UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

50 Preguntas Teóricas

Estadística Aplicada

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura.. Grupo..............................

Apellidos .... Nombre..............................

Ejercicio del día ....

CUESTIONES DE ESTADÍSTICA TEÓRICA

1. Sea A, B y C tres sucesos incompatibles, con $P(A)=0,5$, $P(B)=0,25$, $P(C)=0,15$. Se pide:

a) $P(\bar{A}\cap\bar{B})$ b) $P(\bar{A}\cap\bar{B}\cap\bar{C})$ c) $P(\bar{A}\cap\bar{B}\cap C)$

d) Probabilidad de que exactamente se realice uno de los tres sucesos considerados.

2. La variación porcentual del PIB de 2017, con respecto al año anterior, es una variable aleatoria normal $N(2, 0,5)$. ¿Entre qué limites oscilará la variación del PIB con probabilidad 0,95?

3. Dadas tres variables aleatorias independientes, con características: $X_1 \sim N(0,2)$, $X_2 \sim N(2,2)$, $X_3 \sim N(4,1)$. Calcúlese $P(-7 \le Y \le 10)$, siendo $Y = 4X_1 + 5X_2 - 6X_3 + 6$

4. El peso (en gramos) de unas cajas es una variable aleatoria X, con distribución normal, cuya desviación típica es igual a 12 gramos. Se extraen muestras aleatorias simples de tamaño 100 cajas. ¿Cuál es la probabilidad de que el peso medio de las cajas supere el peso medio real en al menos 300 miligramos?

5. El número de llamadas independientes a la central de una agencia de viajes durante un día es una distribución de Poisson de parámetro $\lambda = 100$. ¿Cuál es la probabilidad de que durante 30 días, el total de llamadas se encuentre entre 2900 y 3100?

6. El peso de un paquete de cigarrillos se distribuye según una $N(80, 4)$. Se dispone de dos lotes de 100 paquetes cada uno de ellos. ¿Cuál será la probabilidad de que el peso del primer lote supere en más de 40 gramos al peso del segundo lote?

7. Sean A y B dos sucesos aleatorios tales que $P(A)=0,6$, $P(B)=0,4$ y $P(\bar{A}\cup\bar{B})=0,7$. Calcula:

a) $P(A\cap\bar{B})$ b) $P(A/B)$ c) $P(\overline{\bar{A}\cap B})$

8. Sea X una variable aleatoria que sigue una distribución χ^2 con 200 grados de libertad. Calcular: (a) $P(X \le 234)$ y (b) $P(X \le 183)$ sin recurrrir a las tablas de la Chi-cuadrado.

9. Dada una variable aleatoria X que se distribuye como una $B(30;\ 0,2)$ y otra variable aleatoria Y que se distribuye como una $B(150;\ 0,3)$.

a) ¿Si X e Y son independientes, se puede afirmar que la variable $X+Y$ se distribuye como una $B(180;\ 0,5)$? Justifica tu respuesta.

b) ¿Se puede afirmar que por el teorema de Moivre la variable Y tiende a una normal? Si fuera así, ¿qué parámetros tiene esa distribución normal?

Asignatura.. Grupo........................
Apellidos .... Nombre........................
Ejercicio del día ....

10. Los errores que se cometen al estimar el ahorro familiar de un país siguen una distribución normal de media 0 y desviación σ_e. Comprobar que la función $\frac{n\bar{e}^2}{\sigma_e^2}$ sigue una distribución Chi-cuadrado, sabiendo que el tamaño muestral es n y el error medio muestral es $\bar{e}$.

11. La probabilidad de que un paciente mejore tras administrarle un medicamento es de 0,82.
a) Se suministra de manera independiente el medicamento a 10 pacientes en la Clínica Salud. ¿Qué modelo de distribución de probabilidad sigue la variable número total de pacientes que mejoran en dicha clínica?
b) Si cada una de las 50 clínicas de una provincia realizan el mismo experimento de manera independiente, ¿Qué modelo de distribución de probabilidad sigue la variable número total de pacientes que mejoran en dicha provincia? ¿Cuál es la probabilidad de que al menos 400 pacientes mejoren en esa provincia?

12. Se ha realizado una muestra aleatoria simple (m.a.s) de tamaño 10 de una población considerada normal, llegando a la conclusión que la varianza muestal es 4. Calcular la probabilidad $P\left[\left|\bar{x}-\mu\right| \leq 1,22\right]$

13. Se desea dar una estimación por intervalos a un 95% de confianza para la proporción de alumnos que están a favor de aumentar en número máximo de convocatorias permitidas por la normativa de permanencia. ¿Qué tamaño de muestra habría que considerar si se desea un intervalo con una amplitud de 0,06?. Suponga que $p = q = 1/2$.

14. Sean A y B dos sucesos tales que $P(A) = 0,6$. Calcular $P(A \cap \bar{B})$ en cada caso:
a) A y B mutuamente excluyentes
b) A está contenido en B
c) B está contenido en A y $P(B) = 0,3$
d) $P(A \cap B) = 0,1$

15. Sean A y B dos sucesos incompatibles con $P(A \cup B) > 0$. Demostrar que

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

16. Las variables aleatorias en estudio son independientes.

Sean: $X_1 \sim N(10,2)$, $X_2 \sim N(15,2)$ y $X_3 \sim N(20,2)$ donde $Y = 5X_1 + 4X_2 - 3X_3$ y

$$U = \left(\frac{X_1 - 10}{2}\right)^2 + \left(\frac{X_2 - 15}{2}\right)^2 + \left(\frac{X_3 - 20}{2}\right)^2$$ **. Calcular: a) $P[20 \leq Y \leq 30]$ b) $P[U \leq 11]$**

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....
Apellidos .... Nombre ....
Ejercicio del día ....

(IMPRESO EN PAPEL RECICLADO) UNICAMENTE PARA USO ESCOLAR

17. El error cometido en expedir tiques por una maquina sigue una normal $N(0,\sigma)$.

a) ¿Cuál es la probabilidad de que el error cometido en valor absoluto de una medida cualquiera sea al menos σ?

b) ¿Cuánto valdría la probabilidad si se toma como medida la media aritmética de 10 medidas independientes?

18. En una población de mujeres, las puntuaciones de un test de ansiedad-riesgo siguen una distribución normal $N(25,10)$. Al clasificar la población en cuatro grupos de igual tamaño, ¿cuales serán las puntuaciones que delimiten estos grupos?.

19. El número de millones de metros cúbicos que tiene un embalse sigue una distribución normal $N(980, 50)$. El consumo diario de las poblaciones que sirve es una normal $N(85, 30)$. Si se sabe que durante una tormenta la cantidad de agua que se embolsa es una normal $N(50, 25)$. Un día han caído dos tormentas, calcular la probabilidad de que al final del día el agua embalsada sea menor o igual que 980 metros cúbicos.

20. Un dispositivo está formado por muchos elementos que trabajaban independientemente, siendo la probabilidad de fallo durante la primera hora de trabajo muy pequeña e iguales en todos los elementos. Si la probabilidad de que en ese tiempo falle por lo menos un elemento es 0,98.

a) Hallar la media y desviación típica del número de elementos que fallen en la primera hora.

b) Calcular la probabilidad de que fallen a lo sumo dos elementos en ese tiempo.

21. El departamento comercial de una industria alimenticia sabe que 2 de cada 10 consumidores reconocen su producto en una prueba a ciegas. ¿Cuántas pruebas a ciegas de sabor deberían hacerse para que la proporción de que los que conocen la marca oscile entre el 16% y el 24% con una probabilidad mínima de 0,8?

22. El p-valor en un contraste de hipótesis ha resultado ser de 0,08. Si el nivel de significación del contraste ha sido de 0.05 , ¿Se debería aceptar o rechazar la hipótesis nula?

23. Defina el Error de Tipo I y el Error de Tipo II. Analice la verdad o no de las siguientes afirmaciones y explique por qué:

a) P(Error tipo I) $+$ P(Error Tipo II) $=$ 1

b) Cuando disminuye la probabilidad de cometer un Error de Tipo I, también disminuye la probabilidad de cometer un Error de Tipo II.

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

UNICAMENTE PARA USO ESCOLAR (IMPRESO EN PAPEL RECICLADO)

24. Las variables aleatorias en estudio son independientes. Analiza si las afirmaciones son verdaderas o falsas:

a) X sigue una distribución binomial $B(1, 0{,}3)$ e Y una distribución binomial $B(1,\ 0{,}2)$ entonces X+Y sigue una binomial $B(2,\ 0{,}5)$. ¿Bajo qué condiciones se dice que la distribución binomial es aditiva o reproductiva?

b) X sigue una distribución de Poisson $P(\lambda = 2)$ e Y una distribución de Poisson $P(\lambda = 3)$ entonces $X + Y \sim P(\lambda = 5)$

c) Si $X \sim N(0,\ 1)$ y F_X es su función de distribución entonces $F_X(-x) = 1 - F_X(x)$

25. Un ascensor limita el peso de sus cuatro ocupantes a 300 kilogramos. Si el peso de una persona sigue una distribución normal $N(71,\ 7)$, calcular la probabilidad de que el peso 4 personas supere los 300 kilogramos.

26. Sean las variables aleatorias independientes X e Y, donde X se distribuye como una binomial $B(15,\ 0{,}4)$ e Y como una binomial $B(85, 0{,}4)$

a) ¿Cómo se distribuye la variable aleatoria $X + Y$?

b) ¿Se puede aproximar la variable aleatoria $X + Y$ a una distribución normal?, ¿Con qué parámetros?

c) ¿Cuando la suma de dos distribuciones binomiales independientes no se puede aproximar a una distribución normal?. Poner un ejemplo.

27. Sean las variables X e Y independientes. La variable X se distribuye como una Poisson con varianza igual a 5. La variable $Z = X + Y$ se distribuye también como una Poisson con esperanza igual a 15. ¿Cuánto vale la esperanza de la variable Y?

Analice bajo qué condiciones se puede afirmar que la distribución de Poisson es aditiva o reproductiva y determine con qué parámetros.

28. El departamento comercial de una industria alimenticia conoce que 2 de cada 10 consumidores reconocen su producto en una prueba a ciegas. ¿Cuántas pruebas ciegas de sabor deberían hacerse para que la proporción de que los que conocen la marca oscile entre el 16% y el 24% con una probabilidad mínima de 0,8?

29. Las puntuaciones obtenidas en la escala de Locus de Control de James por los sujetos depresivos, siguen una distribución normal de media 90 y desviación típica 12. Si se extraen muestras aleatorias simples de 30 sujetos depresivos. ¿Por debajo de que cantidad se encontrará el 90% de las veces el valor de la varianza de la muestra?.

30. Un servicio dedicado a la reparación de electrodomésticos recibe por término medio 15 llamadas diarias. Determinar la probabilidad de que reciba un día más de 20 llamadas.

Asignatura.. Grupo..
Apellidos .... Nombre...
Ejercicio del día ....

31. a) Dada una variable aleatoria X_1 que se distribuye como una $B(30;0{,}2)$ y otra variable aleatoria X_2 que se distribuye como una $B(150;0{,}3)$

- Sí X_1 y X_2 son independientes, ¿se puede afirmar que la variable (X_1+X_2) se distribuye como una $B(180;0{,}5)$?
- Se puede afirmar que por el teorema de Moivre la variable X_2 tiende a una distribución normal. En caso de ser así, ¿qué parámetros tiene esa distribución normal?

b) Dada una variable aleatoria con función de densidad $f(x)=\theta x^{\theta-1}$ $\theta>0$ y $0<x<1$

- Calcule la esperanza de la variable aleatoria X dejándola expresada en función de θ
- Si se propone como estimador del parámetro θ la media muestral. ¿Es este estimador insesgado?

32. La demanda media de un producto es de 100 unidades con una desviación típica de 40 unidades. Calcular la cantidad del producto que se debe tener a la venta para satisfacer la demanda de forma que puedan ser atendidos al menos el 80% de los clientes.

33. Se desea conocer el número de automóviles que se deben poner a la venta durante un periodo determinado para que se satisfaga una demanda media de 300 unidades con una desviación típica de 100 unidades, con una probabilidad no inferior al 75%.

34. La función de densidad de una variable aleatoria es:

$$f(x)=\begin{cases} ax^2+b & 0<x<2 \\ 0 & \text{en el resto} \end{cases} \qquad \text{sabiendo que } P\left[\frac{1}{2}<x<1\right]=0{,}1666.$$

Determinar a y b.

35. Completar la ley de probabilidad , conociendo que la esperanza matemática es 1,8

x_i	0	1	2	3
$P(X=x_i)=p_i$	0,2	a	b	0,3

36. En un cine de verano hay instaladas 800 sillas, sabiendo que el número de asistentes es una variable aleatoria de media 600 y desviación típica 100.
¿Qué probabilidad existe de que el número de personas que vaya al cine un día cualquiera sea superior al número de sillas instaladas?

37. Si a diario e independientemente una variable aleatoria X se distribuye como una Poisson de varianza 7,5. ¿Cómo se distribuirá mensualmente?

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

38. Se tienen dos variables aleatorias cuya distribución es normal: $X \sim N(3;2)$, $Y \sim N(5;1)$

Obtenga la esperanza matemática y la varianza de la variable: $W = 2X + 4Y$ y de la variable $Z = (X+Y)/2$

39. Se supone que la rentabilidad de un producto ofrecido por una entidad bancaria es una variable aleatoria que tiene como función de densidad: $f(x) = \begin{cases} k(x^2-1) & \text{si } 0 \le x \le 2 \\ 0 & \text{en otro caso} \end{cases}$

Hallar el valor de k para que $f(x)$ sea función de densidad.

40. Marca las respuestas correctas.

1. $H_0: \mu \geq 60$, $H_1: \mu < 60$, $t_{25\,,\,0{,}05} = 2{,}485$ y $t_{25} = -1{,}818$

a) Se acepta la hipótesis nula

b) Se rechaza la hipótesis nula

c) El estadístico muestral se ubica a la derecha del valor crítico

d) Se acepta 1 y 3

e) Ninguna opción es correcta

2. $H_0: \mu = 200$, $H_1: \mu \neq 200$, $z_{0{,}005} = 2{,}576$

a) Se acepta la hipótesis nula si $-2{,}576 \leq z \leq 2{,}576$

b) Se rechaza la hipótesis nula si $-2{,}576 \leq z \leq 2{,}576$

c) $z = (\overline{x} - 200)/(\sigma/\sqrt{n})$

d) $\alpha = 0{,}01$

e) Ninguna opción es correcta

3. $H_0: \mu = 60$, $H_1: \mu \neq 60$, $\alpha = 0{,}05$

a) Se acepta la hipótesis nula si p - valor $= 0{,}081$

b) Se rechaza la hipótesis nula si p - valor $= 0{,}04$

c) p - valor = P[rechazar el estadístico muestral observado / H_0 cierta]

d) Se acepta H_0 cuando p - valor $> \alpha$

e) Ninguna opción es correcta

4. Si el contraste $H_0: \mu = 20$, $H_1: \mu \neq 20$ se acepta con un p - valor $= 0{,}06$, el contraste $H_0: \mu \leq 20$, $H_1: \mu > 20$ tiene:

a) p - valor $= 0{,}06$

b) Contraste unilateral cola a la derecha

c) p - valor $= 0{,}03$

d) p - valor $= 0{,}97$

e) Ciertas 2 y 3

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....
Apellidos .... Nombre ....
Ejercicio del día ....

(IMPRESO EN PAPEL RECICLADO) UNICAMENTE PARA USO ESCOLAR

41. Calcular la media y la varianza de una variable aleatoria t_5 de Student

42. Encontrar la moda de una variable aleatoria $X \sim B(14, 0{,}2)$

43. Se consideran dos variables aleatorias independientes X e Y. La variable X tiene una distribución normal $N(0, 1)$, la variable Y tiene una distribución χ^2 con 4 grados de libertad.

Calcular el valor de m, siendo $P\left[\dfrac{2X}{\sqrt{Y}} \le m\right] = 0{,}05$

44. De la distribución $N(5, 2)$ se toma una muestra aleatoria simple de tamaño 20, y de la distribución $N(-4, 10)$ otra muestra aleatoria simple de tamaño 25 e independiente de la primera. Calcular $P[(\bar{x} - \bar{y}) \le 15]$

45. ¿Cuándo se utiliza la desigualdad de Chebychev para obtener intervalos de confianza?. Razonar la respuesta.

46. Se quiere comparar las puntuaciones en estadística teórica de dos Universidades que siguen, respectivamente, una distribución normal, con varianzas desconocidas y se supone que estadísticamente iguales. Se toman al azar dos muestras iguales de estudiantes de ambas Universidades. Los resultados obtenidos en SPSS son:

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ de la diferencia	95% Intervalo de confianza para la diferencia: Inferior	95% Intervalo de confianza para la diferencia: Superior
Calificación	Se han asumido varianzas iguales	,152	,699	,443	28	,661	2,200	4,962	−7,964	12,364

Con un nivel de significación de 0,05, se pide:

a) ¿Existe diferencia significativa sobre las puntuaciones de estadística teórica en las dos Universidades?

b) ¿Cuál es la media ponderada de las cuasivarianzas muestrales?

c) Cuál sería el p-valor si se hubieran contrastado las puntuaciones

$H_0 : \mu_{\text{Autónoma}} \le \mu_{\text{Complutense}} \qquad H_1 : \mu_{\text{Autónoma}} > \mu_{\text{Complutense}}$

47. Calcula el estimador máximo verosímil del parámetro λ de una distribución de Poisson. Comprueba la eficiencia.

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

48. Sea la distribución $N(\mu, \sigma)$, con la media y varianza desconocidas. Calcular los estimadores máximo-verosímiles de μ y σ^2

49.

a) ¿Cuál es el objetivo de la estimación por intervalos de confianza?.

b) ¿Para qué se utiliza el error cuadrático medio de un estimador?

c) ¿Cuál es la interpretación del concepto 'grados de libertad' al utilizar estimadores?

50.

a) ¿Qué es una hipótesis estadística?. ¿Qué es la hipótesis nula?. Razonar la respuesta.

b) Explicar conceptualmente porqué es importante que un estimador sea eficiente.

c) Explicar conceptualmente qué es el sesgo de un estimador y cuál es su interpretación.

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

SOLUCIÓN CUESTIONES DE ESTADÍSTICA TEÓRICA

1. Sea A, B y C tres sucesos incompatibles, con $P(A)=0,5$, $P(B)=0,25$, $P(C)=0,15$. Se pide:

a) $P(\overline{A}\cap\overline{B})$ b) $P(\overline{A}\cap\overline{B}\cap\overline{C})$ c) $P(\overline{A}\cap\overline{B}\cap C)$

d) Probabilidad de que exactamente se realice uno de los tres sucesos considerados.

Solución:

a) $P(A\cup B)=P(A)+P(B)-P(A\cap B) \xrightarrow{\text{incompatibles } P(A\cap B)=0} P(A\cup B)=P(A)+P(B)$

$$P(\overline{A}\cap\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-P(A)-P(B)=1-0,5-0,25=0,25$$

b) Como son incompatibles: $P(A\cup B\cup C)=P(A)+P(B)+P(C)$

$$P(\overline{A}\cap\overline{B}\cap\overline{C})=P(\overline{A\cup B\cup C})=1-P(A\cup B\cup C)=1-P(A)-P(B)-P(C)=1-0,5-0,25-0,15=0,1$$

c) Teniendo en cuenta que A, B y C son incompatibles, se tiene:

$A\cap B=\phi \qquad A\cap C=\phi \qquad B\cap C=\phi \qquad A\cap B\cap C=\phi$

$$\begin{cases} A\cap C=\phi \;\rightarrow\; C\subset\overline{A} \\ B\cap C=\phi \;\rightarrow\; C\subset\overline{B} \end{cases} \rightarrow \quad \overline{A}\cap C=C \qquad \overline{B}\cap C=C \qquad \overline{A}\cap\overline{B}\cap C=C$$

$$P(\overline{A}\cap\overline{B}\cap C)=P(C)=0,15$$

d) Sea el suceso E $\equiv$ ocurre exactamente uno de los tres sucesos A, B, C

$$E=(A\cap\overline{B}\cap\overline{C})\cup(\overline{A}\cap B\cap\overline{C})\cup(\overline{A}\cap\overline{B}\cap C) \quad \begin{cases} A\cap\overline{B}\cap\overline{C}=A \\ \overline{A}\cap B\cap\overline{C}=B \\ \overline{A}\cap\overline{B}\cap C=C \end{cases}$$

con lo cual,

$$P(E)=P\left[(A\cap\overline{B}\cap\overline{C})\cup(\overline{A}\cap B\cap\overline{C})\cup(\overline{A}\cap\overline{B}\cap C)\right]=P(A)+P(B)+P(C)=0,5+0,25+0,15=0,9$$

UAM UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

(IMPRESO EN PAPEL RECICLADO) UNICAMENTE PARA USO ESCOLAR

2. La variación porcentual del PIB de 2017, con respecto al año anterior, es una variable aleatoria normal $N(2, 0,5)$. ¿Entre qué limites oscilará la variación del PIB con probabilidad 0,95?

Solución:

$$P(2-x<PIB<2+x)=0,95$$

$$P(2-x<PIB<2+x)=P\left(\frac{2-x-2}{0,5}<\frac{PIB-2}{0,5}<\frac{2+x-2}{0,5}\right)=P\left(\frac{-x}{0,5}<z<\frac{x}{0,5}\right)=$$

$$=P\left(z>\frac{-x}{0,5}\right)-P\left(z>\frac{x}{0,5}\right)=P\left(z<\frac{x}{0,5}\right)-P\left(z>\frac{x}{0,5}\right)=\left[1-P\left(z>\frac{x}{0,5}\right)\right]-P\left(z>\frac{x}{0,5}\right)=$$

$$=1-2P\left(z>\frac{x}{0,5}\right)=0,95 \quad\rightarrow\quad \frac{x}{0,5}=1,96 \quad\Rightarrow\quad x=0,98$$

En consecuencia, $P(1,02<PIB<2,98)=0,95$

El PIB oscilará entre 1,02 y 2,98

3. Dadas tres variables aleatorias independientes, con características: $X_1 \sim N(0, 2)$, $X_2 \sim N(2, 2)$, $X_3 \sim N(4, 1)$. Calcúlese $P(-7 \le Y \le 10)$, siendo $Y = 4X_1 + 5X_2 - 6X_3 + 6$

Solución:

La variable aleatoria Y sigue una distribución normal:

$$E(Y)=4E(X_1)+5E(X_2)-6E(X_3)+6=4 \text{ x } 0+5 \text{ x } 2-6 \text{ x } 4+6=-8$$

$$V(Y)=4^2\,V(X_1)+5^2\,V(X_2)+(-6)^2\,V(X_3)=16 \text{ x } 4+25 \text{ x } 4+36 \text{ x } 1=200$$

Siendo $Y \sim N(-8, \sqrt{200}) \equiv N(-8, 14,14)$

$$P(-7\le Y\le 10)=P\left(\frac{-7-(-8)}{14,14}\le\frac{Y-(-8)}{14,14}\le\frac{10-(-8)}{14,14}\right)=P(0,07\le z\le 1,27)=$$

$$=P(z\ge 0,07)-P(z\ge 1,27)=0,4721-0,1020=0,3701$$

4. El peso (en gramos) de unas cajas es una variable aleatoria X, con distribución normal, cuya desviación típica es igual a 12 gramos. Se extraen muestras aleatorias simples de tamaño 100 cajas. ¿Cuál es la probabilidad de que el peso medio de las cajas supere el peso medio real en al menos 300 miligramos?

Solución:

$X \sim N(\mu, 12)$ con $n = 100$

Se pide calcular $P(\overline{x} \geq \mu + 0,3) \quad \rightarrow \quad P(\overline{x} - \mu \geq 0,3)$

Se puede determinar que la variable aleatoria $(\overline{x} - \mu)$ sigue una distribución normal con las siguientes características:

$$E(\overline{x} - \mu) = E(\overline{x}) - \mu = \mu - \mu = 0$$

$$V(\overline{x} - \mu) = V(\overline{x}) = \frac{\sigma^2}{n} = \frac{12^2}{100} = \frac{144}{100} = 1,44 \qquad (\overline{x} - \mu) \sim N(0, 1,2)$$

$$P(\overline{x} - \mu \geq 0,3) = P\left(\frac{\overline{x} - \mu}{1,2} \geq \frac{0,3}{1,2}\right) = P(z \geq 0,25) = 0,4013$$

5. El número de llamadas independientes a la central de una agencia de viajes durante un día es una distribución de Poisson de parámetro $\lambda = 100$. ¿Cuál es la probabilidad de que durante 30 días, el total de llamadas se encuentre entre 2900 y 3100?

Solución:

La distribución de Poisson es reproductiva, en el sentido de que la suma de variables de Poisson independientes es otra variable de Poisson.

$$X_i \sim P(\lambda) \quad \rightarrow \quad Y = \sum_{i=1}^{n} X_i \sim P(n\lambda) \quad \text{donde } E(Y) = n\lambda \; , \; V(Y) = n\lambda$$

En 30 días $n = 30$: $Y \sim P(3000)$ donde $E(Y) = 3000$, $V(Y) = 3000$

$$P(2900 \leq Y \leq 3100) = \sum_{k=2900}^{3100} \frac{3000^k}{k!} e^{-3000}$$

Cálculo prácticamente imposible de calcular, por el teorema central del límite se tiene:

$$Y \sim P(n\lambda) \longrightarrow Y \sim N\left(n\lambda, \sqrt{n\lambda}\right) \quad E(Y) = n\lambda \; , \; V(Y) = n\lambda$$

$$P(2900 \leq Y \leq 3100) = P\left(\frac{2900-3000}{\sqrt{3000}} \leq \frac{Y-3000}{\sqrt{3000}} \leq \frac{3100-3000}{\sqrt{3000}}\right) = P(-1,82 \leq z \leq 1,82) =$$

$$= 1 - 2 \times P(z \geq 1,82) = 1 - 2 \times 0,0344 = 0,9312$$

6. El peso de un paquete de cigarrillos se distribuye según una $N(80, 4)$. Se dispone de dos lotes de 100 paquetes cada uno de ellos. ¿Cuál será la probabilidad de que el peso del primer lote supere en más de 40 gramos al peso del segundo lote?

Solución:

Denotando por $\begin{cases} x \equiv \text{peso del paquete de cigarrillos del primer lote} \\ y \equiv \text{peso del paquete de cigarrillos del segundo lote} \end{cases}$

$\sum_{i=1}^{100} x_i \equiv$ peso del primer lote $\qquad \sum_{i=1}^{100} y_i \equiv$ peso del segundo lote

$$P\left(\sum_{i=1}^{100} x_i - \sum_{i=1}^{100} y_i > 40\right) = P\left(\frac{\sum_i x_i}{100} - \frac{\sum_i y_i}{100} > \frac{40}{100}\right) = P(\overline{x} - \overline{y} > 0,4)$$

La distribución $(\overline{x} - \overline{y}) \sim N\left(0, \sqrt{\frac{\sigma^2}{100} + \frac{\sigma^2}{100}}\right) \equiv N\left(0, \sqrt{\frac{4^2}{100} + \frac{4^2}{100}}\right) = N(0, 0,56)$

$$P(\overline{x} - \overline{y} > 0,4) = P\left(\frac{\overline{x} - \overline{y}}{0,56} > \frac{0,4}{0,56}\right) = P(z > 0,71) = 0,2389$$

Asignatura .... Grupo ....

Apellidos .... Nombre ....

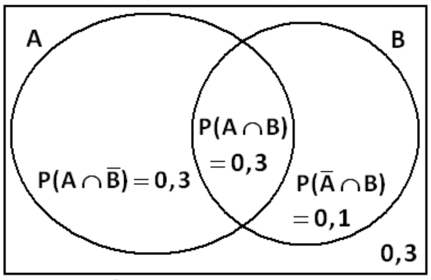
Ejercicio del día ....

7. Sean A y B dos sucesos aleatorios tales que $P(A)=0,6$, $P(B)=0,4$ y $P(\overline{A}\cup\overline{B})=0,7$. Calcula:

a) $P(A\cap\overline{B})$ b) $P(A/B)$ c) $P(\overline{\overline{A}\cap B})$

Solución:

$P(\overline{A}\cup\overline{B})=P(\overline{A\cap B})=0,7 \;\rightarrow\; P(A\cap B)=0,3$



a) $P(A\cap\overline{B})=0,3$

b) $P(A/B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{0,3}{0,4}=\dfrac{3}{4}=0,75$

c) $P(\overline{A}\cap B)=0,1 \;\rightarrow\; P(\overline{\overline{A}\cap B})=0,9$

8. Sea X una variable aleatoria que sigue una distribución χ^2 con 200 grados de libertad. Calcular: (a) $P(X\leq 234)$ y (b) $P(X\leq 183)$ sin recurrrir a las tablas de la Chi-cuadrado.

Solución:

La distribución χ^2 es, asintóticamente, $N\left(n,\sqrt{2n}\right)$, es decir, $\chi^2_{200}\sim N\left(200,\sqrt{400}\right)$

a) $P(X\leq 234)=P\left(\dfrac{X-200}{\sqrt{400}}\leq\dfrac{234-200}{\sqrt{400}}\right)=P(z\leq 1,7)=0,9554$

b) $P(X\leq 183)=P\left(\dfrac{X-200}{\sqrt{400}}\leq\dfrac{183-200}{\sqrt{400}}\right)=P(z\leq -0,85)=P(z\geq 0,85)=0,1977$

9. Dada una variable aleatoria X que se distribuye como una $B(30;\ 0,2)$ y otra variable aleatoria Y que se distribuye como una $B(150;\ 0,3)$.

a) ¿Si X e Y son independientes, se puede afirmar que la variable $X+Y$ se distribuye como una $B(180;\ 0,5)$? Justifica tu respuesta.

b) ¿Se puede afirmar que por el teorema de Moivre la variable Y tiende a una normal? Si fuera así, ¿qué parámetros tiene esa distribución normal?

Solución:

a) No se distribuye como una $B(180;\ 0,5)$, para aplicar la propiedad reproductiva tienen que tener la misma probabilidad p.

b) Según el TCL $Y \sim N\left(150 \times 0,3\, ,\ \sqrt{150 \times 0,3 \times 0,7}\right) \equiv N\left(4,5\, ,\ 5,61\right)$

10. Los errores que se cometen al estimar el ahorro familiar de un país siguen una distribución normal de media 0 y desviación σ_e. Comprobar que la función $\dfrac{n\,\overline{e}^2}{\sigma_e^2}$ sigue una distribución Chi-cuadrado, sabiendo que el tamaño muestral es n y el error medio muestral es $\overline{e}$.

Solución:

Si $X_1, X_2, \cdots, X_n$ son n variables aleatorias independientes $N(0,1)$, la variable Chi-cuadrado de Pearson con n grados de libertad es $\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_2^n$

La variable aleatoria E = "errores al estimar el ahorro familiar" sigue una $N(0,\sigma_e)$

La media muestral de los errores $\overline{e} \sim N\left(0, \dfrac{\sigma_e}{\sqrt{n}}\right)$ con lo que la variable tipificada:

$$z = \frac{\overline{e} - 0}{\sigma_e / \sqrt{n}} \sim N(0,1) \quad \rightarrow \quad z^2 = \left(\frac{\overline{e}}{\sigma_e / \sqrt{n}}\right)^2 = \frac{n\,\overline{e}^2}{\sigma_e^2} \sim \left[N(0,1)\right]^2 = \chi_1^2$$

11. La probabilidad de que un paciente mejore tras administrarle un medicamento es de 0,82.
a) Se suministra de manera independiente el medicamento a 10 pacientes en la Clínica Salud. ¿Qué modelo de distribución de probabilidad sigue la variable número total de pacientes que mejoran en dicha clínica?
b) Si cada una de las 50 clínicas de una provincia realizan el mismo experimento de manera independiente, ¿Qué modelo de distribución de probabilidad sigue la variable número total de pacientes que mejoran en dicha provincia? ¿Cuál es la probabilidad de que al menos 400 pacientes mejoren en esa provincia?

Solución:

a) X = "Mejoría con el medicamento en una Clínica" $X \sim B(n,p) \equiv B(10, 0,82)$

b) Y = "Mejoría con el medicamento en las 50 clínicas" $Y = X_1 + X_2 + \cdots + X_{50} = \sum_{i=1}^{50} X_i$

$$Y \sim B\left(\sum_{i=1}^{50} X_i, 0,82\right) \equiv B(500, 0,82) \sim N\left(500 \times 0,82, \sqrt{500 \times 0,82 \times 0,18}\right) \equiv N(410, 8,59)$$

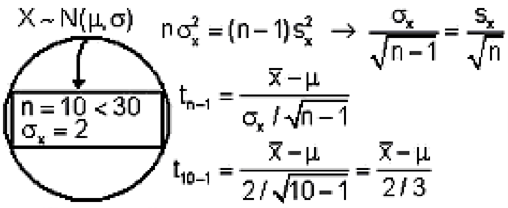
$$P(Y \geq 400) = P\left[\frac{Y-410}{8,59} \geq \frac{400-410}{8,59}\right] = P(z \geq -1,16) = P(z \leq 1,16) = 1 - P(z \geq 1,16) =$$

$$= 1 - 0,1230 = 0,877$$

12. Se ha realizado una muestra aleatoria simple (m.a.s) de tamaño 10 de una población considerada normal, llegando a la conclusión que la varianza muestal es 4. Calcular la probabilidad $P\left[|\bar{x} - \mu| \leq 1,22\right]$

Solución:

$$P\left[|\bar{x} - \mu| \leq 1,22\right] = P\left[\frac{|\bar{x} - \mu|}{2/3} \leq \frac{1,22}{2/3}\right] =$$



$$= P\left[|t_9| \leq 1,83\right] = P\left[-1,83 \leq t_9 \leq 1,83\right] =$$

$$= P\left[t_9 \geq -1,83\right] - P\left[t_9 \geq 1,83\right] = P\left[t_9 \leq 1,83\right] - P\left[t_9 \geq 1,83\right] = 1 - 2P\left[t_9 \geq 1,83\right] = 0,9$$

13. Se desea dar una estimación por intervalos a un 95% de confianza para la proporción de alumnos que están a favor de aumentar en número máximo de convocatorias permitidas por la normativa de permanencia. ¿Qué tamaño de muestra habría que considerar si se desea un intervalo con una amplitud de 0,06?. Suponga que $p = q = 1/2$.

Solución:

$$A = 2 z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad z_{\alpha/2} = z_{0,025} = 1,96 \quad \rightarrow \quad 0,06 = 2 \times 1,96 \times \sqrt{\frac{1}{4n}}$$

$$\frac{0,06}{2 \times 1,96} = \sqrt{\frac{1}{4n}} \quad \mapsto \quad 0,0153 = \sqrt{\frac{1}{4n}} \quad \mapsto \quad 0,000234 = \frac{1}{4n} \quad \mapsto \quad n = 1068$$

14. Sean A y B dos sucesos tales que $P(A) = 0,6$. Calcular $P(A \cap \overline{B})$ en cada caso:

a) A y B mutuamente excluyentes

b) A está contenido en B

c) B está contenido en A y $P(B) = 0,3$

d) $P(A \cap B) = 0,1$

Solución:

a) A y B mutuamente excluyentes $\Rightarrow \quad A \cap B = \phi \quad \mapsto P(A \cap B) = 0$

$P(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A) - 0 = 0,6$

b) $A \subset B \Rightarrow P(A \cap B) = P(A) \quad \mapsto \quad P(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A) - P(A) = 0$

c) $B \subset A \Rightarrow P(A \cap B) = P(B) \quad \mapsto$

$\mapsto \quad P(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A) - P(B) = 0,6 - 0,3 = 0,3$

d) $P(A \cap \overline{B}) = P(A) - P(A \cap B) = 0,6 - 0,1 = 0,5$

15. Sean A y B dos sucesos incompatibles con $P(A\cup B)>0$. Demostrar que

$$P(B/\,A\cup B)=\frac{P(B)}{P(A)+P(B)}$$

Solución:

A y B incompatibles $\mapsto$ $A\cap B=0$ $\mapsto$ $P(A\cup B)=P(A)+P(B)$

$$P(B\,/\,A\cup B)=\frac{P\big[B\cap(A\cup B)\big]}{P(A\cup B)}=\frac{P(B)}{P(A)+P(B)}$$

16. Las variables aleatorias en estudio son independientes.

Sean: $X_1\sim N(10,2)$, $X_2\sim N(15,2)$ y $X_3\sim N(20,2)$ donde $Y=5X_1+4X_2-3X_3$ y

$U=\left(\dfrac{X_1-10}{2}\right)^2+\left(\dfrac{X_2-15}{2}\right)^2+\left(\dfrac{X_3-20}{2}\right)^2$. Calcular: a) $P[20\le Y\le 30]$ b) $P[U\le 11]$

Solución:

a) $Y=5X_1+4X_2-3X_3$ siendo $Y\sim N\left(\displaystyle\sum_{i=1}^{3}\lambda_i\,\mu_i\,,\ \sqrt{\sum_{i=1}^{3}\lambda_i^2\,\sigma_i^2}\right)$

$$Y\sim N\Big[5\,x\,10+4\,x\,15-3\,x\,20\,,\sqrt{5^2\,x\,2^2+4^2\,x\,2^2+(-3)^2\,x\,2^2}\,\Big]\ \mapsto\ Y\sim N\big[50\,,2\sqrt{50}\,\big]$$

$$P[20\le Y\le 30]=P\left[\frac{20-50}{2\sqrt{50}}\le\frac{Y-50}{2\sqrt{50}}\le\frac{30-50}{2\sqrt{50}}\right]=P(-2{,}12\le z\le -1{,}41)=$$

$$=P(1{,}41\le z\le 2{,}12)=P(z\ge 1{,}41)-P(z\ge 2{,}12)=0{,}0793-0{,}0170=0{,}0623$$

b) $\begin{cases} X_1\sim N(10,2)\ \mapsto\ \dfrac{X_1-10}{2}\sim N(0,1)\\[2ex] X_2\sim N(15,2)\ \mapsto\ \dfrac{X_2-15}{2}\sim N(0,1)\\[2ex] X_3\sim N(20,2)\ \mapsto\ \dfrac{X_3-20}{2}\sim N(0,1)\end{cases}$

$$U=\left(\frac{X_1-10}{2}\right)^2+\left(\frac{X_2-15}{2}\right)^2+\left(\frac{X_3-20}{2}\right)^2\ \mapsto\ U=\big[N(0,1)\big]^2+\big[N(0,1)\big]^2+\big[N(0,1)\big]^2=\chi_3^2$$

$$P[U\le 11]=P\big[\chi_3^2\le 11\big]=1-P\big[\chi_3^2\ge 11\big]=1-0{,}05=0{,}95$$

17. El error cometido en expedir tiques por una maquina sigue una normal $N(0,\sigma)$.

a) ¿Cuál es la probabilidad de que el error cometido en valor absoluto de una medida cualquiera sea al menos σ?

b) ¿Cuánto valdría la probabilidad si se toma como medida la media aritmética de 10 medidas independientes?

Solución:

a) Sea la variable aleatoria X= "expedir tiques por la maquina", $N(0,\sigma)$

$$P\left[|X|\geq\sigma\right]=P\left[\left|\frac{X-0}{\sigma}\right|\geq\frac{\sigma-0}{\sigma}\right]=P\left[|z|\geq 1\right]=P\left[(z\leq -1)\cup(z\geq 1)\right]=P(z\leq -1)+P(z\geq 1)=$$

$$=P(z\geq 1)+P(z\geq 1)=2P(z\geq 1)=2\,x\,0,1587=0,3174$$

b) Sea la variable aleatoria $Y=X_1+X_2+\cdots+X_{10} \quad \mapsto \quad \overline{y}=\frac{1}{10}\sum_{i=1}^{10}X_i$

$$E(\overline{y})=E\left[\frac{1}{10}\sum_{i=1}^{10}X_i\right]=\frac{1}{10}E\left[\sum_{i=1}^{10}X_i\right]=\frac{1}{10}10\mu=\mu=0$$

$$\sigma^2_{\overline{y}}=Var(\overline{y})=Var\left[\frac{1}{10}\sum_{i=1}^{10}X_i\right]=\frac{1}{100}Var\left[\sum_{i=1}^{10}X_i\right]=\frac{1}{100}10\sigma^2=\frac{\sigma^2}{10}$$

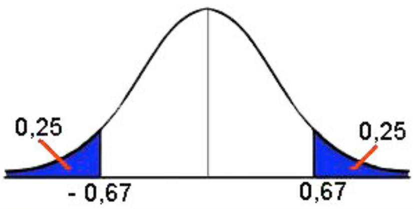
$$\overline{y}\sim N\left[0,\frac{\sigma}{\sqrt{10}}\right]$$

18. En una población de mujeres, las puntuaciones de un test de ansiedad-riesgo siguen una distribución normal $N(25,10)$. Al clasificar la población en cuatro grupos de igual tamaño, ¿cuales serán las puntuaciones que delimiten estos grupos?.

Solución:

Siendo la variable aleatoria X = "puntuaciones en un test de ansiedad-riesgo"

$$P(X \le Q_1) = 0,25 \quad \mapsto \quad P\left(\frac{X-25}{10} \le \frac{Q_1-25}{10}\right) = P\left(z \le \frac{Q_1-25}{10}\right) = 0,25$$



Las puntuaciones que delimitan estos cuatro grupos serán el primer Q_1, segundo Q_2 y tercer cuartil Q_3 de la distribución.

$$P(z \le -0,67) = 0,25 \qquad P(z \ge 0,67) = 0,25$$

$$\frac{Q_1-25}{10} = -0,67 \quad \mapsto \quad Q_1 = 25 - 0,67 \text{x} 10 = 18,3$$

En la distribución normal la media y la mediana son iguales: $\mu = M_e = Q_2 = 25$

$$P(X \le Q_3) = 0,75 \quad \mapsto \quad P\left(\frac{X-25}{10} \le \frac{Q_3-25}{10}\right) = P\left(z \le \frac{Q_3-25}{10}\right) = 0,75 \quad \mapsto$$

$$\mapsto \quad P\left(z \ge \frac{Q_3-25}{10}\right) = 0,25$$

$$\frac{Q_3-25}{10} = 0,67 \quad \mapsto \quad Q_3 = 25 + 0,67 \text{x} 10 = 31,7$$

Por consiguiente, el primer grupo serían las mujeres con puntuaciones inferiores o iguales a 18,3. El segundo grupos son aquellas mujeres con puntuaciones entre 18,3 y 25.
El tercer grupo son las mujeres con puntuaciones entre 25 y 31,7. El cuarto grupo son mujeres que tengan puntuaciones superiores a 31,7.

19. El número de millones de metros cúbicos que tiene un embalse sigue una distribución normal $N(980, 50)$. El consumo diario de las poblaciones que sirve es una normal $N(85, 30)$. Si se sabe que durante una tormenta la cantidad de agua que se embolsa es una normal $N(50, 25)$. Un día han caído dos tormentas, calcular la probabilidad de que al final del día el agua embalsada sea menor o igual que 980 metros cúbicos.

Solución:

Agua del embalse: $X_1 \sim N(\mu_1, \sigma_1) \equiv N(980, 50)$

Consumo diario: $X_2 \sim N(\mu_2, \sigma_2) \equiv N(85, 30)$

$$\begin{cases} \text{Agua una tormenta:} & X_3 \sim N(\mu_3, \sigma_3) \equiv N(50, 25) \\ \text{Agua dos tormentas:} & \begin{cases} E(2X_3) = 2E[X_3] = 2 \times 50 = 100 \\ Var(2X_3) = 4\,Var[X_3] = 4 \times 25^2 \end{cases} \end{cases}$$

$$2X_3 \sim N\left(\sum_{i=1}^{2} \lambda_i \mu_i \,,\, \sqrt{\sum_{i=1}^{2} \lambda_i^2 \sigma_i^2}\right) \equiv N(2 \times 0\,,\, 2 \times 25)$$

Agua embalse un día con dos tormentas: $Y = X_1 - X_2 + 2X_3$ siendo $Y \sim N\left(\sum_{i=1}^{3} \lambda_i \mu_i \,,\, \sqrt{\sum_{i=1}^{3} \lambda_i^2 \sigma_i^2}\right)$

$$Y \sim N\left(980 - 85 + 100,\ \sqrt{50^2 + 30^2 + 4 \times 25^2}\right) \equiv N(995,\ 76{,}81)$$

$$P(Y \le 980) = P\left[\frac{Y - 995}{76{,}81} \le \frac{980 - 995}{76{,}81}\right] = P(z \le -0{,}19) = P(z \ge 0{,}19) = 2P(z \ge 0{,}19) = 0{,}4247$$

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura.. Grupo..............................

Apellidos .... Nombre.............................

Ejercicio del día ....

20. Un dispositivo está formado por muchos elementos que trabajaban independientemente, siendo la probabilidad de fallo durante la primera hora de trabajo muy pequeña e iguales en todos los elementos. Si la probabilidad de que en ese tiempo falle por lo menos un elemento es 0,98.
a) Hallar la media y desviación típica del número de elementos que fallen en la primera hora.
b) Calcular la probabilidad de que fallen a lo sumo dos elementos en ese tiempo.

Solución:

a) Habiendo muchos elementos independientes, el tamaño (n) es muy grande, con una probabilidad de fallo muy pequeña, condiciones para que el dispositivo siga una distribución de Poisson.

Sea la variable aleatoria X = "fallo en la primera hora de un elemento", $X \sim P(\lambda)$ donde

$$P(X=k)=\frac{\lambda^k}{k!}\,e^{-\lambda}$$

$$P(X\geq 1)=0,98 \;\;\mapsto\;\; P(X<1)=P(X=0)=\frac{\lambda^0}{0!}\,e^{-\lambda}=0,02 \;\;\mapsto\;\; e^{-\lambda}=0,02$$

$$-\lambda=\ln 0,02 \;\;\mapsto\;\; \lambda=3,91$$

$$\mu=\lambda=3,91 \qquad \sigma=\sqrt{3,91}=1,98$$

b) $$P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)=\frac{3,9^0}{0!}\,e^{-3,9}+\frac{3,9}{1!}\,e^{-3,9}+\frac{3,9^2}{2!}\,e^{-3,9}=$$

$$=\left[1+3,9+\frac{3,9^2}{2}\right]e^{-3,9}=12,505 \text{ x } e^{-3,9}=0,2531$$

21. El departamento comercial de una industria alimenticia sabe que 2 de cada 10 consumidores reconocen su producto en una prueba a ciegas. ¿Cuántas pruebas a ciegas de sabor deberían hacerse para que la proporción de que los que conocen la marca oscile entre el 16% y el 24% con una probabilidad mínima de 0,8?

Solución:

Reconocen el producto el 20%, $p = 0,2$

$$\hat{p} \approx N\left(p, \sqrt{\frac{pq}{n}}\right) \qquad \hat{p} \approx N\left(0,2, \sqrt{\frac{0,2 \times 0,8}{n}}\right) = N\left(0,2, \frac{0,4}{\sqrt{n}}\right)$$

$$P(0,16 \leq \hat{p} \leq 0,24) = P\left(\frac{0,16 - 0,2}{0,4/\sqrt{n}} \leq \frac{\hat{p} - 0,2}{0,4/\sqrt{n}} \leq \frac{0,24 - 0,2}{0,4/\sqrt{n}}\right) =$$

$$= P\left[-0,1\sqrt{n} \leq z \leq 0,1\sqrt{n}\right] = P\left[z \geq -0,1\sqrt{n}\right] - P\left[z \geq 0,1\sqrt{n}\right] =$$

$$= 1 - 2P\left(z > 0,1\sqrt{n}\right) = 0,8 \quad \mapsto \quad P\left(z \geq 0,1\sqrt{n}\right) = 0,1$$

$$0,1\sqrt{n} = 1,282 \quad \mapsto \quad n = 165$$

Para una probabilidad como mínimo de 0,8 harían falta 165 pruebas.

22. El p-valor en un contraste de hipótesis ha resultado ser de 0,08. Si el nivel de significación del contraste ha sido de 0.05 , ¿Se debería aceptar o rechazar la hipótesis nula?

Solución:

El p-valor es el menor nivel de significación en el que se rechaza una hipótesis.

Para niveles de significación α mayor que el p-valor se rechaza H_0, para niveles de significación menor que el p-valor se acepta H_0.

Como $\alpha = 0,05 < \text{p-valor} = 0,08$ se acepta H_0

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

23. Defina el Error de Tipo I y el Error de Tipo II. Analice la verdad o no de las siguientes afirmaciones y explique por qué:

a) P(Error tipo I) + P(Error Tipo II) = 1

b) Cuando disminuye la probabilidad de cometer un Error de Tipo I, también disminuye la probabilidad de cometer un Error de Tipo II.

Solución:

a) Error Tipo I: $\alpha = P(\text{ET I}) = P(\text{Rechazar } H_0 \,/\, H_0 \text{ cierta})$

Un Error Tipo I es un falso positivo, como una alarma de fuego que suena cuando no existe tal fuego.

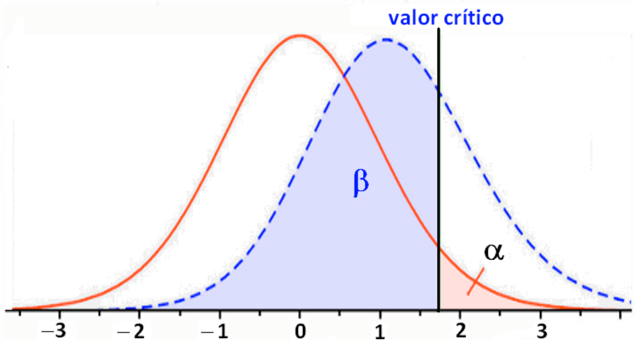
Error Tipo II: $\beta = P(\text{ET II}) = P(\text{Aceptar } H_0 \,/\, H_0 \text{ falsa})$

Un Error Tipo II es el caso de un falso negativo, como una alarma que falla y no suena cuando existe un fuego.

Un Error Tipo I puede tener consecuencias menos aceptables que las que tendría un Error Tipo II. Por otro lado, en términos económicos, cometer un Error Tipo II podría ser menos costoso que un Error Tipo I. En esta línea, no se puede responder a qué error sería peor.

No es verdad P(Error tipo I) + P(Error Tipo II) = 1 porque no son complementarios.

b) Se encuentran inversamente relacionados.



24. Las variables aleatorias en estudio son independientes. Analiza si las afirmaciones son verdaderas o falsas:

a) X sigue una distribución binomial $B(1, 0{,}3)$ e Y una distribución binomial $B(1,\ 0{,}2)$ entonces X+Y sigue una binomial $B(2,\ 0{,}5)$. ¿Bajo qué condiciones se dice que la distribución binomial es aditiva o reproductiva?

b) X sigue una distribución de Poisson $P(\lambda = 2)$ e Y una distribución de Poisson $P(\lambda = 3)$ entonces $X + Y \sim P(\lambda = 5)$

c) Si $X \sim N(0,\ 1)$ y F_X es su función de distribución entonces $F_X(-x)\ =\ 1\ -\ F_X(x)$

Solución:

a) Las distribuciones binomiales son reproductivas de parámetro p, es decir, dadas dos variables aleatorias $X \sim B(n, p)$ e $Y \sim B(m, p)$ siendo independientes se verifica que $X + Y \sim B(n+m, p)$. En consecuencia, una variable aleatoria $X \sim B(n, p)$ se puede descomponer en suma de n variables aleatorias independientes de Bernouilli de parámetro p.

Para poder aplicar la propiedad reproductiva, han de ser independientes y con la misma probabilidad.

b) Es cierto, la distribución de Poisson es reproductiva.

c) Por la simetría de la distribución normal se verifica $F_X(-x)\ =\ 1\ -\ F_X(x)\quad \forall x \in \mathbb{R}$

25. Un ascensor limita el peso de sus cuatro ocupantes a 300 kilogramos. Si el peso de una persona sigue una distribución normal $N(71,\ 7)$, calcular la probabilidad de que el peso 4 personas supere los 300 kilogramos.

Solución:

MÉTODO I: Si el peso de una persona sigue una distribución normal $N(71,\ 7)$, la muestra de 4 personas sigue una distribución normal $N\left(71,\ \dfrac{71}{\sqrt{4}}\right) \equiv N(\ 71,\ 3{,}5)$

$$P\big[(X_1 + X_2 + X_3 + X_4) > 300\big] = P\left[\frac{X_1 + X_2 + X_3 + X_4}{4} > \frac{300}{4}\right] = P\big[\ \overline{x} > 75\big] =$$

$$= P\big[\ \overline{x} > 75\big] = P\left[\frac{\overline{x} - 71}{3{,}5} > \frac{75 - 71}{3{,}5}\right] = P\big[\ z > 1{,}143\big] = 0{,}1265$$

Interpolando: $\dfrac{0,1271-0,1251}{1,14-1,15}=\dfrac{x-0,1251}{1,143-1,15} \mapsto \dfrac{0,002}{-0,01}=\dfrac{x-0,1251}{-0,007}$

$$x=0,1251+\frac{0,002\,.\,0,007}{0,01}=0,1265$$

MÉTODO II: Si $x_i \sim N[71\,;\,7] \mapsto \sum_{i=1}^{4} x_i \sim N\left[\overbrace{4\,.\,71}^{n\,.\,\mu}\,;\,\overbrace{7\,.\sqrt{4}}^{\sigma\,.\sqrt{n}}\right]=N[284\,;\,14]$

$$P\left[\sum_{i=1}^{4} x_i \geq 300\right]=P\left[\frac{\sum_{i=1}^{4} x_i-284}{14} \geq \frac{300-284}{14}\right]=P[z \geq 1,143]=0,1265$$

26. Sean las variables aleatorias independientes X e Y, donde X se distribuye como una binomial B(15, 0,4) e Y como una binomial B(85,0,4)

a) ¿Cómo se distribuye la variable aleatoria $X+Y$?

b) ¿Se puede aproximar la variable aleatoria $X+Y$ a una distribución normal?, ¿Con qué parámetros?

c) ¿Cuando la suma de dos distribuciones binomiales independientes no se puede aproximar a una distribución normal?. Poner un ejemplo.

Solución:

a) Las distribuciones binomiales independientes son reproductivas cuando tienen el mismo parámetro p. En consecuencia, $X \sim B(15,\ 0,4)$ e $Y \sim B(85,0,4)$ se tiene que $X+Y \sim B(100,\ 0,4)$

b) $X+Y \sim B(100,\ 0,4)$ donde $p=0,4<0,5$ y $n\,p=100 \times 0,4=40>5$, por el teorema de Moivre o el Teorema Central del Limite (TCL) la distribución binomial de $X+Y$ se puede aproximar a una distribución normal $N(40,\ 4,89)$ siendo $\mu=np=40$ y $\sigma=\sqrt{100 \times 0,4 \times 0,6}=4,89$

c) Cuando el tamaño fuera pequeño (n pequeña). Ejemplo: $X+Y \sim B(8,\ 0,4)$

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....
Apellidos .... Nombre ....
Ejercicio del día ....

(IMPRESO EN PAPEL RECICLADO) ÚNICAMENTE PARA USO ESCOLAR

27. Sean las variables X e Y independientes. La variable X se distribuye como una Poisson con varianza igual a 5. La variable $Z = X + Y$ se distribuye también como una Poisson con esperanza igual a 15. ¿Cuánto vale la esperanza de la variable Y?
Analice bajo qué condiciones se puede afirmar que la distribución de Poisson es aditiva o reproductiva y determine con qué parámetros.

Solución:

La suma de variables aleatorias de Poisson independientes es otra variable aleatoria de Poisson con parámetros la suma de los parámetros.

$$\begin{cases} X \sim P(\lambda = 5) & \sigma_X^2 = \lambda = 5 \\ Z \sim P(\lambda = 15) & E(Z) = \lambda = 15 \end{cases} \quad \mapsto \quad Y \sim P(\lambda = 10) \quad E(Y) = 10$$

28. El departamento comercial de una industria alimenticia conoce que 2 de cada 10 consumidores reconocen su producto en una prueba a ciegas. ¿Cuántas pruebas ciegas de sabor deberían hacerse para que la proporción de que los que conocen la marca oscile entre el 16% y el 24% con una probabilidad mínima de 0,8?

Solución:

Reconocen el producto el 20%, $p = 0,2 \quad P(0,16 \le \hat{p} \le 0,24) \ge 0,8$

$$\hat{p} \approx N\left(p, \sqrt{\frac{pq}{n}}\right) \quad \to \quad \hat{p} \approx N\left(0,2,\ \sqrt{\frac{0,2 \times 0,8}{n}}\right) = N\left(0,2,\ \frac{0,4}{\sqrt{n}}\right)$$

$$P\left(\frac{0,16 - 0,2}{0,4/\sqrt{n}} < z < \frac{0,24 - 0,2}{0,4/\sqrt{n}}\right) = P\left(-0,1\sqrt{n} < z < 0,1\sqrt{n}\right) = 0,8$$

$$P\left(-0,1\sqrt{n} < z < 0,1\sqrt{n}\right) = 1 - 2P\left(z > 0,1\sqrt{n}\right) = 0,8 \quad \mapsto \quad P\left(z \ge 0,1\sqrt{n}\right) = 0,1$$

$$0,1\sqrt{n} = 1,282 \quad \mapsto \quad n = 165$$

Para una probabilidad como mínimo de 0,8 harían falta 165 pruebas

29. Las puntuaciones obtenidas en la escala de Locus de Control de James por los sujetos depresivos, siguen una distribución normal de media 90 y desviación típica 12. Si se extraen muestras aleatorias simples de 30 sujetos depresivos. ¿Por debajo de que cantidad se encontrará el 90% de las veces el valor de la varianza de la muestra?.

Solución:

En virtud del teorema de Fisher: En el muestreo, si se toman muestras aleatorias de media $\overline{x}$ y desviación típica σ_x de una población $N(\mu,\sigma)$, la variable $\chi^2_{n-1} = \dfrac{(n-1)s^2}{\sigma^2}$,

donde $s^2 \equiv$ cuasivarianza muestral $n\sigma_x^2 = (n-1)s^2$

Las puntuaciones obtenidas siguen una distribución $N(90,12)$

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2} = \frac{n\sigma_x^2}{\sigma^2} \quad \text{con lo que} \quad \chi^2_{29} = \frac{30\sigma_x^2}{144}$$

De las tablas de la Chi-cuadrado, $P(\chi^2_{29} \le k) = 0,9 \Rightarrow P(\chi^2_{29} \ge k) = 0,1 \mapsto k = 39,087$

con lo cual, $P\left(\dfrac{30\sigma_x^2}{144} \le 39,087\right) = 0,9 \Rightarrow P\left(\sigma_x^2 \le \dfrac{39,087 \text{ x } 144}{30}\right) = P\left(\sigma_x^2 \le 187,62\right) = 0,9$

El valor pedido será 187,62

30. Un servicio dedicado a la reparación de electrodomésticos recibe por término medio 15 llamadas diarias. Determinar la probabilidad de que reciba un día más de 20 llamadas.

Solución:

Sea $X =$ "número de llamadas recibidas al día"

La variable aleatoria: $X \sim P\left[\lambda = 15\right]$: $P\left[X = k\right] = \dfrac{15^k}{k!}\,.\,e^{-15}$

$$P\left[\lambda = 15\right] \xrightarrow{\lambda = 15 > 10} N\left[15, \sqrt{15}\right]$$

$$P\left[X > 20\right] = P\left[\frac{X - 15}{\sqrt{15}} > \frac{(20 - 0,5) - 15}{\sqrt{15}}\right] = P\left[z > 1,16\right] = 0,1230$$

31. a) Dada una variable aleatoria X_1 que se distribuye como una $B(30;0,2)$ y otra variable aleatoria X_2 que se distribuye como una $B(150;0,3)$

- **Sí X_1 y X_2 son independientes, ¿se puede afirmar que la variable (X_1+X_2) se distribuye como una $B(180;0,5)$?**
- **Se puede afirmar que por el teorema de Moivre la variable X_2 tiende a una distribución normal. En caso de ser así, ¿qué parámetros tiene esa distribución normal?**

b) Dada una variable aleatoria con función de densidad $f(x)=\theta x^{\theta-1}$ $\quad \theta>0$ y $0<x<1$

- **Calcule la esperanza de la variable aleatoria X dejándola expresada en función de θ**
- **Si se propone como estimador del parámetro θ la media muestral. ¿Es este estimador insesgado?**

Solución:

a) La variable (X_1+X_2) no sigue una distribución $B(180;0,5)$, para aplicar la propiedad reproductiva tienen que tener la misma probabilidad.

- **La variable $X_2 \sim B(150;0,3)$ por el TCL se aproxima a una distribución normal $N\left(150\,x\,0,3, \sqrt{150\,x\,0,3\,x\,0,7}\right) \equiv N(45, 5,61)$ La variable $X_2 \sim B(150;0,3)$ por el TCL se aproxima a una distribución normal $N\left(150\,x\,0,3, \sqrt{150\,x\,0,3\,x\,0,7}\right) \equiv N(45, 5,61)$**

b) $$E(X)=\int_0^1 x\,f(x)\,dx=\int_0^1 x\,\theta x^{\theta-1}\,dx=\int_0^1 \theta x^{\theta}dx=\theta\left(\frac{x^{\theta+1}}{\theta+1}\right)_0^1=\frac{\theta}{\theta+1}$$

$$E(\overline{x})=\frac{nE(x)}{n}=\frac{\theta}{\theta+1}$$ **no es insesgado**

NOTA: El Teorema Central del Limite (TCL) establece que la suma de n variables aleatorias, con el tamaño n suficientemente grande, se aproxima a una distribución normal. Como toda distribución binomial es suma de n variables aleatorias independientes de Bernouilli se cumple, aplicando el TCL para n suficientemente grande.

Teorema de Moivre-Laplace: Sí X es una variable discreta que sigue una distribución binomial de parámetros n y p, $B(n, p)$ y se cumple que $n>10$, $np>5$ y $nq>5$ resulta una aproximación bastante buena suponer que la variable X' se aproxima a la variable normal $N\left(np, \sqrt{npq}\right)$.

Cuando se aproxima una distribución binomial mediante una normal, se transforma una variable discreta X (toma un número determinado de valores) en una variable continua X' (toma valores en un intervalo) se hace necesaria la corrección de continuidad de Yates.

$P(X=a)=P(a-0,5\leq X'\leq a+0,5)$ $P(X<a)=P(X'\leq a-0,5)$ $P(X>a)=P(X'\geq a+0,5)$

$P(X\leq a)=P(X'\leq a+0,5)$ $P(X\geq a)=P(X'\geq a+0,5)$ $P(a\leq X<b)=P(a-0,5\leq X'\leq b-0,5)$

32. La demanda media de un producto es de 100 unidades con una desviación típica de 40 unidades. Calcular la cantidad del producto que se debe tener a la venta para satisfacer la demanda de forma que puedan ser atendidos al menos el 80% de los clientes.

Solución:

$\mu=100\,,\ \sigma=40$

Según Chebyshev: $P\Big[\,|X-\mu_x|\leq k\,\Big]\geq 1-\dfrac{\sigma^2}{k^2} \longrightarrow P\Big[\,\mu_x-k\leq X\leq \mu_x+k\Big]\geq 1-\dfrac{\sigma^2}{k^2}$

$$P\Big[\,100-k\leq X\leq 100+k\Big]\geq \overbrace{1-\frac{40^2}{k^2}}^{0,80}$$

$$0,80=1-\frac{40^2}{k^2}\ \mapsto\ \frac{40^2}{k^2}=0,20\ \mapsto\ k^2=\frac{40^2}{0,20}\ \mapsto\ k=\sqrt{\frac{40^2}{0,20}}=89,44$$

Se deben poner a la venta 90 unidades.

33. Se desea conocer el número de automóviles que se deben poner a la venta durante un periodo determinado para que se satisfaga una demanda media de 300 unidades con una desviación típica de 100 unidades, con una probabilidad no inferior al 75%.

Solución:

Sea la variable aleatoria X = "número de automóviles a la venta"

$\mu=300\,,\sigma=100$

Según Chebyshev: $P\Big[\,|X-\mu_x|\leq k\,\Big]\geq 1-\dfrac{\sigma^2}{k^2} \longrightarrow P\Big[\,\mu_x-k\leq X\leq \mu_x+k\Big]\geq 1-\dfrac{\sigma^2}{k^2}$

$$P\Big[\,300-k\leq X\leq 300+k\Big]\geq \overbrace{1-\frac{100^2}{k^2}}^{0,75}$$

$$0,75 = 1 - \frac{100^2}{k^2} \mapsto \frac{100^2}{k^2} = 0,25 \mapsto k^2 = \frac{100^2}{0,25} \mapsto k = \sqrt{\frac{100^2}{0,25}} = 200$$

$300 + k = 300 + 200 = 500$ automóviles

34. La función de densidad de una variable aleatoria es:

$$f(x) = \begin{cases} ax^2 + b & 0 < x < 2 \\ 0 & \text{en el resto} \end{cases} \qquad \text{sabiendo que } P\left[\frac{1}{2} < x < 1\right] = 0,1666 .$$

Determinar a y b.

Solución:

Hay que calcular dos parámetros (a y b), por lo que se necesitan dos ecuaciones:

- Por ser función de densidad:

$$1 = \int_0^2 f(x)\,dx = \int_0^2 (ax^2 + b)\,dx = \left[a\frac{x^3}{3} + bx\right]_0^2 = \frac{8a}{3} + 2b \mapsto 8a + 6b = 3$$

- $P\left[\frac{1}{2} \le x \le 1\right] = \int_{1/2}^1 f(x)\,dx = \int_{1/2}^1 (ax^2 + b)\,dx = \left[a\frac{x^3}{3} + bx\right]_{1/2}^1 = 0,1666$, con lo que:

$$\left[a\frac{x^3}{3} + bx\right]_{1/2}^1 = \left[\frac{a}{3} + b\right] - \left[\frac{a}{24} + \frac{b}{2}\right] = \frac{7a}{24} + \frac{b}{2} = 0,1666 \mapsto 7a + 12b \approx 4$$

en consecuencia,

$$\left.\begin{array}{r} 8a + 6b = 3 \\ 7a + 12b = 4 \end{array}\right\} \mapsto \left.\begin{array}{r} 16a + 12b = 6 \\ 7a + 12b = 4 \end{array}\right\} \to \boxed{a = \frac{2}{9} = 0,22} \quad 6b = 3 - \frac{16}{9} = \frac{11}{9} \mapsto \boxed{b = \frac{11}{54} = 0,20}$$

35. Completar la ley de probabilidad , conociendo que la esperanza matemática es 1,8

X	0	1	2	3
$P(X=x_i)=p_i$	0,2	a	b	0,3

Solución:

- $\sum_{i=1}^{4} p_i = 0{,}2 + a + b + 0{,}3 = 1 \quad \mapsto \quad a+b=0{,}5$

- $\sum_{i=1}^{4} x_i \,.\, p_i = a + 2b + 0{,}9 = 1{,}8 \quad \mapsto \quad a+2b=0{,}9$

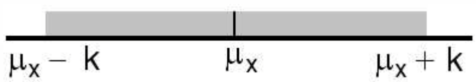
Resolviendo el sistema: $\begin{cases} a+b=0{,}5 \\ a+2b=0{,}9 \end{cases} \Rightarrow \begin{matrix} b=0{,}4 \\ a=0{,}1 \end{matrix}$

36. En un cine de verano hay instaladas 800 sillas, sabiendo que el número de asistentes es una variable aleatoria de media 600 y desviación típica 100.
¿Qué probabilidad existe de que el número de personas que vaya al cine un día cualquiera sea superior al número de sillas instaladas?

Solución:

Sea la variable aleatoria X = "número de sillas del cine", donde $\mu = 600$, $\sigma = 100$

$$P[X > 800] < P\left[\,|X-\mu_x| > k\,\right] \le \frac{\sigma^2}{k^2}$$



$$\mu_x + k = 800 \quad \mapsto \quad k = 800 - 600 = 200$$

$$P[X > 800] \le \frac{100^2}{200^2} = \frac{1}{4} = 0{,}25$$

37. Si a diario e independientemente una variable aleatoria X se distribuye como una Poisson de varianza 7,5. ¿Cómo se distribuirá mensualmente?

Solución:

$$X_i \sim P(\lambda = 7{,}5) \quad \to \quad \sum_{i=1}^{30} X_i \sim P(30\,.\,7{,}5 = 225) \quad \approx \quad N\left[225;\ \sqrt{225} = 15\right]$$

38. Se tienen dos variables aleatorias cuya distribución es normal: $X \sim N(3;2)$, $Y \sim N(5;1)$ Obtenga la esperanza matemática y la varianza de la variable: $W = 2X + 4Y$ y de la variable $Z = (X+Y)/2$

Solución:

$$W = 2X + 4Y \mapsto \begin{cases} E(W) = E(2X+4Y) = 2E(X) + 4E(Y) = 2.3 + 4.5 = 26 \\ V(W) = V(2X+4Y) = 4V(X) + 16V(Y) = 4.2^2 + 16.1^2 = 32 \end{cases}$$

$$Z = \frac{X+Y}{2} \mapsto \begin{cases} E(Z) = E\left(\dfrac{X+Y}{2}\right) = \dfrac{1}{2}E(X+Y) = \dfrac{1}{2}E(X) + \dfrac{1}{2}E(Y) = \dfrac{3}{2} + \dfrac{5}{2} = 4 \\ V(Z) = V\left(\dfrac{X+Y}{2}\right) = \dfrac{1}{4}V(X+Y) = \dfrac{1}{4}V(X) + \dfrac{1}{4}V(Y) = \dfrac{2^2}{4} + \dfrac{1^2}{4} = \dfrac{5}{4} \end{cases}$$

39. Se supone que la rentabilidad de un producto ofrecido por una entidad bancaria es una variable aleatoria que tiene como función de densidad: $f(x) = \begin{cases} k(x^2 - 1) & \text{si } 0 \le x \le 2 \\ 0 & \text{en otro caso} \end{cases}$

Hallar el valor de k para que $f(x)$ sea función de densidad

Solución:

El ejercicio carece de sentido, $f(x)$ no puede ser función de densidad para ningún valor de k:

$$f(x) = \begin{cases} k < 0 & f(x) < 0 & \text{en el intervalo } [1, 2) \\ k = 0 & f(x) = 0 & \forall x \in R \\ k > 0 & f(x) < 0 & \text{en el intervalo } [0, 1) \end{cases}$$

Una función de densidad $f(x)$ es una función real no negativa.

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

(IMPRESO EN PAPEL RECICLADO) UNICAMENTE PARA USO ESCOLAR

40. Marca las respuestas correctas.

1. $H_0: \mu \geq 60$, $H_1: \mu < 60$, $t_{25\,,\,0{,}05} = 2{,}485$ y $t_{25} = -1{,}818$

a) Se acepta la hipótesis nula

b) Se rechaza la hipótesis nula

c) El estadístico muestral se ubica a la derecha del valor crítico

d) Se acepta 1 y 3

e) Ninguna opción es correcta

2. $H_0: \mu = 200$, $H_1: \mu \neq 200$, $z_{0{,}005} = 2{,}576$

a) Se acepta la hipótesis nula si $-2{,}576 \leq z \leq 2{,}576$

b) Se rechaza la hipótesis nula si $-2{,}576 \leq z \leq 2{,}576$

c) $z = (\overline{x} - 200) / (\sigma / \sqrt{n})$

d) $\alpha = 0{,}01$

e) Ninguna opción es correcta

3. $H_0: \mu = 60$, $H_1: \mu \neq 60$, $\alpha = 0{,}05$

a) Se acepta la hipótesis nula si p-valor = 0,081

b) Se rechaza la hipótesis nula si p-valor = 0,04

c) p-valor = P[rechazar el estadístico muestral observado / H_0 cierta]

d) Se acepta H_0 cuando p-valor > α

e) Ninguna opción es correcta

4. Si el contraste $H_0: \mu = 20$, $H_1: \mu \neq 20$ se acepta con un p-valor = 0,06, el contraste $H_0: \mu \leq 20$, $H_1: \mu > 20$ tiene:

a) p-valor = 0,06

b) Contraste unilateral cola a la derecha

c) p-valor = 0,03

d) p-valor = 0,97

e) Ciertas 2 y 3

41. Calcular la media y la varianza de una variable aleatoria t_5 de Student

Solución:

Una variable aleatoria t_n de Student tiene de media $\mu = 0$ y varianza $\sigma^2 = \dfrac{n}{n-2}$, con lo cual la media y la varianza de una t_5 de Student, respectivamente, son $\mu = 0$ y $\sigma^2 = \dfrac{5}{3}$

42. Encontrar la moda de una variable aleatoria $X \sim B(14, 0{,}2)$

Solución:

La moda de una distribución binomial viene dada por el valor (número entero) que verifica $(np - q \leq M_d \leq np + p)$. Generalmente será un valor (la parte entera de la media) y podrán ser dos valores modales cuando $(np - q)$ y $(np + p)$ sea un número entero.

En este caso, $np - q = 14.0{,}2 - 0{,}8 = 2 \quad np + p = 14.0{,}2 + 0{,}2 = 3$

$$np - q \leq M_d \leq np + p \quad \rightarrow \quad 2 \leq M_d \leq 3$$

La distribución es bimodal y las modas son 2 y 3

43. Se consideran dos variables aleatorias independientes X e Y. La variable X tiene una distribución normal $N(0, 1)$, la variable Y tiene una distribución χ^2 con 4 grados de libertad.

Calcular el valor de m, siendo $P\left[\dfrac{2X}{\sqrt{Y}} \leq m\right] = 0{,}05$

Solución:

$$X \sim N(0,1) \quad Y \sim \chi_4^2 \quad t_4 = \frac{z}{\sqrt{\dfrac{1}{4}\chi_4^2}} = \frac{2z}{\sqrt{\chi_4^2}}$$

$$P\left[\frac{2X}{\sqrt{Y}} \leq m\right] = P\left[t_4 \leq m\right] = 0{,}05 = P\left[t_4 \geq -m\right] = 0{,}05 \quad \rightarrow \quad m = -2{,}132$$

44. De la distribución $N(5, 2)$ se toma una muestra aleatoria simple de tamaño 20, y de la distribución $N(-4, 10)$ otra muestra aleatoria simple de tamaño 25 e independiente de la primera. Calcular $P[(\overline{x}-\overline{y})\leq 15]$

Solución:

Considerando como estadístico la diferencia de medias muestrales, siendo el tamaño de la primera muestra $n=20$ y de la segunda muestra $m=25$

$$(\overline{x}-\overline{y}) \sim N\left[\mu_1-\mu_2\,,\ \sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right] \equiv N\left[5-(-4)\,,\ \sqrt{\frac{2^2}{20}+\frac{10^2}{25}}\right] \equiv N[9\,,\ 2{,}05]$$

$$P[(\overline{x}-\overline{y})<15] = P\left[\frac{(\overline{x}-\overline{y})-9}{2{,}05}<\frac{15-9}{2{,}05}\right] = P[z<2{,}93] = 1-P[z>2{,}93] = 1-0{,}00169 = 0{,}9983$$

45. ¿Cuándo se utiliza la desigualdad de Chebychev para obtener intervalos de confianza?. Razonar la respuesta.

Solución:

Cuando se desea obtener un intervalo de confianza para la media y se desconoce la distribución de la población pero se conoce la varianza.

En efecto, para una muestra de tamaño n de una población con media μ desconocida y varianza σ^2 conocida, utilizando como estimador de μ la media muestral $\overline{X}$, se sabe que $E[\overline{X}]=\mu$ y $Var[\overline{X}]=\dfrac{\sigma^2}{n}$

Sustituyendo en la desigualdad de Chebychev: $P\left[\overline{X}-E(\overline{X})\leq k\right]\geq 1-\dfrac{Var(\overline{X})}{k^2}$

Resulta, $P\left[\overline{X}-\mu\leq k\right]\geq 1-\dfrac{\sigma^2}{nk^2}$ que equivale a $P\left[\overline{X}-k\leq\mu\leq\overline{X}+k\right]\geq 1-\dfrac{\sigma^2}{nk^2}$

La expresión $1-\dfrac{\sigma^2}{nk^2}=1-\alpha$ es el coeficiente de confianza, despejando: $k=\dfrac{\sigma}{\sqrt{n\alpha}}$

El intervalo de confianza buscado sería: $\left[\overline{X}-\dfrac{\sigma}{\sqrt{n\alpha}}\,,\ \overline{X}+\dfrac{\sigma}{\sqrt{n\alpha}}\right]$

46. Se quiere comparar las puntuaciones en estadística teórica de dos Universidades que siguen, respectivamente, una distribución normal, con varianzas desconocidas y se supone que estadísticamente iguales. Se toman al azar dos muestras iguales de estudiantes de ambas Universidades. Los resultados obtenidos en SPSS son:

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ de la diferencia	95% Intervalo de confianza para la diferencia: Inferior	Superior
Calificación	Se han asumido varianzas iguales	,152	,699	,443	28	,661	2,200	4,962	−7,964	12,364

Con un nivel de significación de 0,05, se pide:

a) ¿Existe diferencia significativa sobre las puntuaciones de estadística teórica en las dos Universidades?

b) ¿Cuál es la media ponderada de las cuasivarianzas muestrales?

c) Cuál sería el p-valor si se hubieran contrastado las puntuaciones

$$H_0 : \mu_{\text{Autónoma}} \leq \mu_{\text{Complutense}} \qquad H_1 : \mu_{\text{Autónoma}} > \mu_{\text{Complutense}}$$

Solución:

a) En los resultados obtenidos en SPSS la Prueba de Levene permite decidir si las varianzas poblacionales son iguales. Como la probabilidad asociada al estadístico de Levene (0,699) es mayor que 0,05 se acepta la hipótesis nula de que las varianzas poblacionales son iguales.

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm t_{\alpha/2,\,(n_1+n_2-2)} \cdot \underbrace{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}_{\text{error típico diferencia}} \right]$$

Las dos Universidades tienen la misma puntuación en estadística teórica a un nivel de significación del 0,05 ya que el intervalo de confianza cubre el cero.

Un intervalo de confianza sirve para tomar una decisión sobre la hipótesis nula que permite contrastar el estadístico T (t de Student).

Por otra parte,

$p-\text{valor} = \text{Sig.(bilateral)} = 0{,}661 > 0{,}05 \quad \rightarrow$ **Se acepta la hipótesis nula que establece** $H_0 : \mu_{\text{Autónoma}} = \mu_{\text{Complutense}}$

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....
Apellidos .... Nombre ....
Ejercicio del día ....

b) $s_p^2 \equiv$ **Media ponderada cuasivarianzas muestrales:** $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

Error típico de la diferencia $\equiv s_p \cdot \sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}} \;\rightarrow\; 4,962 = s_p \cdot \sqrt{\dfrac{1}{15}+\dfrac{1}{15}}$

$$s_p = \frac{4,962}{\sqrt{\dfrac{1}{15}+\dfrac{1}{15}}} = \frac{4,962}{0,3651} = 13,588 \;\rightarrow\; s_p^2 = 184,633$$

$t \equiv$ **valor experimental del estadístico de contraste:**

$$t = \frac{\text{Diferencia de medias}}{\text{Error típico diferencia}} = \frac{|\bar{x}-\bar{y}|}{13,588 \cdot \sqrt{\dfrac{1}{15}+\dfrac{1}{15}}} = \frac{2,200}{4,962} = 0,443$$

$t = 0,443 \leq t_{0,025\,,\,28} = 2,048$ **se acepta** H_0

$p-$**valor** $=$ **Sig.(bilateral)** $= 0,661$ **. La Significación bilateral muestra el grado de compatibilidad entre el valor poblacional propuesto y la información muestral disponible.**

Si el nivel crítico es pequeño (generalmente menor que 0,05), se concluye que la información recogida en la muestra es incompatible con la hipótesis nula de que las medias poblacionales son iguales.

La Prueba T se basa en el supuesto de normalidad, el cumplimiento de este supuesto sólo es exigible con muestras pequeñas. El supuesto de normalidad carece de relevancia en muestras grandes.

c) SPSS siempre calcula el p-valor para un contraste bilateral o de dos colas, si se desea realizar un contraste unilateral o de una cola hay que dividir entre 2 el contraste bilateral $(p-\text{valor} = 0,699 / 2 = 0,3495)$

Contraste unilateral (cola a la derecha: $H_1 : \mu_{\text{Autónoma}} > \mu_{\text{Complutense}}$**)**

$p-$**valor** $=$ **Sig.(unilateral derecha)** $= 0,3495 > 0,05 \;\Rightarrow$ **Se acepta** H_0

47. Calcula el estimador máximo verosímil del parámetro λ de una distribución de Poisson. Comprueba la eficiencia.

Solución:

En la distribución de Poisson: $P(X=x) = \dfrac{\lambda^x}{x!}\,e^{-\lambda} \quad E(x)=\lambda \quad V(x)=\lambda$

La función de verosimilitud $L(X,\lambda)$ en una muestra aleatoria simple de tamaño n:

$$L(\lambda) = L(X,\lambda) = \prod_{i=1}^{n} P(x_i,\lambda) = \frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\cdots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\,e^{-n\lambda}$$

$$L(X,\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\,e^{-n\lambda} \Rightarrow \ln L(X,\lambda) = \ln\left[\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\,e^{-n\lambda}\right] =$$

$$= \ln\left(\lambda^{\sum_{i=1}^{n} x_i}\right) - \ln\left(\prod_{i=1}^{n} x_i!\right) + \ln(e^{-n\lambda}) = \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln(x_i!) - n\lambda$$

$$\ln L(X,\lambda) = \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln(x_i!) - n\lambda$$

Para obtener el estimador de máxima verosimilitud $EMV(\hat{\lambda})$, se deriva la expresión anterior respecto del parámetro para obtener, sustituyendo λ por $\hat{\lambda}$

$$\left.\frac{\vartheta \ln L(X,\lambda)}{\vartheta\lambda}\right|_{\lambda=\hat{\lambda}} = 0 \equiv \left.\sum_{i=1}^{n} x_i \frac{1}{\lambda} - n\right|_{\lambda=\hat{\lambda}} = 0 \Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

El estimador de máxima verosimilitud (estimador que ofrece mayor credibilidad) viene dado por la media muestral $EMV(\hat{\lambda}) = \overline{x}$

Para que un estimador sea *eficiente* tiene que ser *centrado* y de *varianza mínima*. La varianza mínima se analiza en virtud de la acotación de Cramer-Rao:

$$V(\hat{\lambda}) \geq CCR = \frac{\left[1+b(\hat{\lambda})\right]^2}{n.E\left[\frac{\vartheta \ln P(x\ ;\ \lambda)}{\vartheta\lambda}\right]^2} \overset{b(\hat{\lambda})=0}{\mapsto} V(\hat{\lambda}) \geq CCR = \frac{1}{n.E\left[\frac{\vartheta \ln P(x\ ;\ \lambda)}{\vartheta\lambda}\right]^2}$$

El estimador es eficiente cuando $V(\hat{\lambda}) = CCR$

- **Insesgadez y Varianza**

El estimador $\hat{\lambda}$ es insesgado (centrado) sí $E(\hat{\lambda}) = \lambda$

$$E(\hat{\lambda}) = E\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\lambda) = \lambda$$

$$V(\hat{\lambda}) = V(\overline{x}) = V\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}(n\lambda) = \frac{\lambda}{n}$$

- **Consistencia**

Cuando no es posible emplear estimadores de máxima verosimilitud, el requisito mínimo deseable para un estimador es que sea consistente.

El estimador $\hat{\lambda}$ es *consistente* cuando $\lim_{n\to\infty} E(\hat{\lambda}) = \lambda$ y $\lim_{n\to\infty} V(\hat{\lambda}) = 0$

$$\lim_{n\to\infty} E(\hat{\lambda}) = \lim_{n\to\infty} \lambda = \lambda \quad \text{y} \quad \lim_{n\to\infty} V(\hat{\lambda}) = \lim_{n\to\infty} \frac{\lambda}{n} = 0 \ \mapsto \ \text{Estimador } \hat{\lambda} \text{ es consistente}$$

Siendo, $P(X=x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$

$$\ln P(x,\lambda) = \ln\left[\frac{\lambda^x}{x!} e^{-\lambda}\right] = x\ln\lambda - \ln(x!) - \lambda$$

$$\frac{\vartheta \ln P(x,\lambda)}{\vartheta\lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}$$

$$E\left[\frac{\vartheta \ln P(x,\lambda)}{\vartheta\lambda}\right]^2 = E\left[\frac{x-\lambda}{\lambda}\right]^2 = \frac{1}{\lambda^2} E(x-\lambda)^2 = \frac{1}{\lambda^2} E(x-\overline{x})^2 = \frac{1}{\lambda^2} V(x) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

En consecuencia, $V(\hat{\lambda}) \geq \dfrac{1}{n\,\dfrac{1}{\lambda}} = \dfrac{\lambda}{n}$

El menor valor de la varianza del estimador será $\dfrac{\lambda}{n}$

Se sabe que $V(\hat{\lambda}) = V(\overline{x}) = \dfrac{\lambda}{n}$, lo que muestra $V(\hat{\lambda}) = \text{CCR}$, el estimador empleado es *eficiente*.

48. Sea la distribución $N(\mu, \sigma)$, con la media y varianza desconocidas. Calcular los estimadores máximo-verosímiles de μ y σ^2

Solución:

La función de verosimilitud es:

$$L(X;\mu,\sigma^2) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right]\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_2-\mu)^2}{2\sigma^2}}\right]\cdots\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right] =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\,e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

tomando logaritmos neperianos, se tiene:

$$\text{Ln}\left[L(X;\mu,\sigma^2)\right] = \text{Ln}\left[\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\,e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right] = -\frac{n}{2}\text{Ln}(2\pi) - \frac{n}{2}\text{Ln}(\sigma^2) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

y derivando respecto de μ y σ^2 e igualando a cero:

$$\left[\frac{\vartheta \text{Ln}\, L(X;\mu,\sigma^2)}{\vartheta\mu}\right]_{\mu=\hat{\mu}} = \frac{\sum_{i=1}^{n}(x_i-\hat{\mu})}{\sigma^2} = 0 \;\rightarrow\; \hat{\mu} = \frac{\sum_{i=1}^{n}x_i}{n} = \overline{x}$$

$$\left[\frac{\vartheta \text{Ln}\, L(X;\mu,\sigma^2)}{\vartheta\mu}\right]_{\sigma^2=\hat{\sigma}^2} = -\frac{n}{\hat{\sigma}} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\hat{\sigma}^3} = 0 \;\rightarrow\; \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

resolviendo el sistema resulta: $\hat{\mu} = \overline{x}$ **y** $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} = \sigma_x^2$

Los estimadores máximo-verosímiles de μ y σ^2 son la media y la varianza muestrales.

49.

a) ¿Cuál es el objetivo de la estimación por intervalos de confianza?.

b) ¿Para qué se utiliza el error cuadrático medio de un estimador?

c) ¿Cuál es la interpretación del concepto 'grados de libertad' al utilizar estimadores?

Solución:

a) Establecer un intervalo de poca amplitud y alta probabilidad (coeficiente de confianza) de modo que en su interior se encuentre un determinado parámetro de la distribución de la variable aleatoria.

b) $ECM(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = Var\left[\hat{\theta}\right] + \left(E(\hat{\theta}) - \theta\right)^2 = Var(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$

En el problema de la estimación puntual interesa que el error cuadrático medio sea lo menor posible, lo cual se consigue cuanto menores sean la varianza del estimador $Var(\hat{\theta})$ y el valor absoluto del sesgo $|b(\hat{\theta})|$

Si el estimador es insesgado $b(\hat{\theta}) = 0$, el error cuadrático medio $ECM(\hat{\theta}) = Var(\hat{\theta})$ coincide con la varianza.

c) Si en una muestra aleatoria de tamaño n, las n variables $X_1, X_2, \ldots, X_n$ son independientes, se suele decir que el conjunto de las n variables $(X_1, X_2, \ldots, X_n)$ contiene n grados de libertad.
Ahora bien, es posible que un estimador cualquiera $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ mantenga o no los n grados de libertad.
Por ejemplo, si en una población en la que se desconoce la media poblacional μ, se utiliza como estimador $\hat{\theta} = \sum_{i=1}^{n} (x_i - \overline{x})^2$ (por ejemplo para hacer estimaciones sobre la varianza), ocurre que se ha perdido un grado de libertad puesto que se sabe que $\sum_{i=1}^{n} x_i = n\overline{x}$, concluyendo que conociendo solamente $(n-1)$ valores de la muestra se puede despejar el valor que queda. Así pues, $\hat{\theta}$ posee $(n-1)$ grados de libertad.

Asignatura.. Grupo..............................
Apellidos .... Nombre..............................
Ejercicio del día ....

50.

a) ¿Qué es una hipótesis estadística?. ¿Qué es la hipótesis nula?. Razonar la respuesta.

b) Explicar conceptualmente porqué es importante que un estimador sea eficiente.

c) Explicar conceptualmente qué es el sesgo de un estimador y cuál es su interpretación.

Solución:

a) Una hipótesis estadística es una afirmación verdadera o falsa acerca del valor de alguna característica desconocida de la población.
Para efectuar un contraste de hipótesis, se acepta una hipótesis como verdadera, a la que se denomina hipótesis nula, frente a otra complementaria que se conoce como hipótesis alternativa.

b) Un estimador es eficiente si es insesgado y su varianza alcanza la cota de Cramer-Rao, es decir, tiene menor varianza que cualquier otro estimador insesgado. Ello es importante porque, bajo la hipótesis de eficiencia, el estimador toma, para diferentes muestras, valores próximos unos a otros.

c) Se denomina sesgo de un estimador $\hat{\theta}$ a la diferencia $E(\hat{\theta}-\theta)$, donde θ es el parámetro a estimar. Puede ser positivo, negativo o nulo. Una propiedad deseable del estimador es que el sesgo sea nulo y en ese caso diremos que el estimador es insesgado. En caso contrario el estimador sobreestima o infraestima al prámetro según que el sesgo sea positivo o negativo respectivamente.

Hoja nº ....

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

UNICAMENTE PARA USO ESCOLAR (IMPRESO EN PAPEL RECICLADO)