=E(X)E(Y)$

Si X e Y son independientes entonces $Corr(X,Y)=0$;

Como $Corr(X,Y)=\dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}=0 \Rightarrow Cov(X,Y)=0$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0 \Rightarrow E(XY)=E(X)E(Y)$$

b) Un topógrafo desea marcar en un terreno un cuadrado de longitud L en cada lado. Sin embargo, debido a un error de medición, traza un rectángulo donde los lados N-S tienen una longitud X y los lados E-O tienen una longitud Y. Suponga que X e Y son independientes y que cada una tiene una distribución uniforme en el intervalo $[L-a, L+a]$ (donde $0<a<L$)

¿Cuál es el área del rectángulo resultante?

Área del rectángulo: $XY \rightarrow E(XY)$? Por el resultado anterior:

$$E(XY)=E(X)E(Y);\quad f_X(x)=\begin{cases}\dfrac{1}{L+a-(L-a)}=\dfrac{1}{2a} & L-a\leq x\leq L+a\\ 0 & \text{c.c}\end{cases}$$

$$E(X)=\int_{L-a}^{L+a} x\cdot f_X(x)\,dx=\int_{L-a}^{L+a} x\,\frac{1}{2a}\,dx=\frac{1}{2a}\int_{L-a}^{L+a} x\,dx=\frac{1}{2a}\left(\frac{x^2}{2}\Big|_{L-a}^{L+a}\right)=\frac{1}{2a}\left(\frac{(L+a)^2}{2}-\frac{(L-a)^2}{2}\right)$$

$$=\frac{1}{2a}\left(\frac{(L+a)^2-(L-a)^2}{2}\right)=\frac{1}{2a}\left(\frac{(2a)(2L)}{2}\right)=\frac{1}{2a}(2aL)=L=E(Y)$$

$$E(XY)=2L;$$

Scanned by CamScanner

PRÁCTICO 5 - hoja 3

EJERCICIO 6: Suponga que la función de distribución de probabilidad conjunta de (X, Y) está dada por la siguiente tabla (...) donde se cumple que $a + b = 1/4$

a) Demostrar que $E(XY) = E(X)E(Y)$ y luego $\rho = 0$

$P_X(-1) = a + b + a = 1/4 + a$; $P_Y(-1) = a + b + a = 1/4 + a$

$P_X(0) = b + b = 2b$; $P_Y(0) = b + 0 + b = 2b$

$P_X(1) = a + b + a = 1/4 + a$; $P_Y(1) = a + b + a = 1/4 + a$

$$E(X) = \sum_{x=-1}^{1} x\, P_X(x) = (-1)(1/4 + a) + (1)(1/4 + a) = 0 = E(Y) \Rightarrow E(X)E(Y) = 0$$

$$E(XY) = \sum_{x=-1}^{1} \sum_{y=1}^{1} xy \cdot p(x,y) = (-1)(-1)(a) + (-1)(0)(b) + (-1)(1)(a) + (1)(-1)(a) + (1)(1)(a)$$

$$= a - a - a + a = 0 = E(X)E(Y);$$

$\boxed{\rho = 0}$ $\quad \rho_{X,Y} = \dfrac{cov(X,Y)}{\sigma_X \sigma_Y}$; $\quad cov(X,Y) = E(XY) - E(X)E(Y) = 0 - 0 = 0$

$\Rightarrow \rho_{X,Y} = 0$

b) ¿Son X e Y independientes? Son independientes si para todo par (x,y) se da que $p(x,y) = p(x)p(y)$

$P_X(1) = 1/4 + a$; $P_Y(0) = 2b$

$p(1,0) = b \neq 2b(1/4 + a) = P_X(1) \cdot P_Y(0)$; $\Rightarrow$ NO SON INDEPENDIENTES

EJERCICIO 7: Suponga que Y_1 e Y_2 son V.A. tales que: $E(Y_1) = 2$; $E(Y_2) = -1$;

$\rho(Y_1, Y_2) = \frac{1}{2\sqrt{6}} \Rightarrow \{cov(Y_1, Y_2) = 1\}$; $var(Y_1) = 4$; $VAR(Y_2) = 6$

Obtener $E(3Y_1 - 2Y_2) = 3E(Y_1) - 2E(Y_2) = 3(2) - 2(-1) = 6 + 2 = 8$;

$Var(3Y_1 - 2Y_2) =$ (no entiendo)

$Cov(3Y_1 - 2$

Scanned by CamScanner

EJERCICIO 8: Suponga que la densidad del sedimento (g/cm) de un especimen, seleccionado al azar de cierta región, está normalmente distribuída con una media de 2,65 y desviación estándar de 0,85.

a) Si se le selecciona al azar una muestra aleatoria de 25 especímenes

I - ¿Cuál es la probabilidad de que el promedio muestral de la densidad del sedimento sea a lo sumo 3?

$P(\bar{X} \leq 3)$?

$\bar{X} = \frac{X_1 + \dots X_{25}}{25} \Rightarrow \bar{X} \sim N(2,65 ; \frac{(0,85)^2}{25})$

$P(\bar{X} \leq 3) = P(Z \leq \frac{3-2,65}{\sqrt{0,0289}}) = P(Z \leq 2,06) = 0,9803 = \Phi(2,06)$

II - ¿Y que se encuentre entre 2,65 y 3?

$P(2,65 \leq \bar{X} \leq 3) = P(0 \leq Z \leq 2,06) = P(Z=2,06) - P(Z=0) = \Phi(2,06) - \Phi(0)$

$= 0,9803 - 0,500 = 0,4803$

b) ¿Qué tan grande se requeriría el tamaño muestral para asegurar que la primera probabilidad calculada sea, por lo menos, 0,99?

("por lo menos")

$P(\bar{X} \leq 3) \geq 0,99 ; P(Z \leq \frac{3-2,65}{0,85/\sqrt{n}}) \geq 0,99 \Rightarrow \cancel{P(Z \leq 0,35/(0,85/\sqrt{n})) = 0,99}$

$\Phi(0,35/(0,85/\sqrt{n})) \geq 0,99 ; \Phi(\frac{0,35}{0,85} \cdot \sqrt{n}) \geq 0,99 ; \Phi(0,412 \cdot \sqrt{n}) \geq 0,99 ;$

$0,412 \cdot \sqrt{n} \geq 2,33 ; n \geq (\frac{2,33}{0,412})^2 ; n \geq 32 ;$

EJERCICIO 9: La primera tarea en un curso introductorio de programación por computadora implica correr un breve programa. Si la experiencia indica que el 40% de todos los estudiantes principiantes no cometerán typos, calcular la probabilidad aproximada de que en un grupo de 50 estudiantes:

a) Por lo menos 25 no cometan errores?

X = "número de estudiantes de la muestra que no cometieron errores"

$X \sim b(50; 0,4) \Rightarrow$ pero como $n > 25$, usamos TLC, y $\bar{X} \simeq N(\mu_{\bar{X}}, \sigma^2_{\bar{X}})$

$E(X) = 50 \cdot 0,4 = n \cdot p = 20 ; V(X) = n \cdot p(1-p) = 20 \cdot 0,6 = 12 ; \sigma_X = \sqrt{12} \simeq 3,46 ;$

$P(X \geq 25) = 1 - P(X < 25) = 1 - P(X \leq 24) = 1 - P(Z \leq \frac{24-20}{3,46}) = 1 - P(Z \leq 1,156)$

$= 1 - \Phi(1,16) = 1 - 0,8770 = 0,123 ; P(X \geq 25) \simeq 0,123 ;$

b) Entre 15 y 25 no cometan errores:

$P(15 \leq X \leq 25) = P(-1,44 \leq Z \leq 1,44) = \Phi(1,44) - \Phi(-1,44) = \Phi(1,44) - 1 + \Phi(1,44)$

$= 2\Phi(1,44) - 1 = 0,8502 ;$

húsares

Scanned by CamScanner

PRÁCTICO 5 - hoja 4

EJERCICIO 10: Suponga que el 10% de todos los ejes de acero producidos por cierto proceso están fuera de las especificaciones, pero se pueden volver a trabajar (en lugar de tener que enviarlos a la chatarra). Considere una muestra aleatoria de 200 ejes y sea X el número de los que están fuera de especificaciones y se pueden volver a trabajar ¿Cuál es la probabilidad (aproximada) de que:

a) Sea a lo sumo 30?

$X \sim b(n,p) = b(200; 0{,}10) \Rightarrow E(X) = n.p = 200.0{,}10 = 20;$

$V(X) = n.p.q = 20.0{,}9 = 18$

$P(X \leq 30) = P(Z \leq \frac{30-20}{\sqrt{18}}) = P(Z \leq 2{,}36) = \Phi(2{,}36) = 0{,}9907$

b) Sea menor a 30? $P(X<30) = P(X \leq 29) = P(Z \leq 2{,}12) = \Phi(2{,}12) = 0{,}9830$

c) esté entre 15 y 25 inclusive? $P(15 \leq X \leq 25) = P(\frac{15-20}{\sqrt{18}} \leq Z \leq \frac{25-20}{\sqrt{18}}) =$

$P(-1{,}18 \leq Z \leq 1{,}18) = 2\Phi(1{,}18) - 1 = 2.0{,}8810 - 1 = 0{,}762$

EJERCICIO 11: Un antropólogo quiere estimar la estatura promedio de los hombres de cierta etnia. Se supone que la desviación estándar de la población es de 2,5 pulgadas.

a) Se selecciona al azar a 100 hombres. Encuentre la probabilidad aproximada de que la media muestral diste de la media verdadera de la población en a lo sumo 0,5 pulgadas. $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{2{,}5}{\sqrt{100}}$

$TLC \rightarrow P(\mu - 0{,}5 \leq \bar{X} \leq \mu + 0{,}5) = P(\frac{\mu - 0{,}5 - \mu}{2{,}5/\sqrt{100}} \leq Z \leq \frac{\mu + 0{,}5 - \mu}{2{,}5/\sqrt{100}})$

$= P(-0{,}2.\sqrt{100} \leq Z \leq 0{,}2.\sqrt{100}) = P(-2 \leq Z \leq 2) = 2\Phi(2) - 1 = 2.0{,}9772 - 1 = 0{,}9544$

b) Suponga ahora que el antropólogo quiere que la media muestral de la media de la población en a lo sumo 0,4 pulgadas, con una probabilidad de por lo menos 0,95 ¿Cuántos hombres tendría que seleccionar para alcanzar su objetivo?

$P(\frac{\mu - 0{,}4 - \mu}{2{,}5/\sqrt{n}} \leq Z \leq \frac{\mu + 0{,}4 - \mu}{2{,}5/\sqrt{n}}) = P(-\frac{0{,}4}{2{,}5} \cdot \sqrt{n} \leq Z \leq \frac{0{,}4}{2{,}5}\sqrt{n}) = P(-0{,}16\sqrt{n} \leq Z \leq 0{,}16\sqrt{n})$

$= 2.\Phi(0{,}16\sqrt{n}) - 1 = 0{,}95; \quad \Phi(0{,}16\sqrt{n}) = \frac{0{,}95+1}{2}; \quad \Phi(0{,}16\sqrt{n}) = 0{,}975$

$0{,}16\sqrt{n} = 1{,}96; \quad n = \left(\frac{1{,}96}{0{,}16}\right)^2 = 150{,}1 \simeq \underline{\underline{150}};$

EJERCICIO 12: Suponga que la resistencia esperada a la tensión del acero tipo A es de 106 ksi y la desviación estándar es de 8 ksi. Para el acero tipo B, suponga que la resistencia esperada a la tensión y la desviación estándar son de 104 ksi y 6 ksi, respectivamente. Sean $\bar{X}$ e $\bar{Y}$ las resistencias promedio a la tensión de los aceros tipo A y B con muestras de tamaño 40 y 35 respectivamente.

a) ¿Cuál es la distribución aproximada de $\bar{X} - \bar{Y}$? Por TLC $\Rightarrow$

$\bar{X} \sim N(106; 8^2/40)$ y $\bar{Y} \sim N(104; 6^2/35)$

$= \bar{X} \sim N(106; 1{,}6)$ y $\bar{Y} \sim N(104, 1{,}03)$ $\quad E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = 106 - 104 = \underline{\underline{2}}$

Suponemos que $\bar{X}$ e $\bar{Y}$ son independientes: $V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = 1{,}6 + 1{,}03 = \underline{2{,}63}$;

$(\bar{X} - \bar{Y}) \sim N(2; 2{,}63)$

b) Dar el valor aproximado de $P(-1 \leq \bar{X} - \bar{Y} \leq 1) = P(\frac{-1-2}{\sqrt{2{,}63}} \leq Z \leq \frac{1-2}{\sqrt{2{,}63}}) = P(-1{,}85 \leq Z \leq -0{,}61)$

$= \Phi(-0{,}61) - \Phi(-1{,}85) = (1 - \Phi(0{,}61)) - (1 - \Phi(1{,}85)) = -\Phi(0{,}61) + \Phi(1{,}85) = -0{,}7291 + 0{,}9678$

$= 0{,}2387 \simeq \underline{\underline{0{,}24}}$

húsares

Scanned by CamScanner

c) Dar el valor aproximado de $P(\bar{X}-\bar{Y} \geq 6) = 1 - P(\bar{X}-\bar{Y} \leq 5) = 1 - 0{,}9678 = \underline{0{,}0322}$

$P(\bar{X}-\bar{Y} \leq 5) = P\left(Z \leq \frac{5-2}{\sqrt{2{,}63}}\right) = P(Z \leq 1{,}85) = 0{,}9678$;

<u>EJERCICIO 13</u>: Suponga que cierto consumo de calorías en el desayuno D es una V.A con $E(D) = 500$ y $\sigma_D = 50$, el consumo de calorías al mediodía es A aleatorio con $E(A) = 800$ y $\sigma_A = 100$ y el consumo de calorías en la cena es $E(C) = 1700$ y $\sigma_C = 200$; Si se supone que los consumos son independientes entre sí y se realizan las tres comidas al día, ¿cuál es la probabilidad aproximada de que el promedio de consumo diario de calorías, en un año, esté comprendido entre 2950 y 3050?

$X =$ "Consumo diario" $= D + A + C$
Promedio

$\bar{X} \sim N(E(X), \sigma_X^2/365)$; $E(X) = E(D+A+C) = E(D) + E(A) + E(C) = 500 + 800 + 1700 = \underline{\underline{3000}}$

$V(X) = V(D+A+C) = V(D) + V(A) + V(C) = 50^2 + 100^2 + 200^2 = 52.500$

$\bar{X} \sim N(3000, 143{,}83)$; $P(2950 \leq \bar{X} \leq 3050) = P\left(\frac{2950-3000}{\sqrt{143{,}83}} \leq Z \leq \frac{3050-3000}{\sqrt{143{,}83}}\right)$

$= P(-4{,}17 \leq Z \leq 4{,}17) = 2\Phi(4{,}17) - 1 =$

Scanned by CamScanner

PRÁCTICO 6:

EJERCICIO 1: Se supone que el espesor de pintura (X) tiene distribución normal. Se realizaron las sig. observaciones de espesores (...)

a) Calcule la estimación puntual de la media de la distribución del espesor de pintura por el método de momentos?

$X \sim N(\mu, \sigma^2) \Rightarrow \overline{X} \sim N(\mu, \sigma^2_{\overline{X}})$; tamaño de la muestra: 16

para $m = 1$ $\hat{\mu} = \widehat{E(X)} = \left(\sum_{i=1}^{n} x\right)/n = 21{,}57/16 = 1{,}348$

$\hat{\mu} = 1{,}348 = \overline{X}$

b) Calcule la estimación puntual de la mediana?

$E(\tilde{X}) = \frac{\sum_i^n \tilde{x}}{n} = \frac{\sum_i^n (x_{n/2})}{n} =$

$\hat{\tilde{x}} = \tilde{x}_m$? $= 1{,}31$

c) Calcule la estimación puntual del percentil 90.

Scanned by CamScanner

PRÁCTICO 6:

EJERCICIO 2: $X_1, \dots, X_m$ es una muestra aleatoria de una distribución de resistencia de vigas de concreto con media μ_1 y desviación estándar σ_1, e $Y_1, \dots, Y_n$ es una muestra aleatoria de una distribución de resistencia de cilindros de concreto con media μ_2 y σ_2. Suponga que $X_1, \dots, X_m$ e $Y_1, \dots, Y_n$ son independientes entre sí.

a) Demuestre que $\bar{X} - \bar{Y}$ es un estimador insesgado de $\mu_1 - \mu_2$

$$E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = E\left(\frac{\sum_{i=1}^{m} X_i}{m}\right) - E\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \frac{1}{m} E\left(\sum_{i=1}^{m} X_i\right) - \frac{1}{n} E\left(\sum_{i=1}^{n} Y_i\right)$$

$$= \frac{1}{m} \sum_{i=1}^{m} E(X_i) - \frac{1}{n} \sum_{i=1}^{n} E(Y_i)$$

$$= \frac{1}{m}(m \cdot \mu_1) - \frac{1}{n} n \cdot \mu_2 = \mu_1 - \mu_2 ;$$

b) Obtenga la varianza y la desviación estándar (error estándar) del estimador del inuso a)

i) $$V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = V\left(\frac{\sum_{i=1}^{m} X_i}{m}\right) + V\left(\frac{\sum_{i=1}^{n} Y_i}{m}\right) = \frac{1}{m^2} \sum_{i=1}^{m} V(X_i) + \frac{1}{n^2} \sum_{i=1}^{n} V(Y_i)$$

$$= \frac{1}{m^2} m \sigma_1 + \frac{1}{n^2} n \sigma_2 = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{m}$$

ii) $$\sigma = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

c) Se obtuvieron las siguientes observaciones: (...)
Calcule la estimación de $\mu_1 - \mu_2$ y del error estándar del estimador del inciso a)

$$\widehat{(\mu_1 - \mu_2)} = \bar{X} - \bar{Y} = 7{,}46 - 8{,}575 = -1{,}115$$

$$\sigma = \sqrt{\frac{(0{,}97)^2 + (2{,}1)^2}{20}}$$ (FALTA)

EJERCICIO 3: Se examinan 150 piezas recién fabricadas y se registra el número de imperfecciones por pieza (tabla).

Sea X: número de imperfecciones en una pieza seleccionada al azar, y suponga que X tiene distribución Poisson de parámetro λ.

a) Encuentre un estimador insesgado de λ y calcule la estimación para los datos anteriores.

$\hat{\lambda}$ tal que $E(\hat{\lambda}) = \lambda$

Supongo que $\hat{\lambda} = \bar{X} \Rightarrow E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \lambda = \frac{1}{n} \cdot n \cdot \lambda = \lambda$

$$\bar{X} = \frac{317}{150} = 2{,}11 ;$$

b) ¿Cuál es la desviación estándar de este estimador? Calcule ~~este~~ el error estándar estimado.

$$V(\bar{X}) = V\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2} \cdot V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \lambda = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$$

$$\sigma = \sqrt{V(\bar{X})} = \sqrt{\frac{\lambda}{n}}$$ error estándar del estimador.

Scanned by CamScanner

EJERCICIO 4: Se seleccionan al azar n_1 fumadores (hombres) y n_2 fumadoras. Sean X_1 y X_2 variables aleatorias independientes que denotan, respectivamente, el número de fumadores que fuman cigarrillo con filtro. Denotamos con p_1 y p_2 las respectivas probabilidades de que un hombre y una mujer seleccionados al azar fumen cigarrillos con filtro.

a) Demuestre que $(X_1/n_1)-(X_2/n_2)$ es un estimador insesgado para p_1-p_2

$$X_1 \sim B(n_1,p_1) \; ; \; X_2 \sim B(n_2,p_2)$$

$$E\left(\frac{X_1}{n_1}-\frac{X_2}{n_2}\right) = E\left(\frac{X_1}{n_1}\right) - E\left(\frac{X_2}{n_2}\right) = \frac{1}{n_1}E(X_1) - \frac{1}{n_2}E(X_2) = \frac{1}{\cancel{n_1}}\cancel{n_1}\cdot p_1 - \frac{1}{\cancel{n_2}}\cdot \cancel{n_2}\cdot p_2$$

$$= p_1 - p_2$$

b) ¿Cuál es el error estándar del estimador del inciso a)?

$$V\left(\frac{X_1}{n_1}-\frac{X_2}{n_2}\right) = V\left(\frac{X_1}{n_1}\right) + V\left(\frac{X^2}{n_2}\right) = \frac{1}{n_1^2}V(X_1) + \frac{1}{n_2^2}V(X_2) = \frac{1}{n_1^2}\left(n_1\cdot p_1(1-p_1)\right) + \frac{1}{n_2^2}\left(n_2\cdot p_2(1-p_2)\right)$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \; ; \; \sigma = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

c) Si $n_1 = n_2 = 200$, $x_1 = 127$ y $x_2 = 176$, obtenga una estimación de $p_1 - p_2$.

$$\widehat{(p_1-p_2)} = \frac{x_1}{n_1} - \frac{x_2}{n_2} = \frac{127}{200} - \frac{176}{200} = \frac{127-176}{200} = -0{,}245$$

d) Con los datos del inciso c) obtenga una estimación del error estándar del estimador de $p_1 - p_2$:

$$\sigma = \sqrt{\frac{(127/200)(1-127/200)}{200} + \frac{(176/200)(1-176/200)}{200}} = \sqrt{\frac{0{,}635(0{,}365) + 0{,}88(0{,}12)}{200}}$$

$$= \sqrt{\frac{0{,}232 + 0{,}1056}{200}} = \sqrt{\frac{0{,}3376}{200}} = 0{,}041$$

EJERCICIO 5 Sea $X_1, \dots, X_n$ una muestra aleatoria, donde cada X_i tiene media μ y varianza σ^2.

a) Demuestre que $\overline{X}^2$ no es un estimador insesgado para μ^2.

$$E(\overline{X}^2) = V(\overline{X}) + (E(\overline{X}))^2 = V\left(\frac{\sum_{i=1}^n X_i}{n}\right) + \left(E\left(\frac{\sum_{i=1}^n X_i}{n}\right)\right)^2 = \frac{1}{n^2}\cdot n\cdot\sigma^2 + \frac{1}{n^2}\left(E\left(\sum_{i=1}^n X_i\right)^2\right)$$

$$= \frac{1}{n}\sigma^2 + \frac{1}{n^2}(n\cdot\mu)^2 = \frac{\sigma^2}{n} + \mu^2 \neq \mu^2$$

b) ¿Para qué valor de k es el estimador $\overline{X}^2 - kS^2$ insesgado para μ^2?

$$E(\overline{X}^2 - kS^2) = E(\overline{X}^2) - E(kS^2) = \left(\frac{\sigma^2}{n}+\mu^2\right) - kE(S^2) = \left(\frac{\sigma^2}{n}+\mu^2\right) - kE\left[\frac{\sum_{i=1}^n X_i^2 - \overline{X}^2}{n-1}\right] ;$$

$$E\left[\frac{\sum_{i=1}^n X_i^2 - \overline{X}^2}{n-1}\right] = \frac{1}{n-1}E\left[\sum_{i=1}^n X_i^2 - \overline{X}^2\right] = \frac{1}{n-1}\left(E\left(\sum_{i=1}^n X_i^2\right) - nE(\overline{X}^2)\right) =$$

$$\frac{1}{n-1}\left(E\left(\sum_{i=1}^n X_i^2\right) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) = \frac{1}{n-1}\left[\sum_{i=1}^n E(X^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right] = \frac{1}{n-1}\left[n\cdot(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right]$$

$$= \frac{1}{n-1}\left[n\cdot\left(\sigma^2+\cancel{\mu^2}-\frac{\sigma^2}{n}-\cancel{\mu^2}\right)\right] = \frac{1}{n-1}\left[n\sigma^2 - \sigma^2\right] = \frac{1}{\cancel{n-1}}(\cancel{n-1})\sigma^2 = \sigma^2 ;$$

$$E(\overline{X}^2 - kS^2) = \frac{\sigma^2}{n} + \mu^2 - k\sigma^2 = \mu^2 \Rightarrow k = \left(\cancel{\mu^2} - \cancel{\mu^2} - \frac{\sigma^2}{n}\right)\cdot\frac{-1}{\sigma^2} = \frac{1}{n} \Rightarrow \boxed{k = \frac{1}{n}}$$

húsares

Scanned by CamScanner

PRÁCTICO 6 - hoja 2;

EJERCICIO 6: Sea $X_1, \dots, X_n$ una muestra aleatoria de la función de densidad

$$f(x) = \begin{cases} 0{,}5(1+\theta x) & -1 \le x \le 1 \\ 0 & c.c \end{cases}$$

donde $-1 \le \theta \le 1$. Demostrar que $\hat{\theta} = 3\bar{X}$ es estimador insesgado de θ

Primero calculamos la esperanza para $f(x)$:

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{-1}^{1} x(0{,}5)(1+\theta x)\,dx = \frac{1}{2}\int_{-1}^{1} x + \theta x^2\,dx = \frac{1}{2}\left(\int_{-1}^{1} x\,dx + \int_{-1}^{1} \theta x^2 dx\right)$$

$$= \frac{1}{2}\left(\frac{x^2}{2}\right)\Big|_{-1}^{1} + \frac{1}{2}\left(\theta\frac{x^3}{3}\right)\Big|_{-1}^{1} = \frac{1}{2}\left(\frac{1}{2} - \frac{1}{2}\right) + \frac{1}{2}\left(\theta\frac{1}{3} + \theta\frac{1}{3}\right) = \frac{1}{6}\theta + \frac{1}{6}\theta = \frac{2}{6}\theta$$

$$= \boxed{\frac{\theta}{3}}\ ;$$ Ahora verificamos que $E(3\bar{X}) = \theta$

$$E(3\bar{X}) = 3E(\bar{X}) = 3\frac{\theta}{3} = \theta\ ] \checkmark$$

EJERCICIO 7: Se denota con X la proporción de tiempo que un estudiante, seleccionado al azar, emplea trabajando en cierta prueba de actitud. Se supone que X tiene función de densidad dada por

$$f(x) = \begin{cases} (\theta+1)x^{\theta} & 0 \le x \le 1 \\ 0 & c.c \end{cases}$$ donde $-1 < \theta$

a) Obtenga por el método de momentos un estimador de θ.

Como queremos estimar un sólo parámetro (θ), nuestro $m = 1$:

$$m = 1 \Rightarrow E(X^1) = E(X) = \bar{X} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$E(x) = \int_0^1 x(\theta+1)x^{\theta}\,dx = (\theta+1)\int_0^1 x^{\theta+1}\,dx = (\theta+1)\left(\frac{x^{\theta+2}}{\theta+2}\right)\Big|_0^1$$

$$= (\theta+1)\left(\frac{1^{\theta+2}}{\theta+2} - \frac{0^{\theta+2}}{\theta+2}\right) = (\theta+1)\left(\frac{1}{\theta+2}\right) = \frac{\theta+1}{\theta+2}$$

$$\Rightarrow \bar{X} = \frac{\theta+1}{\theta+2}\ ;\quad \theta+1 = (\theta+2)\bar{X} = \bar{X}\theta + 2\bar{X}\ ;\quad \theta - \bar{X}\theta = 2\bar{X} - 1\ ;$$

$$\theta(1-\bar{X}) = 2\bar{X} - 1\ ;\quad \boxed{\theta = \frac{2\bar{X}-1}{1-\bar{X}}}$$

b) Se toma una muestra aleatoria de 10 estudiantes: (...)
Calcule con esta información una estimación de θ, usando el obtenido en a)

$$\bar{X} = \frac{8}{10} = 0{,}8 \Rightarrow \theta = \frac{2(0{,}8)-1}{1-0{,}8} = \frac{0{,}6}{0{,}2} = 3 \Rightarrow \boxed{\theta = 3}$$

Scanned by CamScanner

EJERCICIO 8: El tiempo de respuesta X (en segundos) de cierta terminal de computadoras tiene distribución exponencial de parámetro $\lambda > 0$.

a) obtenga ~~que~~ el estimador de máxima verosimilitud de λ

valor de λ que incremente al máximo la función de probabilidad, tal que...

$$f(\hat{\lambda}) \geq f(\lambda); \qquad f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & c.c \end{cases}$$

$$f(x_1, \dots, x_n; \lambda) \geq f(x_1, \dots, x_n; \lambda);$$

$$f(x_1, \dots, x_n; \lambda) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x} = \lambda e^{-\lambda} \cdot \lambda e^{-2\lambda} \cdot \dots \cdot \lambda e^{-n\lambda}$$

$$= \lambda^n \cdot e^{-\lambda(x_1 + \dots + x_n)}$$

Aplico log: $\ln[f(x_1, \dots, x_n; \lambda) = \ln[\lambda^n \cdot e^{-\lambda(x_1 + \dots + x_n)}] = n \cdot \ln\lambda + \ln[e^{-\lambda(x_1 + \dots x_n)}]$

$$= n \cdot \ln\lambda + -\lambda(x_1 + \dots + x_n)\ln(e) = n \cdot \ln\lambda - \lambda(x_1 + \dots + x_n);$$

Aplico derivada: (sobre λ)

$$(n \cdot \ln\lambda - \lambda(x_1 + \dots + x_n))' = \frac{n}{\lambda} - (x_1 + \dots + x_n) = 0;$$

$$\frac{n}{\lambda} = \sum_{i=1}^{n} x_i; \quad \frac{1}{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}; \quad \lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$ es un punto crítico;

Luego calculo la derivada segunda y veo si el punto crítico es de un máximo

$$\left(\frac{n}{\lambda} - (x_1 + \dots + x_n)\right)' = \left(\frac{n}{\lambda}\right)' - (x_1 + \dots + x_n)' = (n \cdot \lambda^{-1})' = n(-1)\lambda^{-2} = -\frac{n}{\lambda^2}$$

que siempre será negativa $\Rightarrow$ $\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$ es un máximo.

Luego $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$ es un estimador de Máxima Verosimilitud.

b) Se hacen 10 observaciones: (...). Calcule con estos datos una estimación de λ usando el obtenido en a)

$$\hat{\lambda} = \frac{10}{55,8} = 0,18$$:)

Scanned by CamScanner

PRÁCTICO 6 - hoja 3

EJERCICIO 9 = Se supone que la resistencia al corte de soldaduras eléctricas (lb/plg^2) tiene distribución normal. Se determina la resistencia al corte de 10 soldaduras eléctricas obteniéndose: (...)

a) Estime el verdadero promedio y desviación estándar de la resistencia al corte, usando los estimadores de máxima verosimilitud

Los estimadores de máxima verosimilitud para la distribución normal son, en el caso de $\hat{\mu} = \bar{X}$ y $\sigma = S$

$$\Rightarrow \bar{X} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{3844}{10} = 384,4 = \hat{\mu}$$

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{3556,4}{9}} = \sqrt{395,15} = 19,88 = \hat{\sigma}$$

b) Estime el percentil 95 de la distribución (Use el principio de invarianza)

$$\eta_X(0,95) = \sqrt{\sigma^2} \cdot \eta_Z(0,95) + \mu$$

Usando los estimadores anteriores:

$$\eta_X(0,95) = 19,88 \cdot \eta_Z(0,95) + 384,4 = 19,88(1,645) + 384,4$$

$$= 417,1$$

c) Estime $P(X \leq 400)$ usando el principio de invarianza.

$$P(X \leq 400) = P\left(Z \leq \frac{400 - 384,4}{19,88}\right) = P(Z \leq 0,785) = \ldots$$

Scanned by CamScanner

PRÁCTICO 7:

EJERCICIO 1: La frecuencia de resonancia (en Hz) de las raquetas de tenis de cierto tipo tienen distribución normal con media μ y varianza σ^2 conocida. Se toma una muestra aleatoria de tamaño n.

a) ¿Cuál es el nivel de confianza del intervalo $\bar{x} \pm 2{,}81\,\sigma/\sqrt{n}$

$$P\left(-2{,}81\sigma/\sqrt{n} + x < \mu < \bar{x} + 2{,}81\frac{\sigma}{\sqrt{n}}\right) = 0{,}995$$

$$P(-2{,}81 \leq z \leq 2{,}81) = 2\Phi(2{,}81) - 1 = 0{,}995$$

b) ¿Cuál es el valor de $Z_{\alpha/2}$ que corresponde a un intervalo de confianza para μ del 99,7%?

$$P(-a < \mu < a) = 0{,}9970 \Rightarrow 2\Phi(a) - 1 = 0{,}9970 \; ; \; \Phi(a) = \frac{0{,}9970+1}{2}$$

$$\Phi(a) = 0{,}9985 \; ; \; a = 2{,}965 = Z_{\alpha/2}$$

c) Con la muestra dada se obtuvieron los siguientes intervalos de confianza para μ: (114,4; 115,6) y (114,1 , 115,9)

$$\bar{x} - Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} = 114{,}4 \; ; \; \bar{x} + Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} = 115{,}6 \; ; \; \bar{x} + Z'_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 114{,}1 \; ; \; \bar{x} - Z'_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 115{,}9 \; ;$$

$$\bar{x} - a = 114{,}4 \; ; \; \bar{x} + a = 115{,}6 \; ; \; \bar{x} - b = 114{,}1 \; ; \; \bar{x} + b = 115{,}9 \; ;$$

¿Cuál es el valor de la media muestral?

$\bar{x}_1 = 114{,}4 + a$; $\bar{x}_2 = 114{,}1 + b$

$\bar{x}_1 = 115{,}6 - a$; $\bar{x}_2 = 115{,}9 - b$

$b = \bar{x}_2 - 114{,}1$

$\bar{x}_2 = 115{,}9 - \bar{x}_2 + 114{,}1$

$\bar{x}_2 + \bar{x}_2 = 230{,}3$; $\boxed{x_2 = 115{,}15}$

$114{,}4 + a = 115{,}6 - a$

$2a = 115{,}6 - 114{,}4 = 1{,}2$;

$a = 1{,}2/2 = 0{,}6$

$$\boxed{\bar{x}_1 = 115}$$

¿Cuál de estos intervalos tiene mayor nivel de confianza?

El segundo, pues es más ancho.

EJERCICIO 2: Se desea un intervalo de confianza para la media verdadera μ de pérdida de carga para cierto tipo de motor de introducción (en watts), cuando la corriente de línea se mantiene en 10 amps para una velocidad de 1500 rpm. Suponga que la pérdida de carga está normalmente distribuída con $\sigma = 3$.

A) Calcular un intervalo de confianza del 95% para μ cuando $n = 25$ y $\bar{x} = 58{,}3$

$$P\left(\bar{x} - a\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + a\frac{\sigma}{\sqrt{n}}\right) = 0{,}95 \; ; \; 1-\alpha = 0{,}95 \; ; \; \alpha = 0{,}05 \Rightarrow \Phi(a) = 0{,}9750$$

$\alpha/2 = 0{,}025$ $\quad a = 1{,}96$

$$\Rightarrow \left(\bar{x} \pm a\cdot\frac{\sigma}{\sqrt{n}}\right) = 58{,}3 \pm 1{,}96\cdot\frac{3}{5} \longrightarrow 59{,}476$$

$$\searrow 57{,}124$$

Intervalo del 95% para μ = (57,124; 59,476)

ancho = 59,476 − 57,124 = 2,352

húsares

Scanned by CamScanner

b) lo mismo que antes pero con $n=100$ $(57,71 ; 58,89)$

$58,3 \pm 1,96 \cdot \frac{3}{10}$ → $58,89$ → $57,71$ ancho $= 58,89 - 57,71 = 1,178$

(El anterior es mucho más ancho).

c) Calcular el intervalo de confianza del 99% para μ cuando $n=100$ y $\bar{x}=58,3$

$\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 58,3 \pm Z_{\alpha/2} \cdot \frac{3}{10}$; $1-\alpha = 0,99$; $\alpha = 0,01$; $\alpha/2 = 0,005$

$Z_{0,005} \Rightarrow \Phi(Z_{0,005}) = 0,995 \Rightarrow Z_{0,005} = 2,575$

$58,3 \pm 2,575 \cdot \frac{3}{10}$ → $59,0725$ → $57,5275$ $\Rightarrow (57,5275 ; 59,0725)$

ancho $= 59,0725 - 57,5275 = 1,545$

Este es más ancho que el b) pues el nivel de confianza es mayor.

d) ¿Qué tan grande debería ser n para que la longitud del intervalo del 99% para μ sea a lo sumo 1?

$58,3 + 2,575 \cdot \frac{3}{\sqrt{n}} - \left(58,3 - 2,575 \cdot \frac{3}{\sqrt{n}}\right) \leq 1$;

$2 \cdot 2,575 \cdot \frac{3}{\sqrt{n}} \leq 1$; $\frac{15,45}{\sqrt{n}} \leq 1$; $\frac{\sqrt{n}}{15,45} \leq 1$; $\sqrt{n} \geq 15,45$;

$n \geq (15,45)^2$; $\boxed{n \geq 239}$

EJERCICIO 3: Un artículo analiza el uso de fotografía infrarroja en color para la identificación de árboles normales en bosques. Para una muestra de 69 árboles sanos, la media muestral de densidad de capa de tinte fue de 1,028 y la desviación estándar de 0,163.

a) Calcular un intervalo de confianza del 95% para la media poblacional de la densidad de capa de tinte μ para estos pinos?

$\bar{x} = 1,028$ $n = 69$ $\sigma = 0,163$

$1-\alpha = 0,95 \Rightarrow \alpha = 0,05$; $\alpha/2 = 0,025$ $\Phi(Z_{\alpha/2}) = 1 - 0,025 = 0,975$; $Z_{\alpha/2} = 1,96$

$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1,028 \pm 1,96 \cdot \frac{0,163}{\sqrt{69}}$ → $1,066$ → $0,989$ $\Rightarrow 0,989 < \mu < 1,066$

$\Rightarrow (0,989 ; 1,066)$

b) Suponga que los investigadores habían hecho una estimación de 0,16 para el valor de s antes de reunir los datos, ¿qué tamaño muestral sería necesario para obtener un intervalo de longitud a lo sumo 0,5 y con un nivel de confianza del 95%?

$s = \sigma = 0,16$; $1 - 0,95 = 0,05 \Rightarrow Z_{0,05/2} = 1,96$

$\Rightarrow \left|\overset{1,028}{\bar{x}} \pm 1,96 \cdot \frac{0,16}{\sqrt{n}}\right| \leq 0,5 \Rightarrow \left|1,028 \pm \frac{0,3136}{\sqrt{n}}\right| \leq 0,5 \Rightarrow$ FÓRMULA PARA EL ANCHO: w: ancho

$\left(2 \cdot Z_{\alpha/2} \cdot \frac{s}{w}\right)^2 \leq n \Rightarrow \left(2 \cdot 1,96 \cdot \frac{0,16}{0,05}\right)^2 \leq n \Rightarrow 158 \leq n$

Scanned by CamScanner

PRÁCTICO 7: hoja 2:

EJERCICIO 4: Se seleccionó una muestra aleatoria de 539 sujetos pertenecientes a cierta ciudad. Se determinó que 133 de ellos poseían por lo menos un arma de fuego.

a) Determine un intervalo de confianza del 98% para la verdadera proporción de propietarios de armas en esa ciudad.

Se busca estimar un intervalo para $\hat{p}$

La proporción muestral es $\hat{p}_n = \frac{133}{539} = 0,248$;

X = cantidad de sujetos con armas $\Rightarrow$ $X \sim b(n,p) \Rightarrow \bar{X} = n.p = 539 . 0,248 = 133$

$\sigma^2 = n.p.(1-p) = 539 . 0,248 . 0,752 = 100,52$

$1-0,98 = 0,02 = \alpha$ $Z_{\frac{0,02}{2}} = Z_{0,01} \Rightarrow \Phi(Z_{0,01}) = 1-0,01 = 0,99 \Rightarrow Z_{0,01} = 2,33$

$$\left(\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right. = 0,248 \pm 2,33 . \sqrt{\frac{0,186}{539}} = 0,248 \pm 2,33 . 0,018 \begin{cases} \to 0,291 \\ \to 0,206 \end{cases}$$

Intervalo: $(0,206 ; 0,291)$

b) ¿Qué tamaño de muestra sería necesario para obtener un intervalo de longitud a lo sumo 0,10 y con un nivel de confianza del 98% independiente del $\hat{p}$?

$$n \approx \frac{4 . Z^2_{\alpha/2} \hat{p}.\hat{q}}{w^2} = \frac{4 . (2,33)^2 . 0,248 . 0,752}{0,01}$$

(...)

EJERCICIO 5: Un triatlón es una competencia deportiva que incluye pruebas de natación, ciclismo y carrera. Un artículo reporta una investigación donde participaron triatletas hombres. Se registraron las máximas pulsaciones del corazón (latidos/minutos) durante la actuación en cada uno de los tres eventos. En natación, la media y desviación estándar muestrales fueron 188,0 y 7,2, respectivamente. Se supone que la distribución del ritmo cardíaco es aproximadamente normal.

A) Si participaron 40 atletas:

i- Determine un intervalo de confianza del 98% para la media poblacional de ritmo cardíaco en la prueba de natación. $X \sim N(\mu, \sigma^2)$

$\bar{X}_{n=1} = 188$ $\sigma = 7,2$

$1 - \alpha = 0,98 \Rightarrow \alpha = 0,02 \Rightarrow Z_{\alpha/2} = 2,33$

Intervalo:

$$188 \pm 2,33 . \frac{7,2}{\sqrt{40}} = 188 \pm 2,625 \begin{cases} \to 190,625 \\ \to 185,375 \end{cases}$$

$\Rightarrow 185,375 < \mu < 190,625 \Rightarrow (185,375 ; 190,625)$

ii) Determine un intervalo de confianza del 98% para la varianza poblacional de ritmo cardíaco en la prueba de natación.

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}} , \frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}}\right) = \left(\frac{39.(7,2)^2}{\chi^2_{1-0,01, 39}} , \frac{39(7,2)^2}{\chi^2_{0,01, 39}}\right) = \left(\frac{2021,76}{21,425} , \frac{2021,76}{62,426}\right) = (94,36 ; 32,3)$$

$= (32,39 ; 94,36)$

Scanned by CamScanner

B) Si participaron 9 atletas:

I - Determine un intervalo de confianza del 95% para la media poblacional de ritmo cardíaco en la prueba de natación.

$\bar{x} = 188 \quad \sigma = 7{,}2 \quad T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1) \quad 1-\alpha = 95;\ \alpha = 0{,}05$

$\alpha/2 = 0{,}025$

$$\left(\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}\right) = \left(\bar{x} - t_{0,025,8} \cdot \frac{7{,}2}{3},\ \bar{x} + t_{\alpha/2,n-1} \cdot \frac{7{,}2}{3}\right)$$

$$= (188 - 5{,}5344,\ 188 + 5{,}5344) = (182{,}46\ ;\ 193{,}53)$$

II - Determine un intervalo de confianza del 95% para la varianza poblacional de ritmo cardíaco en la prueba de natación.

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}}\ ;\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = \left(\frac{8.(7{,}2)^2}{\chi^2_{0,025,8}},\ \frac{8.(7{,}2)^2}{\chi^2_{0,975,8}}\right) = \left(\frac{414{,}72}{17{,}534},\ \frac{414{,}72}{2{,}180}\right) = (23{,}65\ ;\ 190{,}24)$$

EJERCICIO 6: El intervalo de confianza del 95% para la media poblacional de una distribución normal con varianza desconocida fue (229,764; 233,504), para una muestra de tamaño 5. Si ahora usted considera que es mejor un intervalo de confianza del 99%. ¿Cuáles son las cotas de ese intervalo?

Intervalo de nivel de confianza del 95% $(229{,}764;\ 233{,}504) \rightarrow \alpha/2 = 0{,}025$

Como $n = 5 \rightarrow$

$$229{,}764 = \bar{x} - t_{0,025,4} \cdot \frac{s}{\sqrt{5}} = \bar{x} - 2{,}776 \cdot \frac{s}{\sqrt{5}}\ ;\quad \bar{x} = 1{,}24 \cdot S + 229{,}764$$

$$\Downarrow$$

$$233{,}504 = \bar{x} + t_{0,025;4} \cdot \frac{s}{\sqrt{5}} = \bar{x} + 2{,}776 \cdot \frac{s}{\sqrt{5}}\ ;\quad (1{,}24 S + 229{,}764) + 1{,}24 \cdot S = 233{,}504$$

$$2{,}48 \cdot S = 3{,}74\ ;\quad \boxed{S = 1{,}51} \Rightarrow \boxed{\bar{x} = 231{,}63}$$

Intervalo de confianza del 99% $\rightarrow 1-\alpha = 0{,}99;\ \alpha = 0{,}01 \quad \alpha/2 = 0{,}005$

$$231{,}63 \pm t_{0,005,4} \cdot \frac{1{,}51}{\sqrt{5}} = 231{,}63 \pm 4{,}604 \cdot 0{,}675 \Rightarrow$$

- cota superior: 234,74
- cota inferior: 228,52

$(228{,}52\ ;\ 234{,}74)$

EJERCICIO 7: El tiempo de reacción TR ante un estímulo se estudia en un artículo, que dice la media muestral del TR para 16 nadadores, al arranque con un disparo, fue de 0,214' y la desviación estándar muestral fue de 0,036'.

Suponiendo que el TR tiene aproximadamente distribución normal, obtener:

a) Un intervalo de confianza del 90% para la media real del TR.

Nivel de confianza 90% $\Rightarrow 1-\alpha = 0{,}90;\ \alpha = 0{,}10,\ \alpha/2 = 0{,}05;\ n = 16;\ S = 0{,}036;\ \bar{x} = 0{,}214$

$$\bar{x} \pm t_{\alpha/2;n-1} \cdot \frac{s}{\sqrt{n}} = 0{,}214 \pm 1{,}753 \cdot \frac{0{,}036}{\sqrt{16}} \Rightarrow (0{,}214 - 0{,}016,\ 0{,}214 + 0{,}016);$$

$$= (0{,}198\ ;\ 0{,}23)$$

b) Un intervalo de confianza del 90% para DE del TR de los nadadores

$$\left(\sqrt{\frac{(n-1).S^2}{\chi^2_{\alpha/2;n-1}}},\ \sqrt{\frac{(n-1).S^2}{\chi^2_{1-\alpha/2;n-1}}}\right) = \left(\sqrt{\frac{15.(0{,}036)^2}{\chi^2_{0,05;15}}},\ \sqrt{\frac{15.(0{,}036)^2}{\chi^2_{0,95;15}}}\right) = (0{,}028\ ;\ 0{,}052)$$

húsares

Scanned by CamScanner

PRÁCTICO 7 - hoja 3

EJERCICIO 8: Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución uniforme en el intervalo $[0, \theta]$. Si $Y = \max(X_i)$, se puede demostrar que la función densidad de $U = Y/\theta$ está dada por:

$$f_U(u) = \begin{cases} n \cdot u^{n-1} & \text{si } 0 \leq u \leq 1 \\ 0 & \text{c.c} \end{cases}$$

A) Pruebe que $P\left((\alpha/2)^{1/n} \leq \frac{Y}{\theta} \leq (1-\alpha/2)^{1/n}\right) = 1 - \alpha$

Scanned by CamScanner

PRÁCTICO 8

EJERCICIO 1: Dos empresas distintas desean establecerse en cierta región y brindar servicios de televisión por cable. Se denota por p la proporción de subscriptores potenciales que prefieren la primera empresa sobre la segunda. Se ~~muestrea una~~ toma una muestra aleatoria de tamaño 25 para probar $H_0: p = 0{,}5$ contra $H_a: p \neq 0{,}5$. Se representa con X el número de suscriptores en la muestra que está a favor de la primera empresa y con x el valor observado de X.

$H_0: p = 0{,}5$ contra $H_a: p \neq 0{,}5$

a) ¿Cuál de las siguientes RR es la más adecuada?

$R_1 = \{x : x \leq 7 \text{ o } x \geq 18\}$ $R_2 = \{x : x \leq 8\}$ $R_3 = \{x : x \geq 17\}$

R_1 pues la hipótesis alternativa es bilateral.

b) ¿Cuál es ~~el estadístico~~ la distribución de probabilidad del estadístico de prueba X cuando H_0 es verdadera?

$H_0 \Rightarrow p = 0{,}5$ y como $n = 25$ no es suficientemente grande para TLC

$\Rightarrow X \sim B(n, p_0) = B(25, 0{,}5)$

c) Calcule la probabilidad de error de tipo II para la región seleccionada cuando $p = 0{,}3$, repita lo mismo para $p = 0{,}4$, $p = 0{,}6$ y $p = 0{,}7$.

ERROR TIPO 2: Aceptar H_0 cuando ésta es falsa.

$\beta(0{,}3) = (\text{Aceptar } H_0 \text{ cuando } X \sim bin(25, 0{,}3) = (8 \leq X \leq 17 \text{ cuando } X \sim bin(0{,}3))$

$$B(17; 25, 0{,}3) - B(8; 25, 0{,}3) = \sum_{i=8}^{17} \binom{17}{i} 0{,}3^{i} (0{,}7)^{17-i} = \binom{17}{8}(0{,}3)^{8}(0{,}7)^{9} + \binom{17}{9}(0{,}3)^{9}(0{,}7)^{8} + \binom{17}{10}(0{,}3)^{10}(0{,}7)^{7}$$
$$+ \binom{17}{11}(0{,}3)^{11}(0{,}7)^{6} + \binom{17}{12}(0{,}3)^{12}(0{,}7)^{5} + \binom{17}{13}(0{,}3)^{13}(0{,}7)^{4} + \binom{17}{14}(0{,}3)^{14}(0{,}7)^{3} + \binom{17}{15}(0{,}3)^{15}(0{,}7)^{2} + \binom{17}{16}(0{,}3)^{16}(0{,}7)^{1}$$
$$+ \binom{17}{17} 0{,}3^{17}$$

d) En ese caso, H_0 es correcta.

Scanned by CamScanner

EJERCICIO 2: La calibración de una balanza debe ser revisada después de pesar 25 veces un espécimen de prueba de 10 kg. Se supone que los resultados de diferentes pesos son independientes entre sí y que el peso en cada intento está normalmente distribuido con $\sigma = 0{,}200$ kg. Se representa con μ el verdadero promedio de lectura de peso de la balanza.

a) ¿Qué hipótesis probaría? $n = 25$

$H_0: \mu = 10$ kg $\quad H_a: \mu \neq 10$ kg

ESTADÍSTICO DE PRUEBA $= \boxed{Z = \dfrac{\bar{X} - 10}{0{,}2\sqrt{25}}}$ bajo H_0

b) Se supone que la balanza debe ser revisada si $\bar{x} \leq 9{,}8968$ o $\bar{x} \geq 10{,}1032$
¿Cuál es la probabilidad de que la revisión se realice cuando no sea necesaria?
Estoy suponiendo que H_0 es falsa cuando no lo es $\Rightarrow$ ERROR TIPO II

R.R sin estandarizar:

$\bar{X} \leq 9{,}8968$;

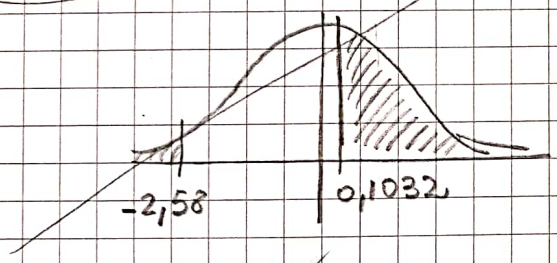
$\dfrac{\bar{X} - 10}{0{,}2\sqrt{25}} \leq \dfrac{9{,}8968 - 10}{0{,}2\sqrt{25}} = -2{,}58$

$Z \leq -2{,}58$

$\bar{X} \geq 10{,}1032$;

$\dfrac{\bar{X} - 10}{1} \geq \dfrac{10{,}1032 - 10}{1} = 0{,}1032$

$Z \geq 0{,}1032$



RR: $Z \leq -2{,}58 \;||\; Z \geq 0{,}1032$

$\Phi(-2{,}58) = 1 - \Phi(2{,}58) = 0{,}9951$

$\alpha = 1 - 0{,}9951 = 0{,}0049$

$\Phi(0{,}10) = 0{,}5398$;

$\beta(10{,}5) = P($

EJERCICIO 3: Considere que el estadístico de prueba Z tiene distribución normal estándar cuando Ho es verdadera. Proporcione el nivel de significación para cada una de las siguientes situaciones

A) $H_a: \mu > \mu_0$, región de rechazo $z \geq 1,88$

$$\alpha = P(Z \geq c) = P(Z \geq 1,88) = 1 - \Phi(1,88) = 1 - 0,9699 = 0,03;$$

B) $H_a: \mu < \mu_0$, región de rechazo $z \leq -2,75$

$$\alpha = P(Z \leq -2,75) = \Phi(-2,75) = 1 - \Phi(2,75) = 1 - 0,9970 = 0,003;$$

C) $H_a: \mu \neq \mu_0$, región de rechazo $z \geq 2,88$ o $z \leq -2,88$

$$\alpha = P(Z \geq 2,88 \text{ o } Z \leq -2,88) = P(Z \geq 2,88) + P(Z \leq -2,88) =$$

$$1 - \Phi(2,88) + (\Phi(-2,88)) = 1 - \Phi(2,88) + 1 - \Phi(2,88) = 2 - 2\Phi(2,88) = 2 - 2(0,9980)$$

$$= 0,004;$$

EJERCICIO 4: Se ha determinado el punto de fusión de cada una de las 16 muestras de cierta marca de aceite vegetal hidrogenado, con resultado $\bar{x} = 94,32$. Suponga que la distribución del punto de fusión es normal con $\sigma = 1,20$.

A) Pruebe $H_0: \mu = 95$ contra $H_a: \mu \neq 95$, utilizando una prueba de nivel 0,01 de dos colas

ESTADÍSTICO: $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{94,32 - 95}{1,20/\sqrt{16}} = -\frac{0,68}{0,3} = -2,26$$

$$\alpha = P(-Z_{\alpha/2} \geq z \text{ o } z \leq Z_{\alpha/2}) = 0,01 \Rightarrow Z_{0,005} = c \Rightarrow \Phi(c) = 1 - 0,005$$

$$= 0,995 \Rightarrow \boxed{Z_{0,005} = 2,575} \Rightarrow RR: \{z \mid z \leq -2,575 \parallel x \geq 2,575\}$$

Como Z observado $= -2,26$ está fuera de la zona de rechazo entonces no hay evidencia suficiente para rechazar Ho.

B) Si se utiliza una prueba de nivel 0,01, ¿Cuál es $\beta(94)$, cuando $\mu = 94$?

$$\beta(94) = P(\text{No rechazar Ho cuando } X \sim (94; 1,20)) =$$

$$\Phi\left(2,575 + \frac{95 - 94}{1,20\sqrt{16}}\right) - \Phi\left(-2,575 + \frac{95 - 94}{1,20\sqrt{16}}\right) = \Phi(2,575 + 3,33) - \Phi(-2,575 + 3,33)$$

$$= \Phi(5,91) - \Phi(0,755) = 1 - 0,7749 = 0,2251$$

C) ¿Qué valor de n es necesario para asegurar que $\beta(94) = 0,1$ cuando $\alpha = 0,01$?

$$n = \left[\frac{\sigma(Z_{\alpha/2} + Z_\beta)}{\mu_0 - \mu'}\right]^2 = \left[\frac{1,20(Z_{0,005} + Z_{0,01})}{95 - 94}\right]^2 = \left[1,20(Z_{0,005} + Z_{0,01})\right]^2 = \left[1,20(2,575 + 2,33)\right]^2$$

$$= 34,64 \approx 35$$

$$Z_{0,01} = c \Rightarrow \Phi(c) = 1 - 0,01 = 0,99 \Rightarrow Z_{0,01} = 2,33$$

Scanned by CamScanner

EJERCICIO 5: Se supone que el diámetro real de unas ruedas de cierto tipo es de 0,5 pulgadas. Para verificar esto, se hará una prueba t bilateral, ¿Qué conclusión es adecuada en cada uno de los casos siguientes?

a) $n=13$; $t_{obs}=1,6$; $\alpha=0,05$ $\quad H_a: \mu \neq 0,5$

ESTADÍSTICO: $t = \dfrac{\bar{x}-0,5}{s\sqrt{13}} = 1,6$

R.R: $H_a\ \mu \neq \mu_0 \Rightarrow t \geq t_{\alpha/2,n-1} \parallel t \leq -t_{\alpha/2,n-1}$

$= 1,6 \geq t_{0,025;12} \parallel 1,6 \leq -t_{0,025;12}$

$\Rightarrow t_{0,025;12} = 2,179$;

$RR: \{t \mid t \geq 2,179 \parallel t \leq -2,179\}$

$\Rightarrow$ Como $-2,179 < 1,6 < 2,179$ entonces H_0 no será rechazada

b) $n=13$ $\quad t_{obs}=-1,6$ $\quad \alpha=0,05$

Misma RR que en a) así que H_0 no será rechazada.

c) $n=25$; $t_{obs}=-2,6$; $\alpha=0,01$

ESTADÍSTICO $t=-2,6$ y $R.R: \{t \mid t \geq t_{\frac{0,01}{2},24} \parallel t \leq -t_{\frac{0,01}{2},24}\}$

$= \{t \mid t \geq 2,797 \parallel t \leq -2,797\}$

$\Rightarrow H_0$ no rechazada.

EJERCICIO 6: Se realizaron $n=26$ observaciones de tiempo de escape, en segundos, para trabajadores petroleros, en un ejercicio simulado, a partir de las cuales se calcularon el promedio y la desviación estándar muestrales, que fueron 370,69 y 24,26 respectivamente. Los investigadores suponen que el tiempo real promedio de escape es de 6 minutos a lo sumo ¿Contradicen los datos esta suposición? Asumiendo normalidad, pruebe las hipótesis apropiadas con un nivel de significación de 0,05.

$H_0: \mu = 360 \qquad H_a: \mu > 360$

Ya que $n=26$, el estadístico a observar será $t \sim t(25)$

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{370,69-360}{24,26/\sqrt{26}} = 0,086$$

$\alpha = 0,05 = P(t \leq -t_{0,025;25} \parallel t \geq t_{0,025;25}) \Rightarrow t_{0,025;25} = 2,060$

$\Rightarrow RR: \{t : t \geq 2,060 \parallel t \leq -2,060\}$

$\rightarrow$ Los datos no son suficientes para rechazar H_0.

EJERCICIO 7: La prueba de impacto Charpy de muesca V es la base para estudiar muchos criterios de resistencia de materiales. Se aplicó esta prueba a una muestra de 42 materiales de una aleación especial a 110°F. El promedio muestral de expansión lateral transversal se calculó en 73,1 milésimas de pulgada y la desviación estándar muestral fue 5,9 milésimas de pulgada. El verdadero promedio de cantidad de expansión debe ser menor de 75 milésimas de pulgada para ser adecuada. La aleación no se utilizará ~~a menos~~ a menos que se satisfaga esta condición.

A) Pruebe las hipótesis pertinentes ~~empleada~~ empleando $\alpha=0,01$ para determinar si la aleación es adecuada.

Scanned by CamScanner

PARÁMETROS DE INTERÉS: μ = promedio de expansión real

$H_0: \mu = 75$ contra $H_a: \mu < 75$

$n = 42 \Rightarrow$ Podemos aplicar el TLC $\quad \bar{x} = 73{,}1 \quad S = 5{,}9$

$$\bar{X} \sim N(\mu, S^2/n) \Rightarrow \left[Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{73{,}1 - 75}{5{,}9/\sqrt{42}} = \frac{-1{,}9}{0{,}91} = -2{,}09 \right.$$

$$\alpha = 0{,}01 = P(Z \leq Z_{0{,}01}) = P(Z \leq -2{,}33)$$

$R.R.: \{z \mid z \leq -2{,}33\} \Rightarrow H_0$ no será rechazada

B) Con $\sigma = 5{,}9$ ¿Cuál es la probabilidad de error tipo II $\beta(\mu')$ para $\mu' = 70$?

$$\beta(\mu') = 1 - \Phi\left(-Z_\alpha + \frac{\mu_0 - \mu'}{\sigma\sqrt{n}}\right);$$

$$\beta(70) = 1 - \Phi\left(-2{,}33 + \frac{75-70}{5{,}9/\sqrt{42}}\right) = 1 - \Phi\left(-2{,}33 + \frac{5}{0{,}91}\right) = 1 - \Phi(3{,}16) = 1 - 0{,}992$$

$= 0{,}0008$

EJERCICIO 8: Se determinó la cantidad de desgaste de un eje, después de un recorrido fijo de millas para cada uno de $n = 8$ motores de combustión interna, que llevan cobre y plomo como material antifricción resultado $\bar{x} = 3{,}72$ y $S = 1{,}25$

A) Se supone que la distribución de desgaste del eje es normal con media μ. Utilice una prueba de hipótesis a nivel 0,05 para probar $H_0: \mu = 3{,}50$ contra $H_a: \mu > 3{,}50$

X = Desgaste del eje $\sim N(\mu, \sigma)$

Como el n es pequeño, y la σ es desconocida, nuestro estadístico a observar $\sim t(n-1) = t(7)$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{3{,}72 - 3{,}50}{1{,}25/\sqrt{8}} = \frac{0{,}22}{0{,}44} = 0{,}5$$

$RR: \{t \mid t > t_{0{,}05;7}\} = \{t \mid t > 1{,}895\} \Rightarrow$ No se rechaza H_0

B) Con $\sigma = 1{,}25$ ¿Cuál es la probabilidad de error de tipo II $\beta(\mu')$ para $\mu' = 4$?

$$\beta(4) = \Phi\left(Z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) = \Phi\left(1{,}895 - \frac{3{,}50 - 4}{1{,}25/\sqrt{8}}\right) = \Phi\left(1{,}895 + \frac{0{,}5}{0{,}44}\right) =$$

$\Phi(3{,}03) =$

Scanned by CamScanner

EJERCICIO 9: Una muestra aleatoria de 150 donaciones en un banco de sangre revela que 92 eran de sangre tipo A. ¿Sugiere esto que el porcentaje real de donadores tipo A difiere del porcentaje de la población con sangre tipo A que se considera del 40%? Haga un prueba de las hipótesis adecuadas, con un nivel de significación de 0,01 ¿Sería distinta su conclusión si se usara un nivel de significación de 0,05?

- Parámetro de interés p = proporción de ~~ban~~ donantes tipo A

- $H_0: p = 0{,}4$ y H_a $p \neq 0{,}4$

$X = 92$ (éxitos muestrales)

- Tamaño de muestra adecuado?

$$n \cdot p_0 = 150 \cdot 0{,}4 = 60 > 10 \quad \& \quad n \cdot (1-p_0) = 150 \cdot 0{,}6 = 90 > 10$$

$\Rightarrow$ ESTADÍSTICO

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{X/n - 0{,}4}{\sqrt{0{,}4 \cdot 0{,}6/150}} = \frac{0{,}213}{0{,}04} = 5{,}325$$

→ Con nivel 0,01:

$$\alpha = 0{,}01 = P(z \geq z_{0{,}005} \,||\, z \leq -z_{0{,}005}) = P(z \geq 2{,}575 \,||\, z \leq -2{,}575)$$

Como $z_{obs} = 5{,}325 \geq 2{,}575$ está dentro de la R.R. $\Rightarrow$ H_0 rechazada.

→ Con nivel 0,05:

$$0{,}05 = \alpha = P(z \geq z_{0{,}025} \,||\, z \leq -z_{0{,}025}) = P(z \geq 2{,}81 \,||\, z \leq -2{,}81)$$

$\Rightarrow$ H_0 rechazada también.

EJERCICIO 10: Algunos científicos piensan que los robots jugarán un papel especial en las fábricas en los próximos 20 años. Supongamos que en un experimento para determinar si es factible el uso de robots para trenzar cables de computadora, se empleó un robot para ensamblar 500 cables. Se examinaron los cables y encontraron 5 defectuosos. Si los ensambladores humanos tienen una tasa de 0,03 (3%) ¿Apoya esta información la hipótesis de que la proporción de partes defectuosas es menor para robots que para humanos? Obtenga el valor p y concluya a un nivel de significación de 0,01.

$$\hat{p} = \frac{X}{n} = \frac{5}{500} \; ; \; p_0 = 0{,}03 \Rightarrow H_0 \; ; \; H_a : p_0 < 0{,}03$$

$500 \cdot 0{,}03 > 10$ y $500 \cdot 0{,}7 > 10 \Rightarrow$

ESTADÍSTICO $$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0{,}01 - 0{,}03}{\sqrt{0{,}03(0{,}97)/500}} = \frac{-0{,}02}{0{,}00763} = -2{,}621$$

P VALOR (para cola inferior) $= \Phi(-2{,}621) = 1 - \Phi(2{,}621) = 1 - 0{,}9956 = 0{,}0044$

$$0{,}01 = \alpha = P(z \leq -z_{0{,}01}) = P(z \leq -2{,}33) \Rightarrow R.R: \{z \mid z \leq -2{,}33\}$$

$z_{obs} = -2{,}621 \leq -2{,}33 \Rightarrow H_0$ es rechazada.

Scanned by CamScanner

EJERCICIO 11: En un experimento para probar los efectos de hormonas en el crecimiento de ganado, ~~de~~ se implantaron 200 mg de progesterona y 20 mg de benzoato de estradiol en el oído externo de 16 terneros seleccionados al azar, cada uno con un peso aproximado de 500 lb.
Se encontró que el promedio muestral de peso ganado por día durante cierto número de días fue de 2,72 lb y la desviación estándar muestral de 0,41 lb por día ¿Sugiere esta información que la media verdadera diaria de ganancia de peso para terneros tratados con el implante de hormonas excede de 2,5? Obtenga el valor p y concluya a un nivel de significación de 0,05

$\bar{X} = 2,72$ $\quad$ X = "peso ganado por día". $\quad$ $S = 0,41$

$H_0: \mu = 2,5 \quad H_a: \mu > 2,5$

No dice qué es la distr. de la población pero suponemos que es normal?

$n = 16$

ESTADÍSTICO: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{2,72 - 2,5}{0,41/4} = \dfrac{0,22}{0,1025} = 2,15$

P VALOR: $P(T > t_{obs}) = P(t \geq 2,15) \Rightarrow 0,025 > \text{p-valor} > 0,01$

R.R: $P(t \geq t_{0,025;15}) = P(t \geq 2,131) \Rightarrow H_0$ es rechazada.

EJERCICIO 12: Se proporcionan pares de valores p y niveles de significancia α. Para cada par exprese si el valor observado p llevaría al rechazo de H_0 al nivel de significancia dado

Para rechazar H_0 a nivel $\Rightarrow$ $\boxed{\text{p-valor} \leq \alpha}$

A) valor $p = 0,084$, $\alpha = 0,05$ No rechaza H_0 $\quad$ B) valor $p = 0,003$; $\alpha = 0,001$ No rechaza H_0
C) valor $p = 0,498$, $\alpha = 0,05$ No rechaza H_0 $\quad$ D) valor $p = 0,084$; $\alpha = 0,10$ Rechaza H_0
E) valor $p = 0,039$, $\alpha = 0,01$ No rechaza H_0 $\quad$ F) valor $p = 0,218$; $\alpha = 0,10$ No rechaza H_0.

EJERCICIO 13: Calcular el valor p para una prueba basada en un estadístico z en cada una de las siguientes situaciones y dar una conclusión a un nivel de significación de $\alpha = 0,05$

A) $H_0: \mu = 5$ contra $H_a: \mu > 5$
(I) $Z_{obs} = 1,42$ $\quad$ p valor: $1 - \Phi(1,42) = 1 - 0,9222 = 0,0778$
(II) $Z_{obs} = 2,48$ $\quad$ p valor: $1 - \Phi(2,48) = 1 - 0,9934 = 0,0066$

(B) $H_0: \mu = 5$ contra $H_a: \mu < 5$
(I) $Z_{obs} = -1,96$ $\quad$ P valor: $\Phi(-1,96) = 1 - \Phi(1,96) = 1 - 0,9750 = 0,025$
(II) $Z_{obs} = -0,11$ $\quad$ P valor: $\Phi(-0,11) = 1 - \Phi(0,11) = 1 - 0,5438 = 0,4562$

(C) $H_0: \mu = 5$ contra $H_a: \mu \neq 5$
(I) $Z_{obs} = 2,10$ $\quad$ P valor: $2(1 - \Phi(2,10)) = 2(1 - 0,9821) = 0,0358$
(II) $Z_{obs} = -1,75$ $\quad$ P valor: $2(1 - \Phi(1,75)) = 2(1 - 0,9599) = 0,08$

(A) R.R: $\{z \mid z \geq -z_{0,05}\} = \{z \mid z \geq 1,645\}$ (I) NO RECHAZA (II) RECHAZA

(B) R.R: $\{z \mid z \leq -1,645\}$ (I) RECHAZA (II) NO RECHAZA

(C) R.R.: $\{z \mid z \leq -z_{0,025} \,||\, z \geq z_{0,025}\} = \{z \mid z \leq -1,96 \,||\, z \geq 1,96\}$

(I) RECHAZA (II) NO RECHAZA

Scanned by CamScanner

EJERCICIO 14: Dar el valor p para una prueba basada en un estadístico t en cada una de las siguientes situaciones y dar una conclusión a un nivel de significación de $\alpha = 0{,}02$.

(A) Prueba cola superior, grados de libertad = 8, $t_{obs} = 2$

$\boxed{\text{RECHAZA } H_0 \text{ si } p\text{-valor} \leq \alpha}$

$p\text{ valor} = P(t > t_{obs}) = P(t > 2) \Rightarrow 0{,}025 < p\text{-valor} < 0{,}05$

$\Rightarrow \alpha = 0{,}02 < 0{,}025 \Rightarrow$ NO RECHAZA H_0

(B) Prueba cola inferior, grados de libertad = 11, $t_{obs} = -2{,}4$

$p\text{ valor} = P(t < t_{obs}) = P(t < -2{,}4) = 1 - P(t$

(C) Prueba bilateral, grados de libertad = 15, $t_{obs} = -1{,}6$

$p\text{ valor} = 2P(t > |t_{obs}|) = 2P(t > 1{,}6) \Rightarrow 2(0{,}05 < p\text{-valor} < 0{,}1) = 0{,}1 < p\text{-valor} < 0{,}2$

$\Rightarrow$ NO RECHAZA H_0

Scanned by CamScanner

PRUEBA DE HIPÓTESIS
- EJEMPLOS

EJEMPLO 8.1: Se sabe que un cierto tipo de automóvil no sufre daños visibles el 25% del tiempo en pruebas de choque a 10 mph. Se ha propuesto un diseño de defensa modificado en un esfuerzo por incrementar el porcentaje.

Sea p la proporción de todos los choques a 10 mph con esta nueva defensa que no producen daños visibles. Las hipótesis a ser tratadas son:

$\boxed{H_0: p = 0{,}25}$ (ninguna mejora) $\boxed{H_a: p > 0{,}25}$

La prueba se basará en un experimento que implica $\boxed{n = 20}$

Intuitivamente H_0 será rechazada si un número sustancial de los choques no muestra daños visibles.

PROCEDIMIENTO DE PRUEBA:

ESTADÍSTICO: X = número de choques sin daños

RR: $R_8 = \{8, 9, 10, \dots, 20\}$; es decir, rechazar H_0 si $x \geq 8$, donde x es el valor observado del estadístico.

Esta región de rechazo se llama de COLA SUPERIOR y se compone sólo de grandes valores del estadístico.

Cuando H_0 es verdadera, la distribución de probabilidad de X es binomial con $n=20$ y $p=0{,}25$.

Entonces,

$$\alpha = P(\text{ERROR I}) = P(H_0 \text{ rechazada cuando V})$$

$$= P(X \geq 8 \text{ cuando } X \sim Bin(20; 0{,}25))$$

$$= 1 - B(7; 20, 0{,}25)$$

$$B(7; 20, 0{,}25) = \sum_{x=1}^{7} b(i; 20, 0{,}25) = \sum_{i=1}^{7} \binom{20}{i} p^i (1-p)^{20-i} = \sum_{i=1}^{7} \binom{20}{i} 0{,}25^i (0{,}75)^{20-i}$$

$$= \binom{20}{1} 0{,}25 \cdot 0{,}75^{20} + \binom{20}{2} 0{,}25^2 \cdot 0{,}75^{18} + \binom{20}{3} 0{,}25^3 \cdot 0{,}75^{17} + \binom{20}{4} 0{,}25^4 \cdot 0{,}75^{16}$$

$$+ \binom{20}{5} 0{,}25^5 \cdot 0{,}75^{15} + \binom{20}{6} 0{,}25^6 \cdot 0{,}75^{14} + \binom{20}{7} 0{,}25^7 \cdot 0{,}75^{13}$$

$$= 0{,}016 + 0{,}067 + 0{,}134 + 0{,}189 + 0{,}202 + 0{,}168 + 0{,}112$$

$$\simeq 0{,}888 \; (= 0{,}898)$$

$$= 1 - 0{,}898 = 0{,}102$$

Es decir, cuando H_0 es verdadera, aproximadamente el 10% de todos los experimentos compuestos de 20 choques darán por resultado que H_0 fuera rechazada incorrectamente (error tipo I)

Scanned by CamScanner

En contraste con α, no hay una sola β. En su lugar hay una β diferente por cada p diferente que exceda de 0,25.

Por ejemplo, para $p=0{,}3 \Rightarrow$

$$\beta(0{,}3) = P(\text{ERROR TIPO II para } p=0{,}3)$$

$$= P(H_0 \text{ no rechazada cuando es falsa pq } p=0{,}3)$$

$$= P(X \leq 7 \text{ cuando } X \sim Bin(20, 0{,}3))$$

$$= B(7; 20, 0{,}3) = 0{,}772$$

Cuando p es en realidad 0,3 y no 0,25 (un pequeño alejamiento de H_0), aproximadamente el 77 % de todos los experimentos de este tipo darían por resultado que H_0 fuera incorrectamente rechazada.

EJEMPLO 8.2 Se sabe que el tiempo de secado de un cierto tipo de pintura en condiciones de prueba especificadas está normalmente distribuído con valor medio de 75 min y desviación estándar de 9 min. Algunos químicos propusieron un nuevo aditivo para reducir el tiempo de secado. Se cree que los tiempos de secado con este nuevo aditivo permanecerán normalmente distribuídos con $\sigma = 9$ min.

Debido al gasto asociado con el aditivo:

$$H_0: \mu = 75 \text{ contra } H_a: \mu < 75.$$

$$n = 25. \qquad \bar{X} \sim N(\mu, 9/\sqrt{25} = 1{,}8).$$

Una región de rechazo razonable sería con $c = 70{,}8 \Rightarrow \bar{X} \leq 70{,}8$.

Para calcular α y β:

- ESTANDARIZACIÓN DE $\bar{X}$:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{25}} = \frac{\bar{X} - 75}{1{,}8}$$

- $\alpha = P(\text{ERROR I}) = P(\bar{X} \leq 70{,}8 \text{ cuando } \bar{X} \sim N(75, 1{,}8))$

$$= \Phi\left(\frac{70{,}8 - 75}{1{,}8}\right) = \Phi(-2{,}33) = 0{,}01$$

- para $\mu' = 72$

$$\beta(72) = P(\text{ERROR II para } \mu = 72)$$

$$= P(H_0 \text{ no es rechazada cuando es falsa porque } \mu = 72)$$

$$= P(\bar{X} > 70{,}8 \text{ cuando } \bar{X} \sim N(72; 1{,}8))$$

$$= 1 - \Phi\left(\frac{70{,}8 - 72}{1{,}8}\right) = 1 - \Phi(-0{,}67) = 1 - 0{,}2514 = 0{,}7486$$

$$\beta(70) = 1 - \Phi\left(\frac{70{,}8 - 70}{1{,}8}\right) = 0{,}3300 \; ; \quad \beta(67) = 0{,}0174$$

Scanned by CamScanner

EJEMPLO 8.3 Utilice el mismo experimento y el estadístico de prueba X como en el primer ejemplo, pero ahora la R.R $= \{9, 10, \ldots, 20\}$. Como X sigue teniendo una distribución binomial con parámetros $n=20$ y p,

$$\alpha = P(H_0 \text{ es rechazada cuando } p=0{,}25)$$

$$= P(X \geq 9 \text{ cuando } X \sim bin(20; 0{,}25))$$

$$= 1 - B(8; 20; 0{,}25) = 0{,}041 \quad (\text{antes era } \alpha = 0{,}102)$$

Sin embargo, se pagó un precio por esta reducción:

$$\beta(0{,}3) = P(X \leq 8 \text{ cuando } X \sim bin(20; 0{,}3) =$$

$$= 1 - B(8; 20, 0{,}3) = 0{,}887 \quad (\text{antes era } 0{,}772)$$

EJEMPLO 8.5 Considere la situación en la cual μ es el contenido de nicotina promedio verdadero de los cigarrillos marca B. El objetivo es probar

$H_0: \mu = 1{,}5$ contra $H_a: \mu > 1{,}5$.

Suponga que se sabe que la distribución del contenido es NORMAL, con $\sigma = 0{,}20$

Entonces $\bar{X}$ está normalmente distribuída con valor medio $\mu_{\bar{X}} = \mu$ y $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0{,}20/\sqrt{32}$

$= 0{,}0354 \quad (n = 32)$

Se estandariza $\bar{X}$ suponiendo H_0:

$$Z = \frac{\bar{X} - 1{,}5}{0{,}0354}; \qquad \text{La forma de la R.R es } z \geq c.$$

Determínese ahora c para $\alpha = 0{,}05$:

$$\alpha = P(\text{ERROR I}) = P(Z \geq c \text{ cuando } Z \sim N(0,1))$$

El valor de c debe capturar el área de cola superior 0,05 bajo la curva z:

$$c = Z_{0{,}05} \Rightarrow \Phi(Z_{0{,}05}) = 0{,}95 \Rightarrow$$

$$Z_{0{,}05} = 1{,}645 = c$$

Que $Z \geq 1{,}645$ equivale a $\bar{X} - 1{,}5 \geq (0{,}0354)(1{,}645) \Rightarrow \bar{X} \geq 1{,}56$

entonces β es la probabilidad de que $\bar{X} < 1{,}56$ y puede ser calculada con cualquier valor de $\mu > 1{,}5$.

Scanned by CamScanner

EJEMPLO 8.6 Un fabricante de sistemas rociadores utilizados como protección contra incendios afirma que la temperatura de activación promedio verdadera es de 130°.

Una muestra de $n=9$ sometida a prueba da una temperatura promedio muestral de 131,08°F. Si la distribución de los tiempos de activación es normal con $\sigma = 1,5°F$, ¿contradicen los datos la afirmación del fabricante a un nivel de significación $\alpha = 0,01$?

1. PARÁMETRO μ = temperatura de activación promedio verdadera
2. $H_0: \mu = 130$ $(\mu_0 = 130)$
3. $H_a: \mu \neq 130$ (nos interesa un alejamiento en ambas direcciones)
4. ESTADÍSTICO: NORMAL

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1,5/\sqrt{9}}$$

5. R.R $z \geq Z_{0,005}$ o $z \leq -Z_{0,005}$

$(\alpha/2 = 0,005)$ $Z_{0,005} = 2,58 \Rightarrow RR: \{z: z \geq 2,58 \text{ o } z \leq -2,58\}$.

6.

$$z = \frac{131,08 - 130}{1,5/\sqrt{9}} = \frac{1,08}{0,5} = 2,16$$

7. $z = 2,16$ no queda en la R.R así que H_0 no puede ser rechazada.

EJEMPLO 8.7 Sea μ la vida promedio verdadera de la banda de rodamiento de un cierto tipo de llanta.

Considere poner a prueba $H_0: \mu = 30.000$ contra $H_a: \mu > 30.000$ basado en un tamaño de muestra $n = 16$ de una distribución de población NORMAL con $\sigma = 1500$. Una prueba con $\alpha = 0,01$ requiere $Z_{0,01} = 2,33$. La probabilidad de cometer un error de tipo II cuando $\mu = 31000$ es

$$\beta(31000) = \Phi\left(2,33 + \frac{30.000 - 31.000}{1500/\sqrt{16}}\right)$$

$$= \Phi(-0,34) = 0,3669.$$

Como $Z_{0,1} = 1,28$, el requerimiento de que el nivel de prueba 0,01 también tenga $\beta(31000) = 0,1$

$$n = \left[\frac{1500(2,33 + 1,28)}{30.000 - 31.000}\right]^2 = (-5,42)^2 = 29,32 \simeq 30$$

$n = 30$

Scanned by CamScanner

EJEMPLO 8.9 PESO MÁXIMO DE IZAMIENTO $\sim N(\mu, \sigma)$

¿Sugieren los datos siguientes que el peso máximo de izamiento medio de la población excede de 25?

25,8 36,6 26,3 21,8 27,2

Efectúe una prueba con nivel de significación. de 0,05

1. parámetro μ
2. $H_0: \mu = 25$
3. $H_a: \mu > 25$
4. $t = \dfrac{\bar{x} - 25}{s\sqrt{n}}$ $\quad n = 5 \quad \bar{x} = 27{,}54 \quad s = 5{,}47$
5. Rechazar H_0 si $t \geq t_{\alpha, n-1} = t_{0{,}05;4} = 2{,}132$
6. $t = \dfrac{27{,}54 - 25}{5{,}47/\sqrt{5}} = \dfrac{2{,}54}{2{,}45} = 1{,}04$
7. Como $t = 1{,}04$ no queda dentro de la zona de rechazo, H_0 no puede ser rechazado a un nivel de significación de 0,05.

EJEMPLO 8.10 Se supone que la caída de voltaje promedio verdadera es de, cuando mucho, 2,5 volts.

Un investigador selecciona una muestra de $n=10$ transistores y utiliza los voltajes resultantes para probar

$H_0: \mu = 2{,}5$ contra $H_a: \mu > 2{,}5$ por medio de una prueba t

con nivel de significación de 0,05

Si la desviación estándar de la distribución de voltaje es $\sigma = 0{,}100$

¿Qué tan probable es que H_0 no será rechazado cuando en realidad $\mu = 2{,}6$?

Con $d = \dfrac{|\mu_0 - \mu'|}{\sigma} = \dfrac{|2{,}5 - 2{,}6|}{0{,}1} = 1$

Scanned by CamScanner

EJEMPLO 8.11: Se reportó que 1276 individuos en una muestra de 4115 adultos fueron encontrados obesos (IMC > 30). (2002)
Una encuesta revela que el 20% se consideran a sí mismos obesos (1998)

¿Sugieren los datos más recientes que la proporción verdadera de adultos obesos es más de 1,5 veces el porcentaje de la encuesta?
Realice una prueba de hipótesis de un nivel de significación de 0,10.

① PARÁMETRO: p = proporción de obesos

② Decir que el porcentaje actual es 1,5 veces el porcentaje reportado equivale a decir que es el 30%.

$H_0: p = 0{,}30$

③ La frase "más que" implica que $H_a: p > 0{,}30$

④ Como $n p_0 = 4115.(0{,}3) > 10$ y $n.q_0 = 4115(0{,}7) > 10$, ciertamente la muestra es grande.

El estadístico es entonces: $z = \dfrac{(\hat{p} - 0{,}3)}{\sqrt{\dfrac{(0{,}3)(0{,}7)}{4115}}} = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 \cdot q_0}{n}}}$

⑤ H_a implica una prueba de cola superior: R.R $z > z_{0{,}10} = 1{,}28$

⑥ $\hat{p} = 1276/4115 = 0{,}310 \Rightarrow z = (0{,}310 - 0{,}3) / \sqrt{(0{,}3)(0{,}7)/4115}$
$= 0{,}010 / 0{,}0071 = 1{,}40$

⑦ Como $z = 1{,}40$ excede el valor crítico 1,28, se justifica el rechazo de H_0.

EJEMPLO 8.12 Un servicio de mensajería anuncia que por lo menos el 90% de todos los paquetes llevados a su oficina a las 9 A.M son entregados ~~antes~~ alrededor de las 12.

Sea p la proporción verdadera de dichos paquetes entregados como se anuncia, y considere las hipótesis $H_0: p = 0{,}9$ contra $H_a: p < 0{,}9$.

Si sólo el 80% de los paquetes son entregados así, ¿qué tan probable es que una prueba de nivel 0,01 basada en $n = 225$ paquetes detectará tal alejamiento de H_0?

¿Cuál debe ser el tamaño de muestra para garantizar que $\beta(0{,}8) = 0{,}01$?

Con $\alpha = 0{,}01$, $p_0 = 0{,}9$, $p' = 0{,}8$ y $n = 225$ $\quad z_{0{,}01} = 2{,}33$

$$\beta(0{,}8) = 1 - \Phi\left(\frac{p_0 - p' - 2{,}33\sqrt{p_0 \cdot q_0 / n}}{\sqrt{p' \cdot q' / n}}\right)$$

$$= 1 - \Phi\left(\frac{0{,}9 - 0{,}8 - 2{,}33\sqrt{0{,}9 \cdot 0{,}1 / 225}}{\sqrt{0{,}8 \cdot 0{,}2 / 225}}\right) = 1 - \Phi(2) = 0{,}0228$$

Entonces, la probabilidad de que H_0 sea rechazada cuando sí debe serlo para $p = 0{,}8$ es

$1 - \beta(0{,}8) = 0{,}9772 \Rightarrow$ Aproximadamente el 98% de las pruebas darán por resultado el rechazo correcto de H_0 cuando p es 0,8.

$$n = \left[\frac{z_\alpha \sqrt{0{,}9 \cdot 0{,}1} + z_\beta \sqrt{0{,}8 \cdot 0{,}2}}{\ldots}\right]^2 \approx 266$$

Scanned by CamScanner

EJEMPLO 8.13

Un fabricante de plástico desarrolló un nuevo tipo de bote de plástico para la basura y propone venderlo con una garantía incodicional de seis años. Para ver si esto es económicamente factible, 20 botes prototipo se someten a una prueba acelerada de duración para simular seis años de uso. La garantía propuesta se modificará sólo si los datos muestrales sugieren fuertemente que menos de 90% sobrevivirán el período de seis años.

Sea p la proporción de todos los botes que sobreviven la prueba acelerada. La hipótesis pertinentes son

$\boxed{H_0: p = 0{,}9}$; $\boxed{H_a: p < 0{,}9}$

ESTADÍSTICOS: X = botes que sobreviven.

Si el nivel de significación es $\alpha = 0{,}05$, c debe satisfacer $B(c; 20; 0{,}9) \leq 0{,}05$. De acuerdo con la tabla, $B(15; 20, 0{,}9) = 0{,}043$, mientras que $B(16; 20; 0{,}9) \leq 0{,}133$. La región de rechazo es, por lo tanto, $\boxed{X \leq 15}$.

La probabilidad de un error de tipo II con el valor alternativo $p' = 0{,}8$ es

$$\beta(0{,}8) = P(H_0 \text{ no rechazada cuando } X \sim bin(20, 0{,}8))$$
$$= P(X \geq 16 \text{ cuando } X \sim Bin(20, 0{,}8))$$
$$= 1 - B(15; 20, 0{,}8) = 1 - 0{,}370 = 0{,}630.$$

Es decir, cuando $p = 0{,}8$, el 63% de todas las muestras compuestas de $n = 20$ botes darían por resultado que H_0 no sea incorrectamente rechazada. Esta probabilidad de error es alta porque $n = 20$ es una muestra pequeña, y $p' = 0{,}8$ se acerca al valor nulo $p_0 = 0{,}9$.

EJEMPLO 8.14

Se sabe que el tiempo promedio verdadero para el alivio inicial del dolor de un analgésico conocido es de 10 min. Sea μ el tiempo promedio verdadero para el alivio del dolor de un analgésico recién desarrollado. Éste será comercializado sólo si proporciona un alivio más rápido que el conocido, así que se debe probar que

$H_0: \mu = 10$ $\quad H_a: \mu < 10$

pero el resultado y la conclusión posterior dependerá del nivel de significancia

EJEMPLO 8.17

El espesor objetivo de obleas de silicio utilizadas en cierto tipo de circuito integrado es de 245 μm. De una muestra de 50 obleas, cada una con un espesor determinado, se obtiene una media de espesor de 246,18 μm y una $\sigma = 3{,}60\,\mu m$. ¿Sugieren estos datos que el espesor de oblea promedio verdadero es algún otro diferente del valor objetivo?

1. PARÁMETRO: μ = espesor de oblea promedio verdadero
2. HIPÓTESIS NULA: $H_0: \mu = 245$ contra $H_a: \mu \neq 245$
3. ESTADÍSTICO:

4 y 5 $\quad z = \dfrac{\bar{x} - 245}{s/\sqrt{n}} = \dfrac{246{,}18 - 245}{3{,}60/\sqrt{50}} = 2{,}32$

6. DETERMINACIÓN DEL VALOR P: $\quad$ valor $P = 2(1 - \Phi(2{,}32)) = 0{,}0204$

7. CONCLUSIÓN: Con un nivel de significación de 0,01, H_0 no sería rechazada puesto que $0{,}0204 > 0{,}01$

Scanned by CamScanner

VARIANZA DE X

Sea $p(x)$ la función masa de probabilidad de X y μ su valor esperado. En ese caso la VARIANZA de X, $V(X)$ o σ_X^2 o σ^2 es

$$V(X) = \sum_D (x-\mu)^2 \cdot p(x) = E[(X-\mu)^2].$$

→ FORMULA ABREVIADA PARA σ^2

$$V(X) = \sigma^2 = \left[\sum_D x^2 \cdot p(x)\right] - \mu^2 = E(X^2) - [E(X)]^2$$

→ REGLAS DE VARIANZA

$$V(aX+b) = \sigma^2_{aX+b} = a^2 \cdot \sigma_X^2 \quad y \quad \sigma_{aX+b} = |a| \cdot \sigma_X$$

En particular, $\sigma_{aX} = |a| \cdot \sigma_X$ y $\sigma_{X+b} = \sigma_X$

DISTRIBUCIÓN DE PROBABILIDAD BINOMIAL

VARIABLE ALEATORIA BINOMIAL Y DISTRIBUCIÓN

X = el número de E entre los n ensayos.

Como la función masa de probabilidad de una variable aleatoria binomial X depende de los dos parámetros n y p, la función masa de probabilidad se denota por $b(x; n, p)$.

$$\boxed{b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots n \\ 0 & \text{de lo contrario} \end{cases}}$$

Para $X \sim Bin(n, p)$, la función de distribución acumulativa será denotada por

$$P(X \leq x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p) \qquad x = 0, 1, \dots, n$$

LA MEDIA Y VARIANZA DE X

Si $X \sim Bin(n, p)$, entonces

$E(X) = np$ y $V(X) = np(1-p) = npq$ y $\sigma_X = \sqrt{npq}$

(donde $q = 1-p$)

Scanned by CamScanner

UNIDAD VI: DISTRIBUCIÓN DE MUESTREO & ESTIMACIÓN PUNTUAL

Observación:

Antes de que los datos estén disponibles, cada observación se considera como una variable aleatoria $(X_1, X_2, \dots, X_n)$.

El valor de la MEDIA de cualquier muestra puede ser considerada como ESTIMACIÓN PUNTUAL de la media de la POBLACIÓN μ.
(y así con el resto de las mediciones posibles sobre la muestra). En resumen, los valores de cada una de las observaciones muestrales siempre varían de una muestra a la otra, así que, en general, el valor de una característica muestral utilizado como estimación casi nunca coincidirá con lo que está siendo estimado.

DEF:

Un ESTADÍSTICO es cualquier cantidad cuyo valor puede ser calculado a partir de datos muestrales. Es una VARIABLE ALEATORIA antes de obtener los datos del muestreo (letra mayúscula); se utiliza letra minúscula para denotarlo una vez calculado.

OBS: A menudo, estadísticos del tipo $\bar{Y} - \bar{X}$ (diferencia entre las medias muestrales de dos muestras diferentes).

Cualquier estadístico, por el hecho de ser una variable aleatoria, tiene una DISTRIBUCIÓN DE PROBABILIDAD. =
La distribución de probabilidad de un estadístico se conoce en ocasiones como DISTRIBUCIÓN DE MUESTREO para enfatizar que describe como varía el valor del estadístico a través de todas las muestras que pudieran ser seleccionadas.

MUESTRAS ALEATORIAS

La d.d.p de cualquier estadístico particular depende no sólo de la de la población y el tamaño de la muestra n sino también del MÉTODO DE MUESTREO:

DEF:

Se dice que las variables aleatorias $X_1, X_2, \dots, X_n$ forman una MUESTRA ALEATORIA simple de tamaño n si

1. Las X_i son variables aleatorias independientes.
2. Cada X_i tiene la misma distribución de probabilidad.

Scanned by CamScanner

Existen dos métodos generales de obtener información sobre una distribución de muestreo estadístico. Uno implica cálculos basados en reglas de probabilidad y el otro implica realizar un experimento de simulación.

DERIVACIÓN DE UNA DISTRIBUCIÓN DE MUESTREO.

Se pueden utilizar reglas de probabilidad para obtener la distribución de un estadístico siempre que sea una función "simple" de las x_i y, o existen relativamente pocos valores x diferentes en la población, o bien la distribución de la población tiene una forma "accesible".

EJEMPLO DISCRETO

utilizando un ejemplo simple de un negocio de afinación de autos, llegamos a que

$$\mu_{\bar{x}} = E(\bar{x}) = \sum \bar{x} \cdot p_{\bar{x}}(\bar{x}) = \dots = \mu \quad \text{y} \quad \sigma^2_{\bar{x}} = V(\bar{x}) = \sum \bar{x}^2 \cdot p_{\bar{x}}(\bar{x}) - \mu^2_{\bar{x}} = \dots = \frac{\sigma^2}{2}$$

(en este caso la varianza de $\bar{x}$ es precisamente la mitad de la varianza original porque $n=2$)

EJEMPLO CONTINUO: Tiempo de servicio para un tipo de transacción bancaria. Distribución exponencial y parámetro λ.

(...)

EXPERIMENTOS DE SIMULACIÓN

Este método casi siempre se utiliza cuando la obtención vía reglas de probabilidad es demasiado difícil o complicado de realizar. Se realiza con ayuda de una computadora.

Scanned by CamScanner

DISTRIBUCIÓN DE LA MEDIA MUESTRAL

La importancia de la media muestral $\bar{X}$ proviene de su uso al sacar conclusiones sobre la media de la población μ.

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con valor medio μ y desviación estándar σ. Entonces

1. $E(\bar{X}) = \mu_{\bar{X}} = \mu$
2. $V(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2/n$ y $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

Además, con $T_0 = X_1 + \cdots + X_n$ (el total de la muestra),
$E(T_0) = n\mu$, $V(T_0) = n\sigma^2$ y $\sigma_{T_0} = \sqrt{n}\,\sigma$

demostración de esto más adelante.

EL CASO DE UNA DISTRIBUCIÓN DE POBLACIÓN NORMAL

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de distribución NORMAL con media μ y desviación estándar σ. Entonces, con cualquier n, $\bar{X}$ está normalmente distribuída (con media μ y desviación estándar $\sigma/\sqrt{n}$), al igual que T_0 (con media $n\mu$ y desviación estándar $\sqrt{n}\,\sigma$).

TEOREMA DEL LÍMITE CENTRAL

Incluso cuando la distribución de la población es altamente no normal, el cálculo de promedios produce una distribución más acampanada que la que está siendo muestreada.

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces si n es suficientemente grande, $\bar{X}$ tiene aproximadamente una distribución normal con $\mu_{\bar{X}} = \mu$ y $\sigma^2_{\bar{X}} = \sigma^2/n$, y T_0 también tiene aproximadamente una distribución normal con $\mu_{T_0} = n.\mu$ y $\sigma^2_{T_0} = n\sigma^2$.
Mientras más grande es n, mejor es la aproximación.

Scanned by CamScanner

REGLA EMPÍRICA

Si $n > 30$, se puede utilizar TLC

OTRAS APLICACIONES DEL TLC:

Puede ser utilizado para justificar la aproximación normal a la distribución binomial. Recuérdese que una variable binomial X es el nro de éxitos en una experiencia binomial compuesta de n ensayos independientes con éxitos/fallas, y $p = P(S)$ para cualquier ensayo particular. Defínanse nuevas V.A. $X_1, X_2, \dots, X_n$ como $X_i = \begin{cases} 1 & \text{éxito} \\ 0 & \text{fallo} \end{cases}$.

Como los ensayos son independientes y $P(S)$ es una constante de un experimento al otro, las X_i son idénticas (una muestra aleatoria de una distribución de BERNOULLI).

El TLC implica entonces que, si n es suf. grande, tanto la suma como el promedio de las X_i tienen distribuciones normales de manera aprox.

> Sean $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución para la cual sólo valores positivos son posibles $[P(X_i > 0) = 1]$. Entonces si n es suficientemente grande, el producto $Y = X_1 \cdot X_2 \cdot \dots \cdot X_n$ tiene aproximadamente una distribución lognormal

DISTRIBUCIÓN DE UNA COMBINACIÓN LINEAL

La media muestral $\bar{X}$ y el total muestral T_0 son casos especiales de un tipo de variables aleatorias que surgen con frecuencia en aplicaciones estadísticas.

> Dado un conjunto de n variables aleatorias $X_1, \dots, X_n$ y n constantes numéricas $a_1, \dots, a_n$, la variable aleatoria.
>
> $$Y = a_1 X_1 + \dots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$
>
> Se llama COMBINACIÓN LINEAL de las X_i

Con $a_1 = a_2 = \dots = a_n = 1$ se obtiene $Y = X_1 + \dots + X_n = T_0$

y con $a_1 = a_2 = \dots = a_n = \frac{1}{n}$ se obtiene $Y = \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n} T_0 = \bar{X}$.

Obsérvese que no se requiere que las X_i sean independientes o que estén idénticamente distribuidas

Scanned by CamScanner

VALOR MEDIO Y VARIANZA ESPERADOS DE UNA COMBINACIÓN LINEAL.

Si $X_1, X_2, \dots, X_n$ tienen valores medios $\mu_1, \dots, \mu_n$ y varianzas $\sigma_1^2, \dots, \sigma_n^2$:

① Sean las x_i independientes o no...

$$\boxed{E(a_1X_1 + a_2X_2 + \dots + a_nX_n)} = E(a_1X_1) + E(a_2X_2) + \dots + E(a_nX_n)$$
$$= a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n)$$
$$= \boxed{a_1\mu_1 + \dots + a_n\mu_n}$$

② Si $X_1, X_2, \dots, X_n$ son INDEPENDIENTES...

$$\boxed{V(a_1X_1 + \dots + a_nX_n)} = a_1^2V(X_1) + \dots + a_n^2V(X_n) = \boxed{a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2}$$

y

$$\boxed{\sigma_{a_1X_1 + \dots + a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2}}$$

③ Con cualquier $X_1, X_2, \dots, X_n$ ...

$$V(a_1X_1 + \dots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \, Cov(X_i, X_j)$$

DIFERENCIA ENTRE DOS VARIABLES ALEATORIAS

Un caso especial importante de una combinación lineal se presenta con $n=2$, $a_1=1$ y $a_2=-1$:

$$Y = a_1X_1 + a_2X_2 = X_1 - X_2$$

Entonces

$\boxed{E(X_1 - X_2) = E(X_1) - E(X_2)}$, y, si X_1 y X_2 son independientes,
$\boxed{V(X_1 - X_2) = V(X_1) + V(X_2)}$

(Existe tanta variabilidad en $X_1 - X_2$ como en $X_1 + X_2$)

Scanned by CamScanner

EL CASO DE VARIABLES ALEATORIAS NORMALES

Si $X_1, \dots, X_n$ son variables aleatorias independientes normalmente distribuidas (con quizás diferentes medias y varianzas), entonces cualquier combinación lineal de las X_i tiene también una distribución normal.
En particular, la diferencia $X_1 - X_2$ también está distribuida normalmente.

El TLC también puede ser generalizado para aplicarlo a ciertas combinaciones lineales. En general, si n es grande y no es probable que algún término individual contribuya demasiado al valor total, entonces Y tiene una distribución normal.

Scanned by CamScanner

CAPÍTULO VI: ESTIMACIÓN PUNTUAL

Cuando se discuten los métodos y conceptos generales de INFERENCIA, es conveniente disponer de un símbolo genérico para el parámetro de interés. Se utilizará θ.

El objetivo de la estimación puntual es seleccionar un sólo número, con base en los datos muestrales, que represente un valor sensible de θ.

> Una ESTIMACIÓN PUNTUAL de un parámetro θ es un número único que puede ser considerado como un valor sensible de θ. Se obtiene una estimación puntual seleccionando un estadístico apropiado y calculando su valor con los datos muestrales dados.
>
> El estadístico seleccionado se llama <u>estimador puntual de θ</u>.

El símbolo $\hat{\theta}$ se utiliza comunmente para denotar tanto la estimación de θ como la estimación puntual que resulta de una muestra dada.

ERROR CUADRÁTICO MEDIO

medida general de precisión — $EMC = E[(\hat{\theta} - \theta)^2]$

Al no conocer el valor exacto de θ por lo general no hay método exacto para determinar el estimador con error mínimo.

ESTIMADORES INSESGADOS

> Se dice que un estimador puntual $\hat{\theta}$ es un ESTIMADOR INSESGADO de θ si $E(\hat{\theta}) = \theta$ con todo valor posible de θ. Si $\hat{\theta}$ no es insesgado, la diferencia $E(\hat{\theta}) - \theta$ se conoce como el SESGO de $\hat{\theta}$.

$\hat{\theta}$ es insesgado si su distribución de probabilidad siempre está "centrada" en el valor verdadero del parámetro.

Scanned by CamScanner

$p \longrightarrow$

Cuando X es una variable aleatoria binomial con parámetros n y p, la proporción muestral $\hat{p} = X/n$ es un estimador insesgado de p

$$E(\hat{p}) = \left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(n.p) = p$$

σ y $\sigma^2 \longrightarrow$

Sean $X_1, \dots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces el estimador

$$\hat{\sigma}^2 = S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1}$$

es un estimador INSESGADO de σ^2

Aún cuando S^2 es insesgado para σ^2, S es un estimador sesgado para σ (aunque con sesgo pequeño para n no pequeña).

$\mu \longrightarrow$

No hay un único estimador insesgado para μ.

Si $X_1, \dots, X_n$ es una variable aleatoria tomada de una distribución con media μ, entonces $\bar{X}$ es un estimador insesgado de μ. Si además la distribución es <u>continua</u> y <u>simétrica</u>, entonces $\tilde{X}$ y cualquier media recortada también son estimadores insesgados de μ.

Scanned by CamScanner

Criterios para elegir entre estimadores insesgados

ESTIMADORES CON VARIANZA MÍNIMA

Entre todos los estimadores de θ insesgados, se selecciona el de varianza mínima. El $\hat{\theta}$ resultante se llama ESTIMADOR INSESGADO CON VARIANZA MÍNIMA (EIVM) de θ.

Metodología para identificar el EIVM en una amplia variedad de situaciones:

Sean $X_1, \dots, X_n$ una muestra aleatoria tomada de una distribución NORMAL con parámetros μ y σ^2. Entonces el estimador $\hat{\mu} = \bar{X}$ es el estimador insesgado con varianza mínima (EIVM) para μ.

Siempre que exista la seguridad de que la población que se está muestreando es normal, el resultado dice que $\bar{X}$ debe usarse para estimar μ.

Algunas complicaciones:

(...) ver ejemplos

REPORTE DE UNA ESTIMACIÓN PUNTUAL: EL ERROR ESTÁNDAR.

El ERROR ESTÁNDAR de un estimador $\hat{\theta}$ es su desviación estándar $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. Si el error estándar implica parámetros desconocidos cuyos valores pueden ser estimados, la sustitución de estas ~~estimaciones~~ estimaciones en $\sigma_{\hat{\theta}}$ da el ERROR ESTÁNDAR ESTIMADO (desviación estándar estimada) del estimador. Este último puede ser denotado $\hat{\sigma}_{\hat{\theta}}$ o por $s_{\hat{\theta}}$.

Cuando la distribución de $\hat{\theta}$ es aproximadamente normal, el valor verdadero de θ queda dentro de 2 errores estándar (aprox.) de $\hat{\theta}$.
Si $\hat{\theta}$ no es necesariamente normal pero es insesgado, se espera que el valor verdadero quede dentro de 4 errores estándar de $\hat{\theta}$.

Scanned by CamScanner

MÉTODOS DE ESTIMACIÓN PUNTUAL

Dos métodos "constructivos": la definición general de cada tipo de estimador sugiere explícitamente cómo obtener el estimador en cualquier problema específico. Aún cuando se prefieren los estimadores de máxima verosimilitud a los de momentos, debido a ciertas propiedades de eficiencia, a menudo requieren significativamente más cálculo que los estimadores de momento. En ocasiones es el caso de que estos métodos dan estimadores insesgados.

EL MÉTODO DE MOMENTOS

Poder igualar ciertas características muestrales a los valores esperados de la población correspondiente. Luego resolviendo estas ecuaciones con valores de parámetros conocidos se obtienen los estimadores.

Si $X_1, \dots, X_n$ constituyen una muestra aleatoria proveniente de una función masa de probabilidad o de una función de densidad de probabilidad $f(x)$. Con $k = 1, 2, 3 \dots$ el k-ÉSIMO MOMENTO DE LA POBLACIÓN o el k-ÉSIMO MOMENTO DE LA DISTRIBUCIÓN $f(x)$ es $E(X^k)$.

El k-ÉSIMO MOMENTO MUESTRAL es $(1/n)\sum_{i=1}^{n} X_i^k$

Por ende, el primer momento de la población es $E(X) = \mu$, y el primer momento muestral es $\sum X_i / n = \bar{X}$... los momentos de la población serán funciones de cualquier parámetro desconocido $\theta_1, \theta_2, \dots$

Si $X_1, \dots, X_n$ son una muestra aleatoria de una distribución con función masa de probabilidad o función de densidad de probabilidad $f(x; \theta_1, \dots, \theta_m)$ donde $\theta_1, \dots, \theta_m$ son parámetros cuyos valores son desconocidos. Entonces los <u>estimadores de momentos</u> $\hat{\theta}_1, \dots, \hat{\theta}_m$ se obtienen igualando los primeros m momentos muestrales con los primeros m momentos de la población correspondientes, y resolviendo para $\theta_1, \dots, \theta_m$.

Scanned by CamScanner

ESTIMACIÓN DE MÁXIMA VEROSIMILITUD

Que $X_1, \dots, X_n$ tengan una función masa de probabilidad o una función de densidad de probabilidad $f(x_1, x_2, \dots, x_n; \theta_1, \dots, \theta_m)$. Cuando $x_1, \dots, x_n$ son los valores muestrales observados y la función f se considera una función de $\theta_1, \dots, \theta_m$ se llama FUNCIÓN DE VEROSIMILITUD.

Las estimaciones de Máxima Verosimilitud (EMV) $\hat{\theta}_1, \dots, \hat{\theta}_m$ son aquellos valores de la θ_i que incrementan al máximo la función de probabilidad, de modo que

$$f(x_1, \dots, x_n; \hat{\theta}_1, \dots, \hat{\theta}_m) \geq f(x_1, \dots, x_n; \theta_1, \dots, \theta_m) \text{ con TODOS los } \theta_i.$$

Cuando se sustituyen las X_i en lugar de las x_i, se obtienen los estimadores de máxima verosimilitud.

La función de verosimilitud dice qué tan probable es que la muestra observada sea una función de los posibles valores del parámetro. Al incrementarse al máximo la probabilidad se obtienen los valores de parámetros con los que la muestra observada es más probable que haya sido generada, es decir, los valores de parámetro que "más concuerdan" con los datos observados.

ESTIMACIÓN DE FUNCIONES DE PARÁMETROS

→ PRINCIPIO DE INVARIANZA

Sean $\hat{\theta}_1, \dots, \hat{\theta}_m$ los estimadores de máxima verosimilitud de los parámetros $\theta_1, \dots, \theta_m$. Entonces el estimador de máxima verosimilitud de cualquier función $h(\theta_1, \dots, \theta_m)$ de estos parámetros es igual a la función $h(\hat{\theta}_1, \dots, \hat{\theta}_m)$ de los estimadores de máxima verosimilitud.

Scanned by CamScanner

COMPORTAMIENTO CON MUESTRA GRANDE DEL ESTIMADOR DE MÁXIMA VEROSIMILITUD

La sig. prop. proporciona razones adicionales fundamentales para el uso de estimadores de máxima verosimilitud.

En condiciones muy generales en relación con la distribución conjunta de la muestra, cuando el tamaño de la muestra, n, es grande, el EMV de cualquier parámetro θ es aproximadamente INSESGADO ($E(\hat{\theta}) \approx \theta$) y su varianza es casi tan pequeña como la que puede ser lograda por cualquier estimador. Expresado de otra manera, el EMV $\hat{\theta}$ es aproximadamente el estimador insesgado con varianza mínima de θ.

<u>OBS</u>! La obtención de un EMV requiere que se especifique la distribución subyacente.

Scanned by CamScanner

CAP 3

DISTRIBUCIONES DE PROBABILIDAD PARA VARIABLES ALEATORIAS DISCRETAS

PARÁMETRO DE UNA DISTRIBUCIÓN DE PROBABILIDAD

Supóngase que $p(x)$ depende de la cantidad que puede ser asignada a cualesquiera de varios valores posibles y cada valor determina una distribución de probabilidad diferente. Tal cantidad es el PARÁMETRO de la distribución.

FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA

Para dos números cualesquiera a y b con $a \leq b$.

$P(a \leq x \leq b) = F(b) - F(a-)$ donde "$a-$" representa el valor posible de X más grande que es estrictamente menor que a.

En particular, si los únicos valores posibles son enteros, y si a y b son enteros, entonces

$$P(a \leq x \leq b) = P(X = a \text{ o } a+1 \text{ o} \ldots \text{o } b) = F(b) - F(a-1)$$

Con $a = b$ se obtiene $P(X=a) = F(a) - F(a-1)$ en este caso.

VALORES ESPERADOS

Sea X una variable aleatoria discreta, con un conjunto de valores posibles D y una función masa de probabilidad $p(x)$. El VALOR ESPERADO o VALOR MEDIO de X, denotado por $E(X)$ o μ_x, es

$$\boxed{E(X) = \mu_x = \sum_{x \in D} x \cdot p(x)}$$

VALOR ESPERADO DE UNA FUNCIÓN

El valor esperado de cualquier función $h(x)$, denotado por $E[h(x)]$ o $\mu_{h(x)}$, se calcula con

$$\boxed{E[h(x)] = \sum_{D} h(x) \cdot p(x).}$$

REGLAS DEL VALOR ESPERADO

La función de interés $h(x)$ con bastante frecuencia es una función lineal $aX + b$. En este caso, calcular $E[h(x)]$ es fácil de calcular a partir de $E(x)$.

$$\boxed{E(aX + b) = a \cdot E(X) + b}$$

30
10,50
2898
000
30 ARS → 2898
1 → 96

Scanned by CamScanner

DISTRIBUCIONES HIPERGEOMÉTRICAS Y BINOMIALES NEGATIVAS

DISTRIBUCIÓN HIPERGEOMÉTRICA

Las suposiciones que conducen a una distribución hipergeométrica son las siguientes:

1. La población o conjunto que se va a muestrear se compone de N individuos, objetos o elementos (población finita)
2. Cada individuo puede ser caracterizado como ÉXITO (E) o FALLA (F), y hay M éxitos en la población.
3. Se selecciona una muestra de n individuos sin reemplazo de modo que cada subconjunto de tamaño n es igualmente probable de ser seleccionado.

La variable aleatoria de interés es X = el nro de ÉXITOS de la muestra.

EJEMPLO: Se reciben 20 solicitudes de servicio de problemas con impresoras, de las cuales 8 eran impresoras láser y 12 eran de inyección de tinta. Se selecciona una muestra de 5 solicitudes completamente al azar ¿Cuál es la probabilidad de que x (x=0, 1, ... ó 5) de las solicitudes de servicio fueran para impresoras de inyección de tinta?

RTA:

El tamaño de la población es N=20, el tamaño de la muestra es n=5 y el número de inyección de tinta E es M=12 y N-M=8 (F). Considérese el valor x=2. Como todos los resultados son igualmente probables:

$$P(X=2) = h(2;5,12,20) = \frac{\text{número de resultados con } x=2}{\text{número de posibles resultados}}$$

$$h(2;5,12,20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}}$$

Si X es el número de éxitos (E) en una muestra completamente aleatoria de tamaño n extraída de la población compuesta de M éxitos y (N-M) fallas, entonces la distribución de probabilidad de X llamada DISTRIBUCIÓN HIPERGEOMÉTRICA es

$$P(X=x) = h(x;n,M,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

MEDIA Y VARIANZA

$$\boxed{E(X) = n \cdot \frac{M}{N}} \qquad \boxed{V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)}$$

Scanned by CamScanner

DISTRIBUCIÓN BINOMIAL NEGATIVA

El experimento continúa hasta que un total de r éxitos hayan sido observados, donde r es un entero positivo especificado.

La variable aleatoria de interés es X= el número de fallas que preceden al r-ésimo éxito.

En contraste con la variable aleatoria binomial, el número de éxitos es fijo y el número de ensayos es aleatorio.

La función masa de probabilidad de la variable aleatoria binomial negativa X con los parámetros r= número de éxitos (E) y P(E) = p es

$$\boxed{nb(x;r,p) = \binom{x+r-1}{r-1} p^r (1-p)^x}$$

En el caso especial r=1,

$$\boxed{nb(x;1,p) = (1-p)^x \cdot p}$$ ← DISTRIBUCIÓN GEOMÉTRICA

Si X es una variable aleatoria binomial negativa con función masa de probabilidad nb(x; r, p) entonces

$$\boxed{E(X) = \frac{r(1-p)}{p}} \quad y \quad \boxed{V(X) = \frac{r(1-p)}{p^2}}$$

Scanned by CamScanner

DISTRIBUCIÓN DE PROBABILIDAD DE POISSON

Contrario a las anteriores, no existe un experimento simple en el cual esté basada la distribución de Poisson:

Se dice que una variable aleatoria X tiene una distribución de Poisson con parámetro λ ($\lambda > 0$) si la función masa de probabilidad de X es

$$P(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad \text{con} \quad x = 0, 1, 2, \ldots$$

El valor de λ es con frecuencia un valor por unidad de TIEMPO o por unidad de ~~área~~ ÁREA.

La letra e es el nro e. Como λ debe ser positiva, $p(x;\lambda) > 0$ con todos los valores posibles x.

EJEMPLO: Sea X el número de criaturas de un tipo particular capturadas en una trampa durante un período determinado. Suponga que X tiene una distribución de Poisson con $\lambda = 4{,}5$, así que en promedio las trampas contendrán 4,5 criaturas. Entonces la probabilidad de que una trampa contenga exactamente 5 criaturas es

$$P(X=5) = P(5;4{,}5) = \frac{e^{-4{,}5}\cdot(4{,}5)^{5}}{5!} = 0{,}1708.$$

La probabilidad de que contenga a lo sumo 5 criaturas es

$$P(X \leq 5) = \sum_{x=0}^{5} \frac{e^{-4{,}5}\cdot(4{,}5)^{x}}{x!} = e^{-4{,}5}\cdot\overbrace{\left[1 + 4{,}5 + \frac{(4{,}5)^{2}}{2!} + \cdots + \frac{(4{,}5)^{5}}{5!}\right]}^{\text{SERIE DE MCLAURIN}} = 0{,}7029.$$

MEDIA Y VARIANZA

Como $b(x;n,p) \to p(x;\lambda)$ a medida que $n \to \infty$, $p \to 0$, $np \to \lambda$, la media y varianza de una variable binomial deberán aproximarse a las de una de Poisson. Estos límites son $np \to \lambda$ y $np(1-p) \to \lambda$:

Si X tiene una distribución de Poisson con parámetro λ, entonces

$$\boxed{E(X) = V(X) = \lambda}$$

Scanned by CamScanner

PROCESO DE POISSON

Una aplicación muy importante de la distribución de Poisson surge en conexión con la ocurrencia de eventos de algún tipo en el transcurso del tiempo.

1. Existe un parámetro $\alpha > 0$ de tal modo que durante cualquier intervalo de tiempo corto Δt, la probabilidad de que ocurra exactamente 1 evento es $\alpha \cdot \Delta t + o(\Delta t)$.

(...) muy difícil, paja.

Scanned by CamScanner

CAPÍTULO 4: VARIABLES ALEATORIAS CONTINUAS Y DISTRIBUCIONES DE PROBABILIDAD

Sea X una variable aleatoria continua. Entonces, una distribución de probabilidad o función de densidad de probabilidad (fdp) de X es una función f(x) tal que para dos números cualesquiera a y b con $a \leq b$,

$$\boxed{P(a \leq X \leq b) = \int_a^b f(x).dx}$$

. (ÁREA SOBRE ESTE INTERVALO Y BAJO LA CURVA DE DENSIDAD f(x))

Para que f(x) sea una función de densidad de probabilidad legítima, debe satisfacer las dos siguientes condiciones:

1. $f(x) \geq 0$ con todas las x
2. $\int_{-\infty}^{\infty} f(x)\,dx$ = área bajo la curva $f(x) = 1$

DISTRIBUCIÓN UNIFORME

Se dice que una variable aleatoria continua X tiene una distribución uniforme en el intervalo [A, B] si la función de densidad de probabilidad de X es

$$\boxed{f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{de lo contrario} \end{cases}}$$

FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,d(y)$$

Sea X una variable aleatoria continua con función de densidad de probabilidad f(x) y función de distribución acumulativa F(x). Entonces con cualquier número a,

$P(X > a) = 1 - F(a)$, y para dos números cualesquiera a y b con $a < b$,

$P(a \leq X \leq b) = F(b) - F(a)$

OBTENCIÓN DE f(x) A PARTIR DE F(x)

El análogo continuo de una diferencia es una derivada.

Si X es una variable continua con función de densidad f(x) y función de distribución acumulativa F(x), entonces con cada x que hace posible que la derivada F'(x),

$$\boxed{F'(x) = f(x)}$$

Scanned by CamScanner

PERCENTILES DE UNA DISTRIBUCIÓN CONTINUA

Sea p un número entre 0 y 1. El $(100p)^{\circ}$ percentil de la distribución de una variable aleatoria continua X, denotada por $\eta(p)$, se define como

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\,dy.$$

$\eta(p)$ es ese valor sobre el eje de medición de tal suerte que el $100p\%$ del área bajo la gráfica de $f(x)$ queda a la izquierda de $\eta(p)$ y $100(1-p)\%$ queda a la derecha.

MEDIANA

La MEDIANA de una distribución continua, denotada por $\tilde{\mu}$, es el 50° percentil, así que $\tilde{\mu}$ satisface $0{,}5 = F(\tilde{\mu})$. Es decir, la mitad del área bajo la curva de densidad se encuentra a la izq. de $\tilde{\mu}$ y la otra a la derecha.

En cumplimiento de lo establecido en el procedimiento de **FONDO ROTATORIO** se agrega el siguiente cuadro comparativo de las ofertas recibidas. Renglones N° 1-2

	PINTURERIA LEON	PINTURERIA VILLA NUEVA	FERRETERIA EL AGUILA
Cotización (Incluye I.V.A.)	$ 23.650.00	$ 15.960,00 R2: No cotiza lo solicitado	NO COTIZA
Forma de Pago	30 Días fecha Factura	30 Días fecha Factura	
Plazo de entrega	7 días	A convenir	

Visto el monto cotizado por la empresa, se concluye en que la oferta de la empresa **PINTURERIA LEON,** es la más conveniente económicamente, por lo que se propone adquirir el producto a la misma.

Agregado:

- Cotización de la empresa: **PINTURERIA LEON**
- Cotización de la empresa: **PINTURERIA "VILLA NUEVA"**
- E mail sin respuesta: **FERRETERIA " EL AGUILA"**

Scanned by CamScanner

DISTRIBUCIÓN NORMAL (ESTÁNDAR)

Z_α

Z_α denotará el valor sobre el eje z para el cual α del área bajo la curva z queda a la derecha de Z_α.
Como α del área bajo la curva z queda a la derecha de Z_α, $1-\alpha$ queda a la izquierda. Por lo tanto, Z_α es el $100(1-\alpha)^\circ$ PERCENTIL.

PERCENTILES Y VALORES CRÍTICOS

PERCENTIL	90	95	97,5	99	99,5	99,9	99,95
	↓	↓	↓	↓	↓	↓	↓
Z_α	1,28	1,645	1,96	2,33	2,58	3,08	3,27

REGLA EMPÍRICA →
Si la distribución de la población de una variable es (aproximadamente) normal, entonces:

1. Aproximadamente 68% de los valores están dentro de 1 DE de la media.
2. Aproximadamente 95% de los valores están dentro de 2DE de la media.
3. Aproximadamente 99,7% de los valores están dentro de 3DE de la media.

Es INUSUAL observar un valor de la población normal que esté mucho más lejos de 2 de. de μ.

PERCENTIL DE NORMAL ARBITRARIA

$$\begin{matrix}(100p)^\circ \text{ percentil} \\ \text{de } (\mu, \sigma) \text{ normal}\end{matrix} = \mu + \left[\begin{matrix}(100p^\circ) \text{ percentil} \\ \text{de normal estándar}\end{matrix}\right] \cdot \sigma$$

APROXIMACIÓN DE LA DISTRIBUCIÓN BINOMIAL

En términos generales, en tanto que el histograma de probabilidad binomial no sea demasiado asimétrico, las probabilidades binomiales pueden ser aproximadas muy bien por áreas de curva normal. Se acostumbra entonces decir que X tiene aproximadamente una distribución normal.

$$\boxed{P(X \leq x) = B(x, n, p) = \Phi\left(\frac{x + 0{,}5 - n.p}{\sqrt{n.p.q}}\right)}$$

Scanned by CamScanner

DISTRIBUCIÓN EXPONENCIAL

Se dice que X tiene una distribución EXPONENCIAL con parámetro λ ($\lambda > 0$) si la función de densidad de probabilidad de X es

$$\boxed{f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & c.c \end{cases}}$$

$E(X) = \int_0^\infty x.\lambda e^{-\lambda x} dx$ para obtener $E(X)$ se requiere integrar por partes, y $E(X^2)$ requiere hacerlo dos veces en sucesión.

Los resultados son los siguientes:

$$\boxed{\mu = \frac{1}{\lambda}} \quad \boxed{\sigma^2 = \frac{1}{\lambda^2}} \Rightarrow \mu = \sigma_x$$

La función de densidad acumulativa se obtiene integrando $f(x)$:

$$\boxed{F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}}$$

Esta distribución se utiliza con frecuencia como modelo de la distribución de tiempos entre la ocurrencia de eventos sucesivos (estrechamente relacionada con la de Poisson).

LA FUNCIÓN GAMA $\Gamma(\alpha)$

Con $\alpha > 0$, la función gama $\Gamma(\alpha)$ se define como

$$\boxed{\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} . e^{-x} . dx}$$

PROPIEDADES

1. Con cualquiera $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1).\Gamma(\alpha - 1)$
2. Con cualquier <u>entero</u> positivo n, $\Gamma(n) = (n-1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Scanned by CamScanner

DISTRIBUCIÓN GAMA

Se dice que una V.A.C. X tiene una distribución gama si la función de densidad de probabilidad de X es

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \text{c.c.} \end{cases}$$ donde $\alpha > 0$ y $\beta > 0$.

La distribución GAMA ESTÁNDAR tiene $\beta = 1$;

(la distr. exponencial se deriva de considerar $\alpha = 1$ y $\beta = 1/\lambda$

$\boxed{E(X) = \mu = \alpha . \beta}$ $\boxed{V(X) = \sigma^2 = \alpha . \beta^2}$

Cuando X es una variable aleatoria gama ESTÁNDAR, la función de distribución acumulativa de X,

$$F(x;\alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy \quad x > 0$$

DISTRIBUCIÓN JI CUADRADA

Sea ν un entero positivo. Se dice entonces que una variable aleatoria X tiene una distribución JI CUADRADA con parámetro ν si la función de densidad de probabilidad de X es la densidad gama con $\alpha = \nu/2$ y $\beta = 2$. Por lo tanto,

$$f(x;\nu) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

El parámetro ν se llama NRO. DE GRADOS DE LIBERTAD (gl) de X.
A menudo se utiliza el símbolo χ^2 en lugar de "ji cuadrada"

Scanned by CamScanner

DISTRIBUCIÓN WEIBULL

Se dice que una variable aleatoria X tiene una distribución WEIBULL con parámetros α y β ($\alpha>0$ y $\beta>0$) si la función de densidad de probabilidad de X es

$$f(x;\alpha,\beta)=\begin{cases}\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-(x/\beta)^{\alpha}} & x\geq 0\\ 0 & x<0\end{cases}$$

$$\boxed{\mu=\beta\Gamma\left(1+\frac{1}{\alpha}\right)}\qquad\boxed{\sigma^2=\beta^2\left\{\Gamma\left(1+\frac{2}{\alpha}\right)-\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2\right\}}$$

$$\boxed{F(x;\alpha,\beta)=\begin{cases}0 & x<0\\ 1-e^{-(x/\beta)^{\alpha}} & x\geq 0\end{cases}}$$

DISTRIBUCIÓN LOGNORMAL

Se dice que una variable aleatoria no negativa X tiene una distribución LOGNORMAL si la variable aleatoria $Y=\ln(X)$ tiene una distribución NORMAL, y ~~cuando~~ con parámetros μ y σ

$$f(x;\mu,\sigma)=\begin{cases}\frac{1}{\sqrt{2\pi}\,\sigma x}e^{-[\ln(x)-\mu]^2/(2\sigma^2)} & x\geq 0\\ 0 & x<0\end{cases}$$

$$\boxed{E(X)=e^{\mu+\sigma^2/2}}\qquad\boxed{V(X)=e^{2\mu+\sigma^2}\cdot(e^{\sigma^2}-1)}$$

$$F(x;\mu,\sigma)=P(X\leq x)=P[\ln(X)\leq\ln(x)]$$

$$=P\left(Z\leq\frac{\ln(x)-\mu}{\sigma}\right)=\boxed{\Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)}\quad x>0;$$

Scanned by CamScanner

CAP 5: DISTRIBUCIONES DE PROBABILIDAD CONJUNTA Y MUESTRAS ALEATORIAS

VARIABLES ALEATORIAS CONJUNTAMENTE DISTRIBUIDAS

→ DISCRETAS (2)

Sean X y Y dos V.A.D definidas en el espacio muestral S de un experimento. La función masa de probabilidad conjunta $p(x,y)$ se define para cada par de números (x,y) como

$$\boxed{p(x,y) = P(X=x \text{ y } Y=y)}$$

Debe cumplirse que $p(x,y) \geq 0$ y $\sum_x \sum_y p(x,y) = 1$.

$$\boxed{P[(X,Y)\in A] = \sum\sum_{(x,y)\in A} p(x,y)}$$

función masa de probabilidad marginal

$$\boxed{P_X(x) = \sum_y p(x,y)} \qquad \boxed{P_Y(y) = \sum_x p(x,y)}$$

→ CONTINUAS (2)

función de densidad conjunta

tal que $f(x,y) \geq 0$ y $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$.

Para cualquier conjunto A de dos dimensiones:

$$\boxed{P[(X,Y)\in A] = \iint_A f(x,y)\,dx\,dy}$$

En particular,

Si A es el rectángulo $\{(x,y): a \leq x \leq b,\ c \leq y \leq d\}$

$$\boxed{P[(X,Y)\in A] = P(a \leq X \leq b,\ c \leq y \leq d) = \int_a^b\int_c^d f(x,y)\,dy\,dx}$$

función de densidad de probabilidad marginal

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy} \qquad \boxed{f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx}$$

Scanned by CamScanner

VARIABLES ALEATORIAS INDEPENDIENTES

Dos variables aleatorias son independientes si por cada par de valores x y y,

$$p(x,y) = p_X(x) \cdot p_Y(y) \quad o \quad f(x,y) = f_X(x) \cdot f_Y(y)$$

MAS DE 2 VARIABLES ALEATORIAS

Si son discretas

$$p(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Si son contínuas, para n intervalos cualesquiera $[a_1, b_1], \dots, [a_n, b_n]$

$$P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \dots, x_n)\, dx_n \dots dx_1$$

Independencia:

Las V.A $X_1, X_2, \dots, X_n$ son INDEPENDIENTES si para cada subconjunto $X_{i_1}, \dots, X_{i_k}$ de las variables, la fmpc o fdpc del subconjunto es igual al producto de las fmp o fdp marginales

DISTRIBUCIONES CONDICIONALES

Sean X y Y dos V.A continuas con fdpc $f(x,y)$ y fdpm $X\, f_X(x)$. Entonces para cualquier valor x de X con el cual $f_X(x) > 0$,
la función de densidad de probabilidad CONDICIONAL DE Y DADO QUE $X = x$ es

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

Si X e Y son DISCRETAS, se reemplazan las $f(x,y)$ y $f_X(x)$ por las $p(x,y)$ y $p_X(x)$.

VALORES ESPERADOS

El VALOR ESPERADO de una función $h(X, Y)$ es

$$\boxed{E[h(X,Y)]} = \begin{cases} \sum_x \sum_y h(x,y) \cdot p(x,y) & \text{DISCRETAS} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \cdot f(x,y) \cdot dx\, dy & \text{CONTINUAS} \end{cases}$$

Scanned by CamScanner

COVARIANZA

Qué tan fuerte están relacionadas las variables entre sí cuando son DEPENDIENTES.

$$\boxed{Cov(X,Y)} = E[(X-\mu_X)(Y-\mu_Y)]$$

$$= \begin{cases} \sum_x \sum_y (x-\mu_x)(y-\mu_y) \cdot p(x,y) & \text{DISCRETAS} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) \cdot f(x,y)\, dx\, dy & \text{CONTINUAS} \end{cases}$$

Como $X-\mu_X$ e $Y-\mu_Y$ son las desviaciones de las dos variables con respecto a sus valores medios, la covarianza es el producto esperado de las desviaciones.

$$\underline{Cov(X,X) = E[(X-\mu_X)^2] = V(X)}$$

- Con una fuerte relación POSITIVA $\rightarrow$ $Cov(X,Y)$ deberá ser bastante positiva.

  NEGATIVA $\rightarrow$ $Cov(X,Y)$ NEGATIVA
  (signos opuestos de las desviaciones)

- Con una relación DÉBIL $\rightarrow$ $Cov(X,Y)$ cerca de 0

$$\boxed{Cov(X,Y) = E(XY) - \mu_X \cdot \mu_Y}$$

CORRELACIÓN

El coeficiente de correlación de X y Y,

$$\boxed{Corr(X,Y) = \rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}}$$

$\rightarrow$ Si a y c son ambas positivas o negativas

$$\boxed{Corr(aX+b, cY+d) = Corr(X,Y)}$$

$\rightarrow$ Para dos v.A cualesquiera X e Y,

$$\boxed{-1 \leq Corr(X,Y) \leq 1}$$

- Si X e Y independientes, entonces $\rho = 0$, pero $\rho = 0$ no implica independencia

- $\rho = 1$ o -1 si y sólo si $Y = aX + b$ con algunos a y b, y $a \neq 0$

$\Rightarrow$ ρ mide el grado de asociación lineal entre X e Y.

Un ρ menor que 1 en valor absoluto indica que la relación no es lineal, pero pueden estar relacionadas fuertemente NO LINEALMENTE.

$\rho = 0$ puede indicar que haya una dependencia no lineal.

Scanned by CamScanner

CAP VII: INTERVALOS ESTADÍSTICOS BASADOS EN UNA SOLA MUESTRA

PROPIEDADES BÁSICAS

Se supone que $x_1, \dots, x_n$ son el resultado de una muestra aleatoria $X_1, \dots, X_n$ tomada de una distribución normal con valor medio μ y desviación estándar σ.
Independientemente del tamaño de la muestra n, la MEDIA MUESTRAL $\bar{X}$ está normalmente distribuida con valor esperado μ y desviación estándar $\sigma/\sqrt{n}$. Si se estandariza $\bar{X}$ se obtiene la variable estándar

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

Como el área bajo la curva normal estándar entre $-1{,}96$ y $1{,}96$ es $0{,}95$,

$$P\left(-1{,}96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1{,}96\right) = 0{,}95 .$$

Ahora manípulense las desigualdades que están adentro del paréntesis de modo que aparezcan de la forma equivalente $l < \mu < u$, donde los puntos l y u implican a $\bar{X}$ y $\sigma/\sqrt{n}$:

1. <u>Multiplíquese por $\sigma/\sqrt{n}$</u>:

$$-1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}$$

2. <u>Réstese $\bar{X}$ de cada término</u>:

$$-\bar{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} - \bar{X}$$

3. <u>Multiplíquese por -1</u>:

$$\bar{X} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$= \boxed{\bar{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1{,}96 \frac{\sigma}{\sqrt{n}}}$$

<u>INTERVALO ALEATORIO:</u>

$$\left(\bar{X} - 1{,}96 \frac{\sigma}{\sqrt{n}} ,\ \bar{X} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right)$$

El intervalo es ALEATORIO porque los dos puntos extremos implican una variable, pero el ancho es $2 \cdot 1{,}96 \cdot \sigma/\sqrt{n}$, el cual no es aleatorio.

Scanned by CamScanner

Si después de observar $X_1 = x_1$, $X_2 = x_2$, ..., $X_n = x_n$. se calcula la media muestral observada $\bar{x}$ y luego se sustituye $\bar{x}$ en lugar de $\bar{X}$, el intervalo FIJO resultante se llama INTERVALO DE CONFIANZA DE 95% PARA μ. Este intervalo de confianza se expresa como

$$\left(\bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right)$$ es un intervalo de 95% de confianza para μ.

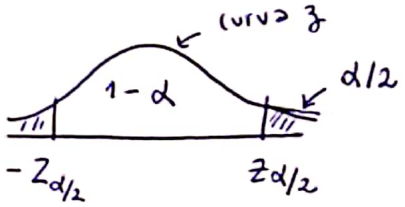
Una expresión concisa para el intervalo es $\bar{x} \pm 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}$.

INTERPRETACIÓN DE UN INTERVALO DE CONFIANZA

Una interpretación correcta de "95% de confianza" se basa en la interpretación de probabilidad de frecuencia relativa a largo plazo. Decir que un evento A tiene una probabilidad del 0,95 implica que si el experimento se repite varias veces, a la larga A ocurrirá el 95% del tiempo, y el 95% de los intervalos de confianza que se vayan calculando contendrán a μ.

OTROS NIVELES DE CONFIANZA

99% Resulta de utilizar 2,58 en lugar de 1,96. Esto sugiere que cualquier nivel de confianza deseado se obtiene reemplazando 1,96 o 2,58 con el valor crítico normal estándar apropiado



La siguiente expresión da un intervalo de confianza de $100(1-\alpha)\%$ para la media μ de una población normal cuando se conoce el valor de σ

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$ o de forma equivalente,

por $\bar{x} \pm Z_{\alpha/2} \cdot \sigma/\sqrt{n}$

Scanned by CamScanner

NIVEL, PRECISIÓN Y TAMAÑO DE MUESTRA

El precio pagado por el nivel de confianza más alto es un intervalo más ancho. Si se considera que el ancho del intervalo especifica su precisión o exactitud, entonces el nivel de confianza (o confiabilidad) del intervalo está relacionado de manera inversa con su precisión.

Una estrategia atractiva es especificar tanto el nivel de confianza deseado como el ancho del intervalo y luego determinar el tamaño de muestra necesario.

La FÓRMULA GENERAL para el tamaño de muestra n necesario para garantizar un ancho de intervalo w se obtiene a partir de $w = 2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$ como

$$\boxed{n = \left(2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2}$$

Mientras más pequeño es el ancho deseado w, más grande debe ser n.

Además, n es una función creciente de σ y del nivel de confianza $100(1-\alpha)$

La mitad del ancho $z_{\alpha/2} \cdot \sigma/\sqrt{n}$ del intervalo de confianza en ocasiones se llama LÍMITE DE ERROR DE ESTIMACIÓN asociado con determinado nivel de confianza.

DERIVACIÓN DE UN INTERVALO DE CONFIANZA

Sean $X_1, X_2, \dots, X_n$ la muestra en la cual se tiene que basar el intervalo de confianza para un parámetro θ. Supóngase que se puede determinar una variable aleatoria que satisface las dos siguientes propiedades:

1. La variable depende funcionalmente tanto de $X_1, \dots, X_n$ como de θ.
2. La distribución de probabilidad de la variable no depende de θ ni de cualesquiera otros parámetros desconocidos.

Sea $h(X_1, \dots, X_n; \theta)$ esta variable aleatoria. Por ejemplo, si la distribución de la población es normal con σ conocida y $\theta = \mu$ la variable $h(X_1, \dots, X_n; \theta) = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ satisface ambas propiedades; Depende funcionalmente de μ, no obstante su distribución de probabilidad es normal estándar, la cual no depende de μ.

En general, la forma de la función h casi siempre se pone de manifiesto al examinar la distribución de un estimador apropiado $\hat{\theta}$.

Scanned by CamScanner

Con cualquier α entre 0 y 1, se ve que las constantes a y b satisfacen

$$P(a < h(X_1, \dots, X_n; \theta) < b) = 1 - \alpha$$

A causa de la segunda propiedad, a y b no dependen de θ. En el ejemplo normal, $a = -Z_{\alpha/2}$ y $b = Z_{\alpha/2}$. Ahora, manipulando las desigualdades de arriba para aislar θ se obtiene la proposición de probabilidad equivalente

$$P(l(X_1, \dots, X_n) < \theta < u(X_1, \dots, X_n)) = 1 - \alpha$$

INTERVALOS DE CONFIANZA DE MUESTRA GRANDE PARA UNA MEDIA Y PROPORCIÓN DE POBLACIÓN

A continuación se presenta un intervalo de confianza que no necesita suponer una distribución normal ni un σ conocido.

INTERVALO DE MUESTRA GRANDE PARA μ

Sean $X_1, X_2, \dots, X_n$ una muestra aleatoria de una población con media μ y desviación estándar σ. Siempre que n es grande, el TLC establece que $\bar{X}$ tiene de manera aproximadamente una distribución normal. Se deduce entonces que $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ tiene aprox. un distr. normal estándar, de modo que

$$P\left(-Z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right) \approx 1 - \alpha$$

Si n es suficientemente grande, la variable estandarizada $Z = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$ tiene aproximadamente una distribución normal estándar. Esto implica que $\bar{x} \pm Z_{\alpha/2} \cdot \dfrac{s}{\sqrt{n}}$ es un INTERVALO DE CONFIANZA DE MUESTRA GRANDE PARA μ con nivel de confianza aproximadamente de $100(1-\alpha)\%$. Esta fórmula es válida sin importar la forma de la distribución de la población.

Scanned by CamScanner

UN INTERVALO DE CONFIANZA DE MUESTRA GRANDE GENERAL

Suponga que $\hat{\theta}$ es un estimador que satisface las siguientes propiedades:

1. Tiene aproximadamente una distribución normal
2. Es insesgado (por lo menos aproximadamente)
3. Una expresión para $\sigma_{\hat{\theta}}$ está disponible.

Por ejemplo, si $\theta = \mu$, $\hat{\mu} = \bar{X}$ que es insesgado y distribuido normalmente para n grande, y $\sigma_{\hat{\mu}} = \sigma_{\bar{X}} = \sigma/\sqrt{n}$

Estandarizando $\hat{\theta}$, se obtiene la variable aleatoria $Z = (\hat{\theta} - \theta)/\sigma_{\theta}$, la cual tiene aproximadamente una distribución normal estándar. Esto justifica la proporción de probabilidad $P\left(-Z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < Z_{\alpha/2}\right) \approx 1 - \alpha$. Suponga primero que $\sigma_{\hat{\theta}}$ parámetros desconocidos. Entonces, si se reemplaza los $<$ por $=$ se obtiene $\theta = \hat{\theta} \pm Z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$. Suponga ahora que $\sigma_{\hat{\theta}}$ no implica θ pero sí implica por lo menos otro parámetro desconocido. Sea $S_{\hat{\theta}}$ la estimación de $\sigma_{\hat{\theta}}$ utilizando estimaciones, entonces en condiciones generales, un intervalo de confianza válido es $\hat{\theta} \pm Z_{\alpha/2} \cdot S_{\hat{\theta}}$. El intervalo muestral grande $\bar{x} \pm Z_{\alpha/2} \cdot s/\sqrt{n}$ es un ejemplo. Por último, suponga que $\sigma_{\hat{\theta}}$ no implica el θ desconocido. Este es el caso de $p = \theta$, una proporción de población.

UN INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN DE POBLACIÓN

Sea p la proporción de éxitos en una población, donde éxito identifica a un individuo u objeto que tiene una propiedad específica. Una variable aleatoria de n individuos que tiene que ser seleccionado y X es el número de éxitos de la muestra.

> UN INTERVALO DE CONFIANZA PARA UNA ~~POBLACIÓN~~ PROPORCIÓN DE POBLACIÓN P con nivel de confianza aproximadamente de $100(1-\alpha)\%$ tiene
>
> límite de confianza inferior $= \dfrac{\hat{p} + \frac{Z^2_{\alpha/2}}{2n} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{Z^2_{\alpha/2}}{4n^2}}}{1 + (Z^2_{\alpha/2})/n}$
>
> límite de confianza superior $= \dfrac{\hat{p} + \frac{Z^2_{\alpha/2}}{2n} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{Z^2_{\alpha/2}}{4n^2}}}{1 + (Z^2_{\alpha/2})/n}$

Si n es suficientemente grande hay ciertos términos que se vuelven insignificantes, y al desecharlos se obtiene

$$\boxed{\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}}$$

w = ancho preespecificado

y $n \approx \dfrac{4Z^2\hat{p}\hat{q}}{w^2}$ → método conservador: suponer que $\hat{p} = \hat{q} = 0.5$. implica que el ancho será A LO MUCHO w.

Scanned by CamScanner

INTERVALOS DE CONFIANZA UNILATERALES
(LÍMITES DE CONFIANZA)

Un LÍMITE DE CONFIANZA SUPERIOR muestral grande para μ es

$\mu < \bar{x} + z_\alpha \cdot \frac{s}{\sqrt{n}}$, y uno INFERIOR $\mu > \bar{x} - z_\alpha \cdot \frac{s}{\sqrt{n}}$.

Se obtiene un LÍMITE DE CONFIANZA UNILATERAL PARA p reemplazando $z_{\alpha/2}$ en lugar de z_α y $\pm$ en lugar de + o − en la fórmula para el intervalo de confianza para p.
En todos los casos el nivel de confianza es aprox $100(1-\alpha)\%$

7.3. INTERVALOS BASADOS EN UNA DISTRIBUCIÓN DE POBLACIÓN NORMAL

Como la distribución normal es más frecuentemente apropiada como modelo de una población que cualquier otro tipo de distribución, nos concentraremos en un intervalo de confianza para esta situación (Estas suposiciones son necesarias para un n no grande).

Cuando $\bar{x}$ es la media de una muestra aleatoria de tamaño n tomada de una distribución normal con media μ, la variable aleatoria

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

tiene una distribución de probabilidad llamada distribución t con n-1 grados de libertad (gl)

PROPIEDADES DE DISTRIBUCIONES t

Una distribución t está regida por sólo un (1) PARÁMETRO, llamado número de GRADOS DE LIBERTAD de la distribución, gl, que se denota con la letra griega ν. Posibles valores de ν son los enteros positivos. Cada diferente valor de ν corresponde a una distribución t diferente.

Sea t_ν, la curva de función de densidad para el grado de libertad ν.

1. Cada curva t_ν tiene forma de campana y con su centro en 0
2. Cada curva t_ν está más esparcida que la curva (2) normal estándar.
3. Conforme ν se incrementa, la dispersión de t_ν correspondiente se disminuye.
4. A medida que $\nu \to \infty$, la secuencia de curvas t_ν tiende a la curva normal estándar (así que z a menudo se llama curva t_ν con $\nu = \infty$)

Scanned by CamScanner

Como se desea utilizar T para obtener un intervalo de confianza, es necesario establecer una notación análoga a Z_α para la distribución t:

Sea $t_{\alpha,\nu}$ = el número sobre el eje de medición con el cuál el área bajo la curva t con ν grados de libertad a la derecha de $t_{\alpha,\nu}$ es α; $t_{\alpha,\nu}$ se llama VALOR CRÍTICO.

INTERVALO DE CONFIANZA t PARA UNA MUESTRA

La variable estandarizada T tiene una distribución t con n-1 grados de libertad y el área bajo la curva de densidad t correspondiente entre $-t_{\alpha/2,n-1}$ y $t_{\alpha/2,n-1}$ es $1-\alpha$ (el área $\alpha/2$ queda en cada cola), por consiguiente

$$P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = 1-\alpha$$

Sean $\bar{x}$ y s la media y la desviación estándar muestrales calculadas con los resultados de una muestra aleatoria tomada de una población normal con media μ. Entonces un INTERVALO DE CONFIANZA DE $100(1-\alpha)\%$ para μ es

$$\left(\bar{x} - t_{\alpha/2,n-1}\cdot\frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2,n-1}\cdot\frac{s}{\sqrt{n}}\right)$$

Por desgracia, no es fácil seleccionar n para controlar el ancho del intervalo t. Esto es porque el ancho implica a la s desconocida y porque n ingresa no sólo a través de $1/\sqrt{n}$ sino tmb en $t_{\alpha/2,n-1}$. Por consiguiente, se obtiene una n apropiada sólo mediante ensayo y error.

UN INTERVALO DE PREDICCIÓN PARA UN SÓLO VALOR FUTURO

En muchas aplicaciones, se desea PREDECIR un sólo valor de una variable que tiene que ser observada en un tiempo futuro, en lugar de ESTIMAR el valor medio de dicha variable.

Un INTERVALO DE PREDICCIÓN (IP) para una sola observación que tiene que ser seleccionado de una distribución de población normal es

$$\bar{x} \pm t_{\alpha/2,n-1}\cdot s\sqrt{1+\frac{1}{n}}.$$ El nivel de predicción es $100(1-\alpha)\%$.

Scanned by CamScanner

INTERVALOS DE TOLERANCIA

Sea k un número entre 0 y 100. Un intervalo de tolerancia para capturar por lo menos el k% de los valores en una distribución de población normal con nivel de confianza de 95% tiene la forma $\bar{x} \pm$ (valor crítico de tolerancia). s

(TABLA A.6)

INTERVALOS BASADOS EN DISTRIBUCIONES NO NORMALES

No SIRVEN. (BOOTSTRAP)

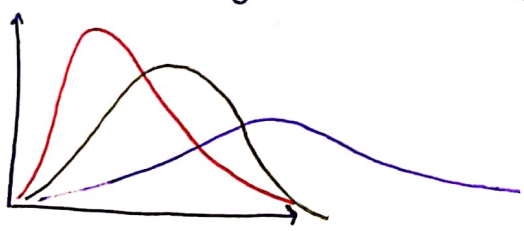
7.4 INTERVALOS DE CONFIANZA PARA LA VARIANZA Y DESVIACIÓN ESTÁNDAR DE UNA POBLACIÓN NORMAL.

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de un distribución normal con parámetros μ y σ^2. Entonces la variable aleatoria

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(x_i - \bar{x})}{\sigma^2}$$

tiene una distribución de probabilidad JI CUADRADA (χ^2) con $n-1$ grados de libertad.

Las gráficas de varias funciones de probabilidad de χ^2 se ilustran así: cada función de distribución de probabilidad $f(x; \nu)$ es positiva sólo con $x > 0$ y cada una tiene una asimetría positiva (una larga cola superior), aunque la distribución se mueve hacia la derecha y se vuelve más simétrica a medida que se incrementa ν.



Sea $\chi^2_{\alpha,\nu}$ llamado VALOR CRÍTICO JI CUADRADO, el número sobre el eje de medición de modo que α del área bajo la curva ji cuadrada con ν grados de libertad quede a la derecha de $\chi^2_{\alpha,\nu}$.

Scanned by CamScanner

La distribución Ji cuadrada no es simétrica, por lo que la tabla contiene valores de $\chi^2_{\alpha,r}$ tanto para α cerca de 0 como cerca de 1.

La variable aleatoria $(n-1)S^2/\sigma^2$ satisface los 2 parámetros en los cuales está basado el método para hallar intervalos de confianza. Es una función del parámetro σ^2, no obstante su distribución (Ji cuadrada) no depende de este parámetro.

Un intervalo de confianza de $100(1-\alpha)\%$ para la varianza σ^2 de una población normal tiene un límite inferior $(n-1)s^2/\chi^2_{\alpha/2,n-1}$ y un límite superior $(n-1)s^2/\chi^2_{1-\alpha/2,n-1}$. Un INTERVALO DE CONFIANZA para σ se obtiene con las raíces cuadradas de éstos límites.

CAP VIII: PRUEBAS DE HIPÓTESIS BASADAS EN UNA SOLA MUESTRA.

Scanned by CamScanner

PRUEBA DE HIPÓTESIS BASADA EN UNA SOLA MUESTRA

$$H_a: \begin{cases} \theta \neq \theta_0 & \text{Prueba bilateral} \\ \theta > \theta_0 & \text{Prueba unilateral a cola superior} \\ \theta < \theta_0 & \text{Prueba unilateral a cola inferior} \end{cases}$$

hipótesis alternativa (contrapuesta a $H_0: \theta = \theta_0$)

→ ESTADÍSTICO DE PRUEBA: Estadístico con distribución CONOCIDA bajo H_0, sobre el cual se basará la decisión a tomar.

→ REGIÓN DE RECHAZO: Conjunto de valores del estadístico de prueba para los cuales H_0 será rechazada. O sea, H_0 será rechazada si el valor OBSERVADO del estadístico pertenece a la RR, en cuyo caso diremos que H_0 es correcta.

Al tomar una decisión se pueden cometer dos tipos de errores:

ERROR TIPO I: Rechazar H_0 cuando H_0 es V $\longrightarrow \alpha$

TIPO II: Aceptar H_0 cuando H_0 es F $\longrightarrow \beta$

La RR determina las probabilidades de cometer cada uno de los errores.

→ NIVEL DE SIGNIFICACIÓN DE UNA PRUEBA

$$\alpha = \max_{\theta \in \Omega_0} \alpha(\theta) \quad \text{donde } H_0: \theta \in \Omega_0$$

PASOS A REALIZAR:

A) Identificar el PARÁMETRO DE INTERÉS.

B) Determinar la H_0 y H_a

C) Determinar el ESTADÍSTICO DE PRUEBA adecuado con distribución conocida, bajo H_0.

D) Fijado un nivel de significancia α, determinar la RR

E) Calcular el VALOR OBSERVADO del estadístico de prueba con la muestra obtenida.

F) Determinar si H_0 debe ser RECHAZADA o no, para el nivel de significancia dado.

~~MUESTRA~~

Scanned by CamScanner

MEDIA POBLACIONAL μ

$H_0: \mu = \mu_0$

CASO A: POBLACIÓN NORMAL CON σ CONOCIDA

ESTADÍSTICO:

$$\boxed{Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}}$$ ← distancia entre $\bar{x}$ y μ_0 expresada en unidades de desviación estándar.

¿VALOR DE CORTE C? Deberá ser elegido para controlar la probabilidad de un error de tipo I al nivel α deseado

⇒ Valor crítico z que captura el área de la cola superior α bajo la curva z

$H_a: \boxed{\mu > \mu_0}$ ⇒ R.R: $Z \geq C = Z_\alpha$ PRUEBA COLA SUPERIOR

$H_a: \boxed{\mu < \mu_0}$ ⇒ R.R: $Z \leq C = -Z_\alpha$ ($\bar{x}$ aparece muy debajo de μ_0 si y solo si z es bastante NEGATIVO)

$H_a: \boxed{\mu \neq \mu_0}$ ⇒ R.R: ~~$Z \geq C$ || $Z \leq -C$~~

$Z \geq Z_{\alpha/2}$ || $Z \leq -Z_{\alpha/2}$ PRUEBA DE DOS COLAS

α: nivel de significancia = P(error de tipo I)

$\beta(\mu')$: P(error tipo II) para una prueba de nivel α

$H_a: \boxed{\mu > \mu_0}$ ⟹ $\Phi\left(Z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

$\boxed{\mu < \mu_0}$ ⟹ $1 - \Phi\left(-Z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

$\boxed{\mu \neq \mu_0}$ ⟹ $\Phi\left(Z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-Z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

TAMAÑO DE MUESTRA N con el cual una prueba de nivel α también tiene $\beta(\mu') = \beta$

$$n = \begin{cases} \left[\frac{\sigma(Z_\alpha + Z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{UNA COLA} \\ \left[\frac{\sigma(Z_{\alpha/2} + Z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{DOS COLAS (solución aproximada)} \end{cases}$$

Scanned by CamScanner

CASO B: MUESTRA GRANDE $n > 40$

Un tamaño de muestra grande implica que $Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ tiene aprox. $\sim N(0,1)$

Lo mismo que el caso anterior, pero todo es APROXIMADO.

ESTADÍSTICO: $\boxed{Z = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}} \sim N(0,1)$

CASO III: POBLACIÓN NORMAL, N PEQUEÑA, σ DESCONOCIDA

ESTADÍSTICO: $t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$

→ REGIÓN DE RECHAZO PARA UNA PRUEBA DE NIVEL α

H_0: $\boxed{\mu > \mu_0} \Rightarrow t \geq t_{\alpha, n-1}$ (cola superior)

$\boxed{\mu < \mu_0} \Rightarrow t \leq t_{\alpha, n-1}$ (cola inferior)

$\boxed{\mu \neq \mu_0} \Rightarrow t \geq t_{\alpha/2, n-1} \;\|\; t \leq -t_{\alpha/2, n-1}$ (dos colas).

→ β Y DETERMINACIÓN DEL TAMAÑO DE MUESTRA

Scanned by CamScanner

PROPORCIÓN DE LA POBLACIÓN (p)

CON MUESTRA GRANDE

$$H_0: p = p_0$$

ESTADÍSTICO:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1) \qquad \hat{p} = X/n$$

REGIÓN DE RECHAZO

H_a: $p > p_0 \Rightarrow Z \geq Z_\alpha$ cola superior

$p < p_0 \Rightarrow Z \leq -Z_\alpha$ cola inferior

$p \neq p_0 \Rightarrow Z \geq Z_{\alpha/2}$ o $Z \leq -Z_{\alpha/2}$ dos colas

{ESTOS PROCEDIMIENTOS DE PRUEBA SON VÁLIDOS SIEMPRE QUE $n.p_0 \geq 10$ y $n(1-p_0) \geq 10$}

β y TAMAÑO DE MUESTRA

$\beta(p')$:

H_a: $p > p_0 \Rightarrow \Phi\left[\frac{p_0 - p' + Z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

$p > p_0 \Rightarrow 1 - \Phi\left[\frac{p_0 - p' - Z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

$p \neq p_0 \Rightarrow \Phi\left[\frac{p_0 - p' + Z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right] - \Phi\left[\frac{p_0 - p' - Z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

n:

$$n = \begin{cases} \left[\frac{Z_\alpha\sqrt{p_0(1-p_0)} + Z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{- prueba de UNA COLA} \\ \left[\frac{Z_{\alpha/2}\sqrt{p_0(1-p_0)} + Z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{- prueba de DOS COLAS} \end{cases}$$

Scanned by CamScanner

PRUEBAS CON MUESTRA PEQUEÑA

DISTRIBUCIÓN BINOMIAL

$H_0: p = p_0$

ESTADÍSTICO: $\boxed{X = \sum_{i=1}^{n} x_i} \sim B(n, p_0)$

REGIÓN DE RECHAZO

PARA $H_a: p > p_0$

$\alpha = P(\text{ERROR TIPO I}) = P(H_0 \text{ RECHAZADA CUANDO V})$

$= P(X \geq c \text{ cuando } X \sim bin(n, p_0))$

$= 1 - P(X \leq c-1 \text{ cuando } X \sim bin(n, p_0))$

$= 1 - B(c-1; n, p_0)$

$B(x, n, p) \rightarrow$ acumulada de $b(x)$

Se utiliza la RR más grande de la forma $\{c, c+1, \dots, n\}$ que satisfaga $1 - B(c-1; n, p_0) \leq \alpha$

$\beta(p')$

$\beta(p') = P(\text{error tipo II}) \text{ cuando } p = p')$

$= P(X < c \text{ cuando } X \sim Bin(n, p'))$

$= B(c-1; n, p')$

PARA $H_a: p < p_0$

Scanned by CamScanner

El tiempo que requiere una rata de cierta subespecie seleccionada al azar para encontrar su camino a través de un laberinto es una variable aleatoria normalmente distribuida con $\mu = 1.5$ min y $\sigma = 0.35$ min. Suponga que se seleccionan 5 ratas. Sean $X_1, \dots, X_5$ sus tiempos. Suponiendo que las X_i son una muestra aleatoria de esta distribución normal, ¿cuál es la probabilidad ~~total~~ de que el tiempo total $T_0 = X_1 + \dots + X_5$ de las cinco sea de entre 6 y 8 min?

De acuerdo con la proposición, T_0 tiene una distribución normal con $\mu_{T_0} = n\mu = 5(1,5) = 7,5$, y varianza $\sigma^2_{T_0} = n\sigma^2 = 5(0,1225) = 0,6125$.

Por tanto, $\sigma_{T_0} = 0,783$.

Para ESTANDARIZAR T_0 reste μ_{T_0} y divida entre σ_{T_0}:

$$P(6 \leq T_0 \leq 8) = P\left(\frac{6-7,5}{0,783} \leq Z \leq \frac{8-7,5}{0,783}\right)$$

$$= P(-1,92 \leq Z \leq 0,64)$$

$$= \Phi(0,64) - \Phi(-1,92) = 0,7389 - (\Phi(1,92) - \Phi(0))$$

$$= 0,7389 - (0,9726 - \quad = 0,7389 - (1 - \Phi(1,92))$$

$$= 0,7389 - (1 - 0,9726) = 0,7389 - 0,0274$$

$$= \underline{0,7115.}$$

La determinación de la probabilidad de que el tiempo promedio muestral $\bar{X}$ sea cuando mucho de 2 min requiere que

$\mu_{\bar{X}} = \mu = 1,5$, y $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0,35/\sqrt{5} = 0,1565$. Entonces

$$P(\bar{X} \leq 2,0) = P\left(Z \leq \frac{2-1,5}{0,1565}\right) = P(Z \leq 3,19) = \Phi(3,19) = 0,9993. \blacksquare$$

Scanned by CamScanner

TEOREMA DEL LÍMITE CENTRAL

EJEMPLO 5.26: La cantidad de una impureza particular en un lote de cierto producto es una variable aleatoria con valor medio de 4.0 g y desviación estandar de 1,5 g. Si se preparan 50 lotes en forma independiente ¿Cuál es la probabilidad aprox. de que la cantidad promedio muestral de la impureza $\bar{X}$ sea de 3,5 a 3,8 gr? (considérese n=50 suficiente grande).

Por el teorema TLC, $\bar{X}$ tiene aprox una media de $\mu_{\bar{X}} = \mu = 4$ gr. y una desviación estándar de $\sigma_{\bar{X}} = 1,5/\sqrt{50} = 0,2121$. Por lo tanto,

$$P(3,5 \leq \bar{X} \leq 3,8) \simeq P\left(\frac{3,5-4}{0,2121} \leq Z \leq \frac{3,8-4}{0,2121}\right) = P(-2,36 \leq Z \leq -0,94)$$

$$= \Phi(-0,94) - \Phi(-2,36) = (1-\Phi(0,94)) - (1-\Phi(2,36)) = (1-0,8254) - (1-0,9909)$$

$$= \underline{\underline{0,1655}}.$$

EJEMPLO 5.27: Una organización reporta cotidianamente el nro. de defectos mayores de cada automóvil nuevo que prueba. Suponga que el nro de tales defectos en cierto modelo es una variable aleatoria con valor medio de 3,2 y desviación estándar de 2,4. Entre 100 carros elegidos al azar de este modelo, ¿qué tan probable es que el número promedio muestral de defectos exceda de 4?

Sea X_i el nro de defectos mayores del i-ésimo carro en la muestra aleatoria. Obsérvese que X_i es una variable aleatoria discreta, pero el TLC es aplicable tanto si continua como si no. Además, con una desviación estándar tan grande respecto al valor medio, puede suponerse que su distribución es positivamente asimétrica. El gran tamaño de muestra implica que $\bar{X}$ sí tiene distribución normal con $\mu_{\bar{X}} = 3.2$ y $\sigma_{\bar{X}} = 0,24$.

$$P(\bar{X} > 4) \approx P\left(Z > \frac{4-3.2}{0,24}\right) = P(Z > 3,33) = 1 - \Phi(3,33) = 1 - 0,9996 = \underline{0,0004}.$$

DISTRIBUCIÓN DE UNA COMBINACIÓN LINEAL

EJEMPLO 5.28: Una gasolinería vende tres grados de gasolina: regular, extra y súper. Éstas se venden a \$21,20, \$21,35 y \$21,50 por galón respectivamente. Sean X_1, X_2 y X_3 las cantidades (galones) de estas gasolinas compradas en un día particular. Suponga que las X_i son independientes con $\mu_1 = 1000$, $\mu_2 = 500$ y $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$ y $\sigma_3 = 50$. El ingreso por las ventas es $\boxed{Y = 21,2X_1 + 21,35X_2 + 21,5 \cdot X_3}$ y

$$E(Y) = 21,2 \cdot \mu_1 + 21,35 \cdot \mu_2 + 21,5 \cdot \mu_3 = 21,2 \cdot 1000 + 21,35 \cdot 500 + 21,5 \cdot 300$$

$$= 38325$$

$$V(Y) = (21,2)^2\sigma_1^2 + (21,35)^2\sigma_2^2 + (21,5)^2\sigma_3^2 = 449,44 \cdot 100 + 455,82 \cdot 80 + 462,25 \cdot 50$$

$$= 104\,025$$

(ESTÁ COMO EL DATO EN EL LIBRO?)

Scanned by CamScanner

DIFERENCIA ENTRE DOS VARIABLES ALEATORIAS

EJEMPLO 5.29 Una compañía automotriz equipa un modelo particular con un motor de seis cilindros o uno de cuatro cilindros. Sean X_1 y X_2 eficiencias de combustible de carros de seis o cuatro seleccionados al azar independientemente, respectivamente.

Con $\mu_1 = 22$, $\mu_2 = 26$ y $\sigma_1 = 1{,}2$ y $\sigma_2 = 1{,}5$

$$E(X_1 - X_2) = \mu_1 - \mu_2 = 22 - 26 = -4$$

$$V(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 = (1{,}2)^2 + (1{,}5)^2 = 3{,}69$$

$$\sigma_{X_1 - X_2} = \sqrt{3{,}69} = 1{,}92.$$

DISTR. DE COMBINACIÓN LINEAL DE V.A DISTRIBUÍDAS NORMALMENTE.

(CONTINUACIÓN DE 5.28)

EJEMPLO 5.30: El ingreso total por la venta de los tres grados de gasolina en un día en particular fue $Y = 21{,}2X_1 + 21{,}35X_2 + 21{,}5X_3$ y se calculó $\mu_Y = 4125$ y (suponiendo independencia) $\sigma_Y = 322{,}53$. Si las X_i están normalmente distribuídas, la probabilidad de que el ingreso sea de más de 4500 es

$$P(Y > 4500) = P\left(Z > \frac{4500 - 4125}{322{,}53}\right) = P(Z > 1{,}16) = 1 - \Phi(1{,}16) = 1 - 0{,}8770$$

$$= \underline{\underline{0{,}123}}$$

ESTIMACIÓN PUNTUAL

EJEMPLO 6.1: Un fabricante automotriz ha producido un nuevo tipo de defensa. El fabricante ha utilizado esta defensa en una secuencia de 25 choques controlados con un muro, cada uno a 10 mph, utilizando uno de sus modelos. Sea X = el número de choques que no provocaron daños visibles al automóvil. El parámetro que tiene que ser estimado es p = la proporción de todos los choques que no produjeron daños. [alternativamente, $p = P(\text{ningún daño})$].

Si se observa que X es $x = 15$, el estimador y estimación ~~posible~~ mas razonables son

estimador $\hat{p} = \dfrac{X}{n}$ estimación $= \dfrac{x}{n} = \dfrac{15}{25} = 0{,}60.$

Scanned by CamScanner

El VM

Ejemplo 6.6 Anteriormente, se argumentó que cuando $X_1, \ldots, X_n$ es una variable aleatoria tomada de una distribución uniforme en el intervalo $[0, \theta]$, el estimador

$$\hat{\theta}_1 = \frac{n+1}{n} \cdot \text{máx}(X_1, \ldots, X_n)$$ es insesgado para θ. Éste no es el único estimador insesgado de θ.

El valor esperado de una variable uniformemente distribuida es simplemente el punto medio del intervalo de densidad positiva, por lo tanto $E(X_i) = \theta/2$. Esto implica que $E(\bar{X}) = \theta/2$, a partir de la cual $E(2\bar{X}) = \theta$. Es decir, el estimador $\hat{\theta}_2 = 2\bar{X}$ es insesgado para θ.

Si X está uniformemente distribuida en el intervalo A, B, en ese caso $V(X) = \sigma^2 = (B-A)^2/12$. Así pues, en esta situación, $V(X_i) = \theta^2/12$, $V(\bar{X}) = \sigma^2/n = \theta^2/12n$, y $V(\hat{\theta}_2) = V(2\bar{X}) = 4V(\bar{X}) = \theta^2/3n$ ...

ERROR ESTÁNDAR

EJEMPLO 6.9 (Cont. 6.2): Suponiendo que el voltaje de ruptura está normalmente distribuído, $\hat{\mu} = \bar{X}$ es la mejor estimación de μ. Si se sabe que el valor de σ es 1.5, el error estándar de $\bar{X}$ es $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 1{,}5/\sqrt{20} = 0{,}335$.
Si, como casi siempre es el caso, el valor de σ es desconocido, la estimación $\hat{\sigma} = s = 1{,}462$ se sustituye en $\sigma_{\bar{X}} = s_{\bar{X}} = s/\sqrt{n} = 1.462/\sqrt{20} = 0{,}327$.

EJEMPLO 6.10 (cont. 6.1): El error estándar de $\hat{p} = X/n$ es

$$\sigma_{\hat{p}} = \sqrt{V(X/n)} = \sqrt{\frac{V(X)}{n^2}} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$$

Como p y $q = 1-p$ son desconocidas, se sustituye $\hat{p} = X/n$ y $\hat{q} = 1 - X/n$ en $\sigma_{\hat{p}}$ para obtener el error estándar estimado $\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n} = \sqrt{(0{,}6)(0{,}4)/25} = 0{,}098$.

Alternativamente, como el valor más grande de pq se obtiene ~~como~~ cuando $p = q = 0{,}5$, un límite superior en el error estándar es $\sqrt{1/4(n)} = 0{,}10$.

Scanned by CamScanner

MÉTODO DE MOMENTOS

EJEMPLO 6.12: Si $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria de tiempos de servicio de n clientes en una instalación, donde la distribución subyacente se supone EXPONENCIAL con parámetro λ.

Como sólo hay un parámetro que tiene que estar estimado, el estimador se obtiene igualando $E(X) \Rightarrow \bar{X}$. Como $E(X) = 1/\lambda$ con distribución exponencial, ésta da $\bar{X} = 1/\lambda$, o $\lambda = 1/\bar{X}$.
El estimador de este momento de λ es entonces $\boxed{\hat{\lambda} = 1/\bar{X}}$

EJEMPLO 6.13 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución gama con parámetros α y β. De acuerdo con la sección 4.4, $E(X) = \alpha\beta$ y $E(X^2) = \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)}$ $= \beta^2(\alpha+1)\alpha$. Los estimadores de momento α y β se obtienen resolviendo

$$\bar{X} = \alpha\beta \qquad \frac{1}{n}\sum X_i^2 = \alpha(\alpha+1)\beta^2$$

Como $\alpha(\alpha+1)\beta^2 = \alpha^2\beta^2 + \alpha\beta^2$, y la primera ecuación implica $\alpha^2\beta^2 = \bar{X}^2$, la segunda ecuación se vuelve

$$\frac{1}{n}\sum X_i^2 = \bar{X}^2 + \alpha\beta^2.$$ Ahora se divide cada miembro de esta 2ª ecuación entre el miembro correspondiente de la primera ecuación y se sustituye otra vez

$$\hat{\alpha} = \frac{\bar{X}^2}{(1/n)\sum X_i^2 - \bar{X}^2} \qquad \hat{\beta} = \frac{(1/n)\sum X_i^2 - \bar{X}^2}{\bar{X}}$$

$$\frac{\frac{1}{n}\sum X_i^2}{\bar{X}} =$$

Scanned by CamScanner

EJEMPLO 6.14 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución binomial negativa generalizada con parámetros r y p. Como $E(X) = r(1-p)/p$ y $V(X) = r(1-p)/p^2$,
$E(X^2) = V(X) + [E(X)]^2$

$$= r(1-p)/p^2 + (r(1-p)/p)^2 = \frac{r(1-p)}{p^2} + \frac{r^2(1-p)^2}{p^2}$$

$$= \frac{r(1-p)(r-rp+1)}{p^2}.$$

Si se iguala $E(X)$ a $\bar{X}$ y $E(X^2)$ a $(1/n)\sum X_i^2$ a la larga se obtiene

$$\boxed{\bar{X} = r(1-p)/p} \qquad \boxed{\frac{1}{n}\sum X_i^2 = \frac{r(1-p)(r-rp+1)}{p^2}}$$

$$\hat{p} = \frac{\bar{X}}{r(1-p)} \qquad \bar{X} = \frac{r-rp}{p} = \frac{r}{p} - \frac{r\cancel{p}}{\cancel{p}} = \frac{1}{p}r - r = \bar{X}$$

$$\frac{1}{p}r = \bar{X} + r;$$

Scanned by CamScanner

ESTIMACIÓN DE MÁXIMA VEROSIMILITUD

EJEMPLO 6.15 Se obtuvo una muestra de diez cascos. El 1º, 3º y 10º están agrietados, en tanto que los demás no. Sea $p = P(\text{casco agrietado})$ y defina $X_1, \dots, X_{10}$ como $X_i = 1$ si el i-ésimo casco está agrietado y 0 lo contrario.

En este caso las x_i son 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, así que la función masa de probabilidad conjunta de la muestra es

$$f(x_1, \dots, x_{10}; p) = p^3(1-p)^7$$

Ahora se hace la pregunta: "¿con qué valor de p es más probable que la muestra observada haya ocurrido?". Es decir, se desea encontrar el valor de $\hat{p}$ que incrementa al máximo la función de masa de probabilidad o, en forma equivalente, que incrementa al máximo el logaritmo natural de la función.

Como

$$\ln[f(x_1, \dots, x_{10}; p)] = 3.\ln(p) + 7\ln(1-p)$$

la cual es una función DERIVABLE de p, igualando la derivada de esta función a 0 se obtiene el valor maximizante

$$\frac{d}{dp}\ln[f(x_1, \dots, x_{10}; p)] = \frac{3}{p} - \frac{7}{1-p} = 0 \Rightarrow p = \frac{3}{10} = \frac{x}{n}$$

donde x es el nro de éxitos registrados. La estimación de p es ahora $\hat{p} = 3/10$. Se llama estimación de máxima verosimilitud porque para $x_1, \dots, x_{10}$ establecido, es el valor del parámetro que maximiza la probabilidad (función de masa de probabilidad) de la muestra observada.

OBS! Obsérvese que si sólo se hubiera dicho que entre los 10 cascos había 3 agrietados, la ecuación de la función de probabilidad sería reemplazada por la función masa de probabilidad binomial

$$\binom{10}{3} p^3(1-p)^7,$$ la cual también se incrementa al máximo con $\hat{p} = \frac{3}{10}$

EJEMPLO 6.16 Suponga que $X_1, \dots, X_n$ es una muestra aleatoria de una distribución exponencial con parámetro λ. Debido a la independencia, la función de verosimilitud es un producto de las funciones de densidad de probabilidad individuales:

$$f(x_1, \dots, x_n; \lambda) = (\lambda e^{-\lambda x_1}) \dots (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \sum x_i}$$

El ln(verosimilitud) es

$$\ln[f(x_1, \dots, x_n; \lambda)] = \ln(\lambda^n e^{-\lambda \sum x_i}) = n.\ln(\lambda) - \lambda \sum x_i.$$

Si se iguala $(d/d\lambda)[\ln(\text{veros.})]$ a cero se obtiene

$$\frac{n}{\lambda} - \sum x_i = 0; \quad \lambda = n/\sum x_i = 1/\bar{X}.$$ Por consiguiente el ~~máximo~~ EMV es

$$\boxed{\hat{\lambda} = 1/\bar{X}};$$

$$\frac{1}{\lambda} = \frac{\sum x_i}{n}$$

$$\lambda = \frac{n}{\sum x_i}$$

OBS! no es insesgado puesto que $E(1/\bar{X}) \neq 1/E(\bar{X})$

húsares

Scanned by CamScanner

EJEMPLO 6.17: Sean $X_1, \dots, X_n$ una muestra aleatoria de una distribución normal. La función de probabilidad es

$$f(x_1, \dots, x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/2\sigma^2} \cdot \dots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/2\sigma^2} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\Sigma(x_i-\mu)^2/2\sigma^2}$$

por consiguiente

$$\ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2.$$

Para determinar los valores maximizantes de μ y σ^2 se deben tomar las derivadas parciales de $\ln(f)$ con respecto a μ y σ^2, igualarlas a cero y resolver las dos ecuaciones resultantes. Omitiendo detalles, los EMV resultantes son

$$\hat{\mu} = \bar{X} \quad \text{y} \quad \hat{\sigma}^2 = \frac{\Sigma(X_i-\bar{X})^2}{n}$$

. El EMV de σ^2 no es insesgado, por consiguiente dos principios diferentes de estimación dan dos estimadores diferentes.

EJEMPLO 6.18: Suponga que cuando el área de la región R que se está muestreando es $a(R)$, el número X de eventos que ocurren en R tiene una distribución de POISSON con parámetro $\lambda a(R)$ (donde λ es el número esperado de eventos por unidad de área) y que las regiones no traslapantes dan X independientes.

Suponga que un ecólogo selecciona n regiones no traslapantes $R_1, \dots, R_n$ y cuenta el nro de plantas de una especie de cada región. La función masa de probabilidad conjunta es entonces

$$p(x_1, \dots, x_n; \lambda) = \frac{[\lambda \cdot a(R_1)]^{x_1} e^{-\lambda \cdot a(R_1)}}{x_1!} \cdot \dots \cdot \frac{[\lambda a(R_n)]^{x_n} e^{-\lambda a(R_n)}}{x_n!}$$

$$= \frac{[a(R_1)]^{x_1} \cdot \dots \cdot [a(R_n)]^{x_n} \cdot \lambda^{\Sigma x_i} \cdot e^{-\lambda \Sigma a(R_i)}}{x_1! \cdot \dots \cdot x_n!}$$

El ln(verosimilitud) es $\ln[p(x_1, \dots, x_n; \lambda)] = \Sigma x_i \cdot \ln[a(R_i)] + (\Sigma x_i)\ln(\lambda) - \lambda\Sigma a(R_i) - \Sigma\ln(x_i!)$

Con ~~derivar~~ $d/d\lambda$ $\ln(p)$ e igualándola a cero da

$$\frac{\Sigma x_i}{\lambda} - \Sigma a(R_i) = 0 \text{ ; por consiguiente } \quad \lambda = \frac{\Sigma x_i}{\Sigma a(R_i)} \quad \text{y} \quad \hat{\lambda} = \frac{\Sigma X_i}{\Sigma a(R_i)}$$

Scanned by CamScanner

EJEMPLO 6.19 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una función de densidad de probabilidad WEIBULL

$$f(x;\alpha,\beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} \cdot x^{\alpha-1} \cdot e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Si se escribe la verosimilitud y el ln(verosimilitud) y luego con $\partial/\partial\alpha[\ln(f)] = 0$ y $\partial/\partial\beta[\ln(f)] = 0$ se obtienen las ecuaciones

$$\alpha = \left[\frac{\sum x_i^\alpha \cdot \ln(x_i)}{\sum x_i^\alpha} - \frac{\sum \ln(x_i)}{n}\right]^{-1} \quad y \quad \beta = \left(\frac{\sum x_i^\alpha}{n}\right)^{1/\alpha}$$

Estas dos ecuaciones no pueden ser resueltas explícitamente para obtener fórmulas generales de los estimadores de máxima verosimilitud $\hat{\alpha}$ y $\hat{\beta}$. En su lugar, por cada muestra $x_1, \ldots, x_n$, las ecuaciones deben ser resueltas con un procedimiento numérico iterativo.

PRINCIPIO DE INVARIANZA

EJEMPLO 6.20: (conti. 6.17) En el caso normal, los estimadores de máxima verosimilitud de μ y σ^2 son $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = \sum(x_i - \bar{x})^2/n$. Para obtener el EMV de $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$ sustituya los estimadores en la función

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum(X_i - \bar{X})^2\right]^{1/2}.$$

El EMV de σ NO es la desviación estándar muestral S, y se aproximan bastante cuando n es bastante pequeño.

EJEMPLO 6.21 El valor medio de una variable aleatoria X que tiene una distribución Weibull es $\mu = \beta \cdot \Gamma(1 + 1/\alpha)$.
El EMV de μ es, por consiguiente, $\hat{\mu} = \hat{\beta} \cdot \Gamma(1 + 1/\hat{\alpha})$ donde $\hat{\alpha}$ y $\hat{\beta}$ son los EMV de α y β.
En particular $\bar{X}$ no es el EMV de μ, aunque sí es insesgado. Por lo menos con n grande, $\hat{\mu}$ es un mejor estimador que $\bar{X}$.

INTERVALOS DE CONFIANZA

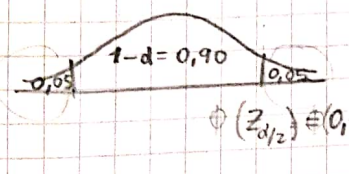
EJEMPLO 7.1: Las cantidades requeridas para calcular el intervalo de 95% de confianza para la altura preferida promedio verdadera son $\sigma = 2{,}0$, $n = 31$ y $\bar{x} = 80{,}0$.
El intervalo resultante es

$$\bar{x} \pm 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} = 80{,}0 \pm 1{,}96 \cdot \frac{2}{\sqrt{31}} = 80{,}0 \pm 0{,}7 = (79{,}3 \,;\, 80{,}7)$$

Es decir, $79{,}3 < \mu < 80{,}7$

Scanned by CamScanner

EJEMPLO 7.3: El proceso de producción de una caja de control de un tipo particular de motor fue modificado. Antes de esa modificación, datos históricos sugirieron que la distribución de diámetros de agujeros para bujes era normal con desviación estándar de 0,100 mm. Se cree que la modificación no ha afectado la forma de la distribución ni la desviación estándar, pero que el valor del diámetro puede haber cambiado. Se selecciona una muestra de 40 cajas y se determina el diámetro de agujero para cada una, y el resultado es un diámetro medio muestral de 5,426 mm.
Calcúlese un intervalo de confianza de 90%. Esto requiere que $100(1-\alpha) = 90$, de donde $\alpha = 0,10$ y $Z_{\alpha/2} = Z_{0,05} = 1,645$ (correspondiente a un área de curva z acumulativa de 0,9500).
El intervalo deseado entonces es



$$5,426 \pm (1,645)\frac{100}{\sqrt{40}} = 5,426 \pm 0,026$$
$$= (5,400,\ 5,452)$$

Este intervalo es angosto debido a la pequeña cantidad de variabilidad del diámetro del agujero ($\sigma = 0,100$).

NIVEL DE CONFIANZA, PRECISIÓN Y TAMAÑO DE MUESTRA.

EJEMPLO 7.4: El tiempo de respuesta a un comando de edición particular está normalmente distribuido con desviación estándar de 25 ms. Se instaló un nuevo SO y se desea estimar el tiempo de respuesta promedio verdadero μ en el nuevo entorno. Suponiendo que los tiempos de respuesta siguen estando normalmente distribuidos con $\sigma = 25$. ¿Qué tamaño de muestra es necesario para asegurarse de que el intervalo de confianza de 95% resultante tiene un ancho de, cuando mucho, 10.

El tamaño de la muestra debe satisfacer que
$10 = 2.(1,96)(25/\sqrt{n})$; $10 = 98/\sqrt{n}$; $\frac{1}{\sqrt{n}} = 10/98$; $\sqrt{n} = 98/10$;

$\cancel{n = \sqrt{9,8}}$ $n = 9,8^2 = 96,04$.
En vista de que n debe ser un entero, se requiere un tamaño de muestra de 97.

EJEMPLO 7.5: Un modelo teórico sugiere que el tiempo hasta la ruptura de un fluido aislante entre electrodos a un voltaje particular tiene una distribución exponencial con parámetro λ.
Una muestra aleatoria de $n = 10$ tiempos de ruptura da los siguientes datos muestrales: (en min)
$x_1 = 41,53$; $x_2 = 18,73$; $x_3 = 2,99$; $x_4 = 30,34$; $x_5 = 12,33$; $x_6 = 117,52$; $x_7 = 73,02$; $x_8 = 223,63$; $x_9 = 4,00$; $x_{10} = 26,78$. Se desea un intervalo de confianza del 95% para λ y para el tiempo de ruptura promedio verdadero.

Sea $h(x_1, \dots, x_n; \lambda) = 2\lambda\sum x_i$. Se puede demostrar que esta variable aleatoria tiene una distribución de probabilidad llamada JI CUADRADA (con 2n grados de libertad (gl)) ($v = 2n$ donde v es el parámetro de una distr. Ji cuadrada). ~~La tabla~~
El número pertinente de grados de libertad en este caso es $2(10) = 20$. La fila $v = 20$ de la tabla muestra que 34,170 captura un área de cola superior de 0,025, y 9,591 captura un área de cola inferior de 0,025 (área de cola superior de 9,591)
Por consiguiente, para $n = 10$,

$$P(9,591 < 2\lambda\sum x_i < 34,170) = 0,95;\ P\left(\frac{9,591}{2\sum x_i} < \lambda < \frac{34,170}{2\sum x_i}\right) = 0,95;$$

como $2\sum x_i = 550,87 \cdot 2 = 1101,74$, da el intervalo $(0,00871,\ 0,03101$

Scanned by CamScanner

EJEMPLO 7.6 El voltaje de ruptura de corriente alterna (CA) de un líquido aislante indica su resistencia eléctrica. Se dan las observaciones muestrales adjuntas de voltaje de ruptura (kV) de un circuito particular en ciertas condiciones.
(...)
Una gráfica de caja de los datos muestra una alta concentración a la mitad de la parte media de los datos (ancho de caja angosto). Hay un sólo valor apartado en el extremo superior, pero éste está en realidad un poco más cerca de la mediana (55), que la observación muestral más pequeña.

Las cantidades resumidas incluyen $n=48$, $\sum x_i = 2626$ y $\sum x_i^2 = 144950$, a partir de las cuales $\bar{x} = 54,7$ y $s = 5,23$. El intervalo de confianza de 95% es entonces

$$54,7 \pm 1,96 \cdot \frac{5,23}{\sqrt{48}} = 54,7 \pm 1,5 = (53,2 ; 56,2).$$

Es decir, $53,2 < \mu < 56,2$

EJEMPLO 7.7 Remítase al ejemplo 7.6 sobre voltaje de ruptura. Suponga que el investigador ~~que~~ cree que virtualmente todos los valores de la población se encuentran entre 40 y 70. Entonces $(70-40)/4 = 7,5$ da un valor razonable para s. El tamaño de muestra apropiado para estimar el voltaje de ruptura promedio verdadero a dentro de 1 kV con nivel de confianza de 95%, es decir, para que el intervalo de confianza de 95% tenga un ancho de 2 kV, es

$$n = [(1,96)(7,5)/1]^2 \approx 217$$

INTERVALOS UNILATERALES

EJEMPLO 7.10 Una muestra de 48 observaciones de resistencia al esfuerzo cortante dio una resistencia media muestral de 17,17 N/mm² y una desviación estándar muestral de 3,28 N/mm². Un límite de confianza inferior para ~~la~~ μ con nivel de confianza de 95% es

$$17,17 - (1,645)\frac{3,28}{\sqrt{48}} = 17,17 - \cancel{1,645}\ 0,78 = 16,39.$$

Es decir, con un nivel de confianza de 95% el valor de μ queda en el intervalo $(16,39 ; \infty)$

DISTRIBUCIÓN t.

EJEMPLO 7.11 Considérese las observaciones de módulo de elasticidad (MPa) obtenidas un minuto después de cargar la configuración: (ver libro).

???

Scanned by CamScanner

INTERVALO DE PREDICCIÓN

EJEMPLO 7.12 Considere la siguiente muestra de contenido de grasa (porcentaje) de n=10 perros ~~calleteres~~ caliente seleccionados al azar: 25,2 – 21,3 – 22,8 – 17,0 – 29,8 – 21,0 – 25,5 – 16,0 – 20,9 – 19,5
Supongamos que la distr. de la población es normal, un intervalo de confianza de 95% para (la estimación del intervalo de) el contenido de grasa medio de la población es

$$\bar{x} \pm t_{0,025,9} \cdot \frac{s}{\sqrt{n}} = 21,90 \pm 2,262 \cdot \frac{4,134}{\sqrt{10}} = 21,90 \pm 2,96$$

$$= (18,94 ; 24,86)$$

$\alpha = 0,5$
$\nu = n-1 = 10-1 = 9$

Suponga, sin embargo, que se va a comer un sólo perro caliente de este tipo y desea PREDECIR el contenido de grasa resultante. Una predicción PUNTUAL, análoga a una estimación puntual, es simplemente $\bar{x} = 21,90$.

EJEMPLO 7.13 Con $n=10$, $\bar{x} = 21,90$, $s = 4,134$ y $t_{0,025,9} = 2,262$, un intervalo de predicción de 95% para el contenido de grasa de un sólo perro caliente es

$$21,90 \pm (2,262)(4,134)\sqrt{1+\frac{1}{10}} = 21,9 \pm 2,262 \cdot 4,134 \cdot 1,05 = 21,9 \pm 9,81$$

$$= (12,09 ; 31,71)$$

El intervalo es bastante ancho, lo que indica una incertidumbre sustancial en cuanto al contenido de grasa. Obsérvese que el ancho del intervalo de predicción es más de tres veces ~~de~~ el del intervalo de confianza.

EJEMPLO 7.14

Scanned by CamScanner

EJEMPLO 7.15 Los datos adjuntos sobre voltaje de ruptura de circuitos eléctricamente sobrecargados se tomaron de un diagrama de probabilidad normal. La linealidad del diagrama se apoyó de manera firme la suposición de que el voltaje de ruptura está aproximadamente distribuído en forma normal:

1470 - 1510 - 1690 - 1740 - 1900 - 2000 - 2030 - 2100 - 2190 - 2200 - 2290 - 2380 - 2390
2480 - 2500 - 2580 - 2700.

Sea σ^2 la varianza de la distribución del voltaje de ruptura. El valor calculado de la VARIANZA MUESTRAL es $s^2 = 137324,3$, la estimación puntual de σ^2.

Con grados de libertad $= n-1 = 16$, un intervalo de confianza de 95% requiere $\chi^2_{0,975,16} = 6,908$ y $\chi^2_{0,025,16} = 28,845$. El intervalo es

$$\left(\frac{16 \cdot 137324,3}{28,845}, \frac{16 \cdot 137324,3}{6,908}\right) = (76172,3 ; 318064,4)$$

Tomando la raíz cuadrada de cada punto extremo se obtiene (276,0; 564,0) como el intervalo de confianza de 95% para σ. Estos intervalos son bastante anchos, lo que refleja la variabilidad sustancial del voltaje de ruptura en combinación con un tamaño de muestra pequeño.

Scanned by CamScanner

	Parte A				Parte B		
	Ejercicio 1	Ejercicio 2	Ejercicio 3		Ejercicio 4	Ejercicio 5	Ejercicio 6
Total A				Total B			

Justifique claramente los resultados obtenidos.

Parte A

Ejercicio 1: Según el Código Alimentario Argentino (CAA) un alimento es considerado bajo en sodio si posee a lo sumo 120 mg por cada 100 gr de producto. Se seleccionaron aleatoriamente muestras de galletas rotuladas bajas en sodio para cierta marca para determinar el cumplimiento de las normas del CAA. Los resultados de sodio (en mg) obtenidos por cada 100 gr de galleta fueron:

118,96	119,50	120,58	118,49	118,40	120,99
119,50	121,56	118,79	120,23	123,51	

a) Calcular el promedio muestral, desvío estándar muestral, primer y tercer cuartil muestral, rango intercuartil muestral, mediana muestral y varianza muestral. 0.5
b) Clasificar como medida de posición o medida de variabilidad a cada uno de los valores hallados en el ítem a). 0.2
c) Determinar si hay datos atípicos en la muestra y justificar. 0.3

Ejercicio 2: Sea X una variable aleatoria con distribución acumulada dada por:

$$F(x) = \begin{cases} 0 & si\, x < -1 \\ a & si -1 \leq x < 0{,}5 \\ \frac{3}{4} & si\, 0{,}5 \leq x < 1 \\ b & si\, x \geq 1 \end{cases}$$

a) Determinar los valores de las constantes a y b siendo $E(X) = -1/8$. Clasificar la variable aleatoria como discreta o continua. Justifique sus respuestas.
b) Determinar la función de probabilidad de X.
c) Calcular la varianza de X.
d) Si $W = -5\,X^2 + 15$ determinar el valor esperado de W.

Ejercicio 3: Se sabe que la longitud del cuerpo de una especie de insecto tiene distribución normal. La longitud para insectos que habitan en un cultivo A (X) tiene un valor medio y varianza de 15 mm y 18 mm² mientras que en un cultivo B (Y) tiene un valor medio y varianza de 8 mm y 20 mm².

a) ¿Cuál es la probabilidad que la longitud de un insecto que habita en el cultivo B se encuentre comprendido entre 6 y 14 mm?

2.865

Scanned by CamScanner

b) ¿Cuál es la probabilidad que la longitud de un insecto que habita en el cultivo A sea mayor que la longitud de un insecto que habita en el cultivo B?

c) Se tomaron dos muestras aleatorias de tamaño 9 en cada uno de los cultivos A y B.

   i) ¿Cuál es la probabilidad que exactamente cinco de los nueve insectos que habitan en el cultivo A tenga una longitud de por lo menos 20 mm?

   ii) ¿Cuál es la probabilidad que el promedio de las longitudes para los insectos que habitan en el cultivo A sea mayor que el promedio de las longitudes para los insectos que habitan en el cultivo B?

Parte B

Ejercicio 4: Usando los datos del Ejercicio 1 y suponiendo que la variable cantidad de sodio, por cada 100gr ~~de~~ galleta de la marca elegida, tiene distribución normal con media y varianza μ y σ^2 respectivamente.

a) Dar las estimaciones por Máxima Verosimilitud para μ, σ^2 y para el percentil 2%. 0.25

b) Dar un intervalo de confianza del 99% para μ. 0.25

c) Dar un intervalo de confianza del 95% para σ^2. 0.25

d) ¿Existe evidencia suficiente para decir que esta marca de galletas no cumple con las normas del CAA? Plantear las hipótesis pertinentes, dar la región de rechazo al 5% y concluir en el contexto del problema. 0.25

Ejercicio 5: Un comercio desea tener una idea de cuál es la verdadera proporción de clientes a favor de la ampliación del horario de atención al público (p). Se seleccionó una muestra aleatoria de 360 clientes, de los cuales 290 estaban a favor de la ampliación del horario de atención.

a) Obtenga un intervalo de confianza del 98% para p.

b) Si el comerciante resuelve ampliar el horario de atención si p es superior a 0,75.

   i)¿Existe evidencia suficiente para decidir la ampliación del horario de atención? Plantear las hipótesis pertinentes, dar la región de rechazo al 2% y concluir en el contexto del problema.

   ii)Calcular el p valor para las hipótesis planteadas y usando el valor hallado tomar una decisión al 1%.

Ejercicio 6: Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria con función densidad dada por

$$f(x) = \begin{cases} \frac{1}{2}(1 + x\,\theta) & si -1 \leq x \leq 1 \\ 0 & en\ caso\ contrario \end{cases}$$

para $-1 \leq \theta \leq 1$.

a) Hallar el estimador por el método de los momentos para θ y determinar si es insesgado.

b) Obtenga el desvío estándar para el estimador obtenido en el ítem a).

Patricia 351 - 740 - 0829 0.005

of. 311 13 - 14.30

Scanned by CamScanner

EXAMEN FINAL PROB y EST (I)

PARTE A

EJERCICIO 1: Según el código alimentario argentino (CAA) un alimento es considerado bajo en sodio si posee a lo sumo 120 mg por cada 100 gr de producto. Se seleccionaron aleatoriamente muestras de galletas rotuladas bajas en sodio para cierta marca para determinar el cumplimiento de las normas del CAA. Los resultados de sodio (en mg) obtenidos por cada 100 gr de galleta fueron: (...)

(A) Calcular el promedio muestral, desvío estándar muestral, primer y tercer cuartil muestral, rango intercuartil muestral, mediana muestral y varianza muestral.

Sea $X_1, \dots, X_n$ una muestra aleatoria $n = 11$

PROMEDIO MUESTRAL: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1320,51}{11} = 120,046 \approx 120,04$

DESVÍO ESTÁNDAR MUESTRAL: $\sigma = \sqrt{\frac{\left(\sum_{i=1}^{n} x_i^2\right) - n.\bar{x}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - 11.(120,04)^2}{10}} = \sqrt{\frac{158546,41 - 158505,62}{10}}$

$\boxed{\sigma = \sqrt{\frac{40,79}{10}} \approx 2,02}$ ⇒ VARIANZA $\boxed{\sigma^2 = 4,079}$

DATOS ORDENADOS: 118,4 - 118,49 - 118,79 - 118,96 - 119,5 - 119,5 - 120,23 - 120,58 - 120,99 - 121,56 - 123,51

$[Q_1 = X_{11(0,25)} \approx X_3 = 118,79]$

$[Q_3 = X_{11(0,75)} \approx X_8 = 120,58]$

$[IRQ = Q_3 - Q_1 = 120,58 - 118,79 = 1,79]$

MEDIANA MUESTRAL: $\tilde{X} = X_{(n/2)+1} = X_6 = 119,5$

(n/2) → división entera

(B) Clasificar como medida de posición o medida de variabilidad a los ítems anteriores.

DE VARIABILIDAD: Miden rangos y dispersiones entre los datos.
σ, σ^2 y RIC.

DE POSICIÓN →

- DE TENDENCIA CENTRAL: Marcan puntos donde los datos tienden a agruparse: $\bar{X}$, $\tilde{X}$
- DE POSICIÓN RELATIVA: Dividen los datos ordenados en cuatro partes
  CUARTILES → Q_1, Q_3

(C) Determinar si hay datos atípicos en la muestra y justificar

$x < Q_1 - 1,5.RIC = 118,79 - 1,5.(1,79) = 116,105$ no existe tal x

$x > Q_3 + 1,5.RIC = 123,26$ sí existe tal $x = x_{11} = 123,51$

(ATÍPICOS LEVES - NO HAY ATÍPICOS EXTREMOS)

Scanned by CamScanner

EJERCICIO 2: Sea una variable X aleatoria con distribución acumulada ~~por~~: dada por:

$$F(x) = \begin{cases} 0 & \text{si } x < -1 \\ a & \text{si } -1 \leq x < 0{,}5 \\ 3/4 & \text{si } 0{,}5 \leq x < 1 \\ b & \text{si } x \geq 1 \end{cases}$$

(A) Determinar los valores de las constantes a y b siendo $E(X) = -1/8$. Clasificar la variable aleatoria como discreta o continua. Justifique sus respuestas

La variable aleatoria X es discreta ya que su función de probabilidad está definida sólo para los valores discretos ~~de~~ -1, $0{,}5$ y 1, demostrado por su función de distribución acumulada definida por casos en vez de ser ella integral de la función de probabilidad.

Sabemos que $\boxed{b=1}$ ya que la suma de todas las probabilidades, en este caso $F(1)$, debe dar 1.

Ahora, $E(X) = a(-1) + (3/4 - a).(0{,}5) + (1 - 3/4).(1) = -1/8$
ya que $E(X) = \sum_{x \in X} x.p(x)$.

$\Rightarrow -a + 0{,}375 - a.0{,}5 + 1 - 0{,}75 = -0{,}125$;
$-1{,}5.a + 0{,}625 = -0{,}125$; $-1{,}5a = -0{,}75$; $\boxed{a = 0{,}5}$

(B) Determinar la función de probabilidad de X.

x	-1	0,5	1
f(x)	0,5	0,25	0,25

(C) Calcular la varianza de X.

$V(X) = E(X^2) - [E(X)]^2$; $E(X^2) = 0{,}5.(-1)^2 + 0{,}25.(0{,}5)^2 + 0{,}25(1)^2 = 0{,}5 + 0{,}0625 + 0{,}25$
$= 0{,}8125$

$V(X) = 0{,}8125 - 0{,}125 = 0{,}6875$

(D) Si $W = -5X^2 + 15$ determinar el valor esperado de W

$E(W) = E(-5X^2 + 15) = -5E(X^2) + 15 = -5(0{,}8125) + 15 = 10{,}9375$

EJERCICIO 3: Se sabe que la longitud del cuerpo de una especie de insecto tiene distribución normal. La longitud para insectos que habitan en un cultivo A (X) tiene un valor medio de 15 mm y varianza de 18 mm², mientras que en un cultivo B (Y) tiene un valor medio de 8 mm y 20 mm².

(A) ¿Cuál es la probabilidad de que la longitud de un insecto del cultivo B se encuentre comprendido entre 6 y 14 mm?

$Y \sim N(\mu = 8\text{mm}, \sigma^2 = 20\text{mm}^2)$

$\Rightarrow P(6 \leq Y \leq 14) = \{\text{estandarizamos } Y\} = P\left(\frac{6-8}{\sqrt{20}} \leq Z \leq \frac{14-8}{\sqrt{20}}\right)$

$= P(-0{,}447 \leq Z \leq 1{,}34) \approx P(-0{,}45 \leq Z \leq 1{,}34) = \Phi(1{,}34) - \Phi(-0{,}45)$

$= \Phi(1{,}34) - (1 - \Phi(0{,}45)) = 0{,}9099 - 1 + 0{,}6736 = \underline{0{,}5835}$;

Scanned by CamScanner

Ⓑ ¿Cuál es la probabilidad de que la longitud de un insecto del cultivo A sea mayor que la longitud de uno que habita en el cultivo B?

$P(Y < X) =$

Ⓒ Se tomaron dos muestras aleatorias de tamaño 9 en cada uno de los cultivos A y B

I) ¿Cuál es la probabilidad de exactamente 5 de los 9 insectos del cultivo A tengan una longitud de por lo menos 20 mm?

E = longitud de por lo menos 20 mm $\qquad X' =$ "cantidad de éxitos" $\sim bin(n, P(E))$

$$p = P(E) = P(X \geq 20) = 1 - P(X \leq 20) = 1 - P\left(Z \leq \frac{20-15}{\sqrt{18}}\right) = 1 - P(1,18) = 1 - 0,8810$$

$= 0,119$

$$\Rightarrow P(X'=5) = \binom{9}{5} p^5 (1-p)^{9-5} = \binom{9}{5}(0,119)^5(0,881)^4 = 0,00181$$

II) ¿Cuál es la probabilidad de que ~~exactamente~~ el promedio de las longitudes para los insectos del cultivo A sea mayor que el promedio de las longitudes para los del cultivo B?

PARTE B:

EJERCICIO 4: Usando los datos del EJERCICIO 1, y suponiendo que la variable cantidad de sodio por cada 100 gr de galleta de la marca elegida, tiene distribución normal con media μ y varianza σ^2

A) Dar las estimaciones por Máxima Verosimilitud para μ y σ^2, y para el percentil 2%.

Está demostrado que los estimadores de Máxima Verosimilitud para μ y σ^2 de una variable distribuida normalmente son el promedio muestral ($\bar{X}$) y la desviación muestral S^2

Por lo tanto $\hat{\mu} = \bar{X} = 120,04$

$\hat{\sigma}^2 = S^2 = 4,079$

Con respecto al percentil $\eta(p)$, llamémosle a la variable aleatoria de la consigna X, por propiedad de la Invarianza:

$$\boxed{\eta_X(p) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}} \cdot \eta_Z(p) + \bar{x}}$$

mo — 11

2 — 0,22

$\eta_Z(2) \Rightarrow \Phi(\eta_Z(2)) = 0,22 \Rightarrow$

$\Rightarrow \eta_Z(2) = -0,775$

$$\eta_X(2) = S \cdot \eta_Z(2) + \bar{x} = 2,02 \cdot (-0,775) + 120,04 = 118,47$$

Scanned by CamScanner

B) Dar un intervalo de confianza del 99% para μ

nivel de confianza del 99% $\Rightarrow 1-\alpha = 0{,}99$; $\alpha = 0{,}01$

$Z_{\alpha/2} = Z_{0{,}01/2} = Z_{0{,}005} \Rightarrow \Phi(Z_{0{,}005}) = 0{,}99 \Rightarrow Z_{0{,}005} = 2{,}33$

Intervalo de confianza para μ del 99%: $\bar{x} \pm Z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}} = 120{,}04 \pm 2{,}33 \cdot \frac{2{,}02}{\sqrt{11}} \Rightarrow$

$\Rightarrow (118{,}62 < \mu < 121{,}46)$

C) Dar un intervalo de confianza para σ^2: del 95%

IC para $\sigma^2 = \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2, n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2, n-1}} \right)$; IC del 95% $\Rightarrow 1-\alpha = 0{,}95$;

$\alpha = 0{,}05$

$\alpha/2 = 0{,}025$

$\Rightarrow$ IC para σ^2 del 95% $= \left(\frac{(10)(4{,}079)}{\chi^2_{0{,}025, 10}}, \frac{(10)(4{,}079)}{\chi^2_{0{,}975, 10}} \right) = \left(\frac{40{,}79}{20{,}483}, \frac{40{,}79}{3{,}247} \right)$

$= (1{,}991 , 12{,}56)$

d) ¿Es evidencia suficiente para decir que esta marca de galletas no cumple con las normas del CAA?

Plantear las hipótesis pertinentes, dar la región de rechazo del 5% y concluir con el contexto del problema.

Nuestro parámetro de interés será la cantidad media de sodio en las muestras por cada 100 gr de galleta. es decir μ

$H_0: \mu = 120$ $\quad H_a: \mu > 120$

Como la distribución de la población es normal pero la n es pequeña y la D.E es desconocida, usaremos el estadístico t

$$T = \frac{(\bar{x} - \mu_0)}{s/\sqrt{n}} \sim t_{n-1}$$

$$T_{obs} = \frac{120{,}04 - 120}{2{,}02/\sqrt{11}} = \frac{0{,}04}{0{,}61} = 0{,}065$$

$\boxed{R.R: \{t : t \geq t_{\alpha; n-1}\}}$ nivel $\alpha = 0{,}05$ (5% de significancia)

$\Rightarrow \{t : t \geq t_{0{,}05, 10}\} = \{t : t \geq 3{,}169\}$ $\quad t_{obs} < 3{,}169 \Rightarrow$

Con este nivel de significancia no hay evidencia suficiente para rechazar la hipótesis inicial, que es que dicha marca cumple con las normas del CAA.

húsares

Scanned by CamScanner

EJERCICIO 5: Un comercio desea tener una idea de cuál es la verdadera proporción de clientes a favor de la ampliación del horario de atención al público (p). Se seleccionó una muestra aleatoria de 360 clientes, de los cuales 290 estaban a favor de la ampliación del horario de atención.

(A) Obtenga un intervalo de confianza del 98% para p:

Primero comprobamos que el tamaño de la muestra es suficiente:

$n.\hat{p} \geq 10$ & $n.\hat{q} \geq 10$

$\hat{p} = \frac{X}{n}$ donde X = cantidad de clientes a favor = 290

$= \frac{290}{360} = 0,805 \Rightarrow \hat{q} = (1-\hat{p}) = 0,195$;

$n.\hat{p} = 360 . 0,805 = 289,8 > 10$ y $n.\hat{q} = 360 . 0,195 = 70,2 > 10$

Entonces, un intervalo de confianza para p es

$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}.\hat{q}}{n}}$, donde $1-\alpha = 0,98 \Rightarrow \alpha = 0,02 \Rightarrow \alpha/2 = 0,01$,

entonces

$IC_{98\%} = \left(0,805 \pm Z_{0,01} \sqrt{\frac{0,16}{360}}\right)$; $\Phi(Z_{0,01}) = 0,99 \Rightarrow Z_{0,01} = 2,33$;

$= 0,805 \pm 2,33 . 0,02 = 0,805 \pm 0,046$

$\Rightarrow (0,759 ; 0,851) = IC_{98\%}$ para p.

(B) Si el comerciante resuelve ampliar el horario de atención si p es superior a 0,75

(I) ¿Existe evidencia suficiente para ampliar el horario?

Plantear las hipótesis pertinentes, dar la región de rechazo al 2% y concluir en el contexto del problema.

- PARÁMETRO DE INTERÉS: p

- HIPÓTESIS: $H_0: p = 0,75$ contra $H_a: p > 0,75$

- Estadístico: Como la muestra es lo suficientemente grande, podemos utilizar un estadístico con distribución normal

$$Z = \frac{(\hat{p} - p_0)}{\sqrt{p_0 . q_0}/\sqrt{n}} = \frac{(0,805 - 0,75)}{\sqrt{(0,75)(0,25)/360}} = \frac{0,055}{0,023} = 2,39$$

Región de Rechazo al 2% $\Rightarrow \alpha = 0,02$

$RR = \{z \mid Z \geq Z_{0,02}\}$; $\Phi(Z_{0,02}) = 0,98 \Rightarrow Z_{0,02} = 2,055$

$= \{z \mid Z \geq 2,055\}$ Como nuestro Z observado $= 2,39$ es mayor que $2,055$, está dentro de la RR y nos permite concluir que la proporción de clientes a favor del cambio de horario es mayor al 75% y debería de ampliarse

Scanned by CamScanner

II) Calcular el p valor para las hipótesis planteadas y usando el valor hallado tomar una decisión al 1%.

P-valor para nuestro Z observado $Z_{obs} = 2{,}39$
para una prueba de cola SUPERIOR

$$p\text{-valor} = 1 - \Phi(2{,}39) = 1 - 0{,}9916 = 0{,}0084$$

Si p-valor $\leq \alpha$ entonces H_0 puede ser rechazada; $\alpha = 0{,}01$

$0{,}0084 \leq 0{,}01 \Rightarrow$ H_0 puede ser rechaza también a este nivel de significancia

EJERCICIO 6: Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria con función de densidad dada por

$$f(x) = \begin{cases} \frac{1}{2}(1 + x\theta) & \text{si } -1 \leq x \leq 1 \\ 0 & \text{c.c} \end{cases} \quad \text{para } -1 \leq \theta \leq 1.$$

a) Hallar el estimador por el Método de Momentos para θ y determinar si es insesgado.

MÉTODO DE MOMENTOS $\quad m =$ parámetros a estimar $= 1$ ("θ")

Como hay que calcular los $m = 1$ momentos, sólo tenemos que calcular

$$E(X) = \int_{-1}^{1} \frac{1}{2}(1 + x\theta) \cdot x \, dx = \frac{1}{2}\int_{-1}^{1} x + x^2\theta \, dx = \frac{1}{2}\left(\int_{-1}^{1} x \, dx + \int_{-1}^{1} x^2\theta \, dx\right)$$

$$= \frac{1}{2}\left(\frac{x^2}{2}\Big|_{-1}^{1} + \frac{x^3}{3}\theta\Big|_{-1}^{1}\right) = \frac{1}{2}\left(\frac{1}{3}\theta + \frac{1}{3}\theta\right) = \frac{1}{2} \cdot \frac{2}{3}\theta = \frac{\theta}{3}$$

$$\Rightarrow \boxed{\hat{\theta} = 3X}$$

$\hat{\theta}$ será insesgado si $E(\hat{\theta}) = \theta$; $E(\hat{\theta}) = E(3X) = \int_{-1}^{1} \frac{1}{2}(1 + 3x\theta)x \, dx$

$$= \frac{1}{2}\int_{-1}^{1}(x + 3x^2\theta)dx = \frac{1}{2}\left(\int_{-1}^{1} x \, dx + 3\theta\int_{-1}^{1} x^2 dx\right) = \frac{1}{2}\left(\frac{x^2}{2}\Big|_{-1}^{1} + 3\theta\left(\frac{x^3}{3}\Big|_{1}^{-1}\right)\right)$$

$$= \frac{1}{2} \cdot 3 \cdot \theta\left(\frac{2}{3}\right) = \theta$$ SÍ ES INSESGADO

b) obtenga el desvío estándar para el estimador obtenido en el ítem a)

$$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})} \; ; \quad V(\hat{\theta}) = E(\hat{\theta}^2) - [E(\hat{\theta})]^2 ;$$

$$E(\hat{\theta}^2) = E(9x^2) = \int_{-1}^{1} \frac{1}{2}(1 + 9x^2\theta)x \, dx = \frac{1}{2}\int_{-1}^{1} x + 9x^3\theta \, dx = \frac{1}{2}\left(\frac{x^2}{2}\Big|_{-1}^{1} + 9\theta\frac{x^4}{4}\Big|_{-1}^{1}\right)$$

$= 0$

$$V(\hat{\theta}) = 0 - [9x^2] = -9x^2$$

??

$$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})} = \sqrt{-9x^2} = 3x$$

húsares

Scanned by CamScanner

1	
2	
3	
4	
5	
6	

Probabilidad y Estadística
Profesorados y Licenciatura en Computación
Examen 04/12/2013

Apellido y Nombre:
Carrera:

Problema 1: Se seleccionaron 18 detectores de Radón de cierto tipo y cada uno de ellos se expuso a 100 pCi/L de Radón. Las lecturas de Radón obtenidas con estos detectores fueron:

101,05	100,10	99,09	100,19	99,56	100,96	99,72	99,76	100,85
99,89	98,98	100,15	99,11	100,48	99,81	99,61	101,02	98,90

a) Obtener dos medidas de posición o tendencia central y dos de dispersión o variabilidad para estos datos.
b) Realizar el gráfico de caja para estos datos, ¿hay datos atípicos o anómalos? Justifique su respuesta.

Problema 2: En la producción de cierto artículo se pueden presentar sólo dos tipos de defectos (A y B). Se sabe que el defecto tipo A ocurre en un 5 % de los artículos; el defecto tipo B se presenta en un 3 % de los artículos y en un 1 % ocurren ambos defectos.
a) Calcular la probabilidad de que un artículo tomado al azar de la producción:
i) presente sólo el defecto tipo A.
ii) presente al menos un defecto.
iii) no presente ningún defecto.
b) Sea Y la variable que cuenta el número el número de defectos encontrados en el artículo seleccionado al azar.
i) Dar la función de distribución de probabilidad y la distribución acumulada para la variable Y.
ii) Calcular el valor esperado de $X = 2 - Y^2$.

Problema 3: Una unidad de radar es utilizada para medir la velocidad de los automóviles en una vía rápida, durante la hora de mayor congestionamiento. La velocidad de los automóviles está distribuída normalmente con una media y desviación estándar poblacional de $\mu = 100$ y $\sigma = 8{,}5$ kilómetros por hora, respectivamente.
a) ¿Cuál es la probabilidad que un automóvil elegido al azar viaje a una velocidad de a lo sumo 88 kilómetros por hora?
b) ¿Cuál es la probabilidad que un automóvil elegido al azar viaje a una velocidad comprendida entre 88 y 110 kilómetros por hora?
c) ¿Cuál es la probabilidad de que al menos uno de diez automóviles, elegidos al azar, viaje a una velocidad mayor de 88 kilómetros por hora?

Problema 4: Considerando los datos del Problema 1 y suponiendo que la distribución de lecturas de Radón en este tipo de detectores es **normal**, responda los siguientes ítems.
a) Dar estimaciones por Máxima Verosimilitud para la media poblacional, varianza poblacional y percentil 5 % para la variable lecturas de Radón en este tipo de detectores.
b) Construya un intervalo de confianza del 99 % para la verdadera lectura media de Radón de ese tipo de detector.

Scanned by CamScanner

c) Construya un intervalo de confianza del 95 % para la varianza de la lectura de Radón de ese tipo de detector.

d) Según la información observada, ¿puede decirse que la lectura media de Radón para este tipo de detector difiere del valor al cual fueron expuestos los detectores? Para responder esta pregunta plantear las hipótesis pertinentes, dar la región de rechazo y concluir usando un nivel de significancia del 0,01.

Problema 5: En el diseño de mascarillas para bomberos es importante que resistan altas temperaturas debido a que por lo común trabajan en temperaturas de 200 a 500°F. En una prueba realizada en un conjunto de 120 mascarillas de cierta marca, a 48 se les desprendieron las lentes a una dada temperatura de referencia (250°F).

a) Construya un Intervalo de Confianza del 90 % para la proporción poblacional p de mascarillas de ese tipo, cuyas lentes se desprenderían a la temperatura de referencia.

b) Determinar el tamaño de muestra necesario para que un Intervalo de Confianza del 90 % para p tenga una longitud de a lo sumo la mitad de la obtenida en el ítem anterior, **independientemente** del valor de $\hat{p}$.

c) Se quiere determinar si existe evidencia suficiente para decir que la proporción poblacional p de mascarillas de ese tipo cuyas lentes se desprenderían a la temperatura de referencia, es menor a 0,5. Para ello:

i) Plantear las hipótesis para este problema.

ii) Establecer la región de rechazo para un nivel de significación del 5 % y tomar una decisión.

iii) Calcular el $p-$valor y usando este valor tomar una decisión considerando $\alpha = 0{,}01$.

Problema 6: Sea $X_1, X_2, ..., X_n$ una muestra aleatoria con $n \geq 3$ y distribución Poisson de parámetro λ.

a) Hallar el Estimador por el Método de los Momentos para λ.

b) Hallar el Estimador por el Método de Máxima Verosimilitud para λ.

c) Considerando los siguientes estimadores de λ

$$\widehat{\lambda}_1 = X_1 \ , \quad \widehat{\lambda}_2 = \frac{X_1 + X_n}{2} \ , \quad \widehat{\lambda}_3 = \frac{X_1 + 2\,X_2 + X_3}{3} \ , \quad \widehat{\lambda}_4 = \overline{X}.$$

i) ¿Cuáles de estos estimadores son insesgados para λ? Justifique su respuesta.

ii) Entre los estimadores insesgados encontrados, ¿cuál tiene menor varianza? Justifique su respuesta.

Ayuda: Si $X \sim P(\lambda)$ entonces su función de distribución de probabilidad está dada por

$$P(X = k) = e^{-\lambda}\,\frac{\lambda^k}{k!} \ , \text{ para todo } k \in N \cup \{0\}$$

Scanned by CamScanner

1	
2	
3	
4	
5	
6	

Probabilidad y Estadística

Profesorados y Licenciatura en Computación

Examen 18/12/2013

Apellido y Nombre:

Carrera:

Problema 1: Las máquinas M_1, M_2 y M_3 fabrican en serie determinado artículo. Las producciones de M_1, M_2 y M_3 son de 450, 450 y 600 artículos diarios y los porcentajes de defectuosos son del 2 %; 3,5 % y 2,5 % respectivamente. De la producción total de las tres máquinas se toma una pieza al azar al final del día.

a) ¿Cuál es la probabilidad de que el artículo haya sido producido por la máquina M_1?

b) ¿Cuál es la probabilidad de que sea defectuoso y haya sido producido por la máquina M_3?

c) ¿Cuál es la probabilidad de que el artículo sea defectuoso?

d) Si se sabe que el artículo seleccionado es defectuoso, ¿cuál es la probabilidad de que haya sido producido por la máquina M_2?

Problema 2: Una secretaria tiene dos líneas telefónicas A y B. El número de veces que recibe llamadas a las líneas A y B en una mañana son variables aleatorias independientes X_A y X_B con distribución de Poisson de parámetros $\lambda_A = 1$ y $\lambda_B = 2$ respectivamente.

a) Dar la distribución de la variable número total de llamadas que recibe en una mañana.

b) Dar el valor esperado y varianza del número total de llamadas que recibe en una mañana.

c) Se observan 250 mañanas y se registra el número de mañanas que la secretaria recibe sólo una llamada. Calcular la probabilidad aproximada de que el número de mañanas que la secretaria recibe sólo una llamada sea mayor a 50, usando el teorema central del límite.

Problema 3: La altura (X) de los hombres de una población tiene una distribución normal de media $\mu_X = 170\ cm$ y desvío estándar $\sigma_X = 7\ cm$ y la altura (Y) de las mujeres tienen una distribución normal de media $\mu_Y = 160\ cm$ y desvío estándar $\sigma_Y = 6\ cm$.

a) Se elige un hombre al azar de esta población, ¿cuál es la probabilidad de que su altura esté entre 168 y 173 cm?

b) ¿Para qué valor de c el intervalo $(160 - c\ ,\ 160 + c)$ incluye el 96 % de las alturas de las mujeres?

c) Se elige un hombre y una mujer al azar de esta población, ¿cuál es la probabilidad de que el hombre sea más alto que la mujer?

d) Si se eligen 20 mujeres al azar, ¿cuál es la probabilidad de que el promedio muestral de las alturas esté entre 158 y 160 cm?

Problema 4: Se propone un nuevo método para la determinación de concentraciones de fósforo en el suelo. Con este nuevo método se realizaron 20 mediciones de concentraciones de fósforo en un suelo cuya concentración verdadera es de 585 mg/Kg. Los datos obtenidos fueron:

579,4	570,5	583,4	598,3	586,8	582,6	593,4	584,0	594,5	605,6
581,0	591,6	585,4	576,7	570,7	582,5	594,6	594,3	575,7	588,4

Se puede suponer que la variable concentración de fósforo tiene una distribución normal.

a) Dar estimaciones por Máxima Verosimilitud para la media poblacional, varianza poblacional y percentil

Scanned by CamScanner

3 % para la variable concentración de fósforo con este nuevo método.

b) Hallar un intervalo de confianza del 99 % para la concentración media de fósforo en suelo, con el nuevo método.

c) ¿Hay evidencia suficiente para decir que la concentración promedio de fósforo obtenida con el nuevo método difiere significativamente del valor real? Plantear las hipótesis de interés, dar la región de rechazo al 5 %, calcular el valor observado del estadístico de prueba y concluir.

d) Acotar el p-valor de la prueba planteada en el ítem c) y concluir usando un $\alpha = 0,01$.

Problema 5: Antes de tomar la decisión de ampliar el horario de atención al público, una institución bancaria desea tener una idea de cuál sería la verdadera proporción p de clientes interesados en este cambio de horario. En una muestra aleatoria de 250 clientes, 190 estaban de acuerdo con la extensión del horario de atención.

a) Construya un Intervalo de Confianza del 90 % para la proporción poblacional p de clientes interesados en este cambio de horario.

b) Determinar el tamaño de muestra necesario para que un Intervalo de Confianza del 90 % para p tenga una longitud de a lo sumo la mitad de la obtenida en el ítem anterior, independientemente del valor de $\hat{p}$.

c) El directorio del banco resuelve extender el horario de atención si p es superior a 0,70.

i) ¿Existe evidencia suficiente para que el banco decida extender el horario de atención? Plantear las hipótesis pertinentes y tomar una decisión considerando $\alpha = 0,05$.

ii) Calcular el p-valor para la prueba planteada en el ítem (i) y usando este valor tomar una decisión al 1 %.

Problema 6: Si X tiene una distribución binomial de parámetros n y p, entonces X/n es un estimador insesgado de p. Para estimar la varianza de X, utilizamos en general $\hat{\theta} = n(X/n)(1 - X/n)$.

a) Probar que el estimador $\hat{\theta}$ no es un estimador insesgado para $V(X)$.

b) Modificar $\hat{\theta}$ levemente para obtener un estimador insesgado para $V(X)$.

Scanned by CamScanner