PORTAL ESTADÍSTICA APLICADA

Vídeo Contrastes | Algebra Lineal | Decisiones en Bolsa | Matrices Determinantes | Métodos Integración | Ecuaciones Diferenciales

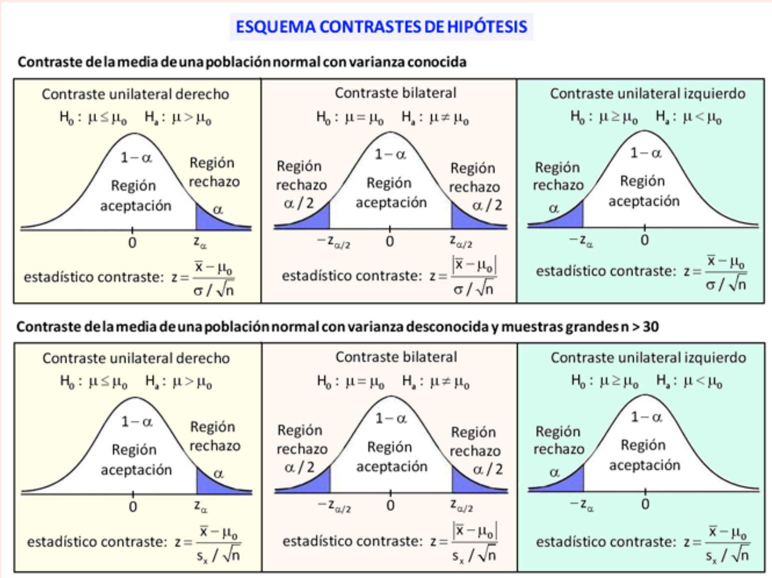
PAU MATEMÁTICAS

Destrezas Matemáticas

Sistema Ecuaciones | Programación Lineal | Matrices Determinantes | Análisis Matemático | Geometría | Estadística Probabilidad

ESTADÍSTICA Y PROBABILIDAD

Si te fijas en la analogía con los intervalos de confianza y observas los gráficos



Teorema Central del Límite

Distribuciones muestrales

Una muestra grande $n > 30$ de una población que no sigue una $N(\mu, \sigma)$, por el Teorema Central del Limite, puede considerarse que sigue una distribución normal.	En una población $N(\mu, \sigma)$, la media $\bar{x}$ de una muestra de tamaño n se distribuye según una normal $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$, es decir, $\bar{x} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$	Para una muestra grande $n > 30$, la distribución muestral de la proporción $\hat{p} \sim N\left(p, \sqrt{\dfrac{p.q}{n}}\right)$

Si una variable aleatoria $X \sim N(\mu, \sigma)$, la suma de n variables aleatorias sigue una distribución normal $\sum_{i=1}^{n} X \sim N(n.\mu, \sqrt{n}.\sigma)$	Sean las variables aleatorias $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$, independientes entre sí, la nueva variable aleatoria $X \pm Y \sim N\left(\mu_1 \pm \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$

Intervalo de Confianza para la media μ de una distribución normal $N(\mu, \sigma)$ de varianza conocida σ^2	Intervalo de Confianza para la media μ de una distribución normal de varianza desconocida, con muestras grandes, $n > 30$	Intervalo de Confianza para la media μ de una distribución normal de varianza desconocida, con muestras pequeñas, $n \leq 30$
$I_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}\right]$	$I_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{\alpha/2} \dfrac{s_x}{\sqrt{n}}\right]$	$I_{1-\alpha}(\mu) = \left[\bar{x} \pm t_{\alpha/2,\, n-1} \dfrac{s_x}{\sqrt{n}}\right]$

Intervalo de Confianza para la diferencia de medias $\mu_1 - \mu_2$ de dos distribuciones normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ con varianzas conocidas

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

Intervalo de Confianza para la diferencia de medias $\mu_1 - \mu_2$ de dos distribuciones normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ con varianzas desconocidas, y muestras grandes $n_1 + n_2 > 30$

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$$

Intervalo de Confianza para la diferencia de medias $\mu_1 - \mu_2$ de dos distribuciones normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ con varianzas desconocidas pero iguales, en muestras pequeñas $n_1 + n_2 \leq 30$

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm t_{\alpha/2,\, n_1+n_2-2}\; s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

media ponderada de las cuasivarianzas muestrales

$$s_p^2 = \frac{(n_1 - 1)\,.\,s_1^2 + (n_2 - 1)\,.\,s_2^2}{n_1 + n_2 - 2}$$

Intervalo de Confianza para la diferencia de medias $\mu_1 - \mu_2$ de dos distribuciones normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ con varianzas desconocidas y distintas, en muestras pequeñas $n_1 + n_2 \leq 30$

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm t_{\alpha/2,\, f} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$$

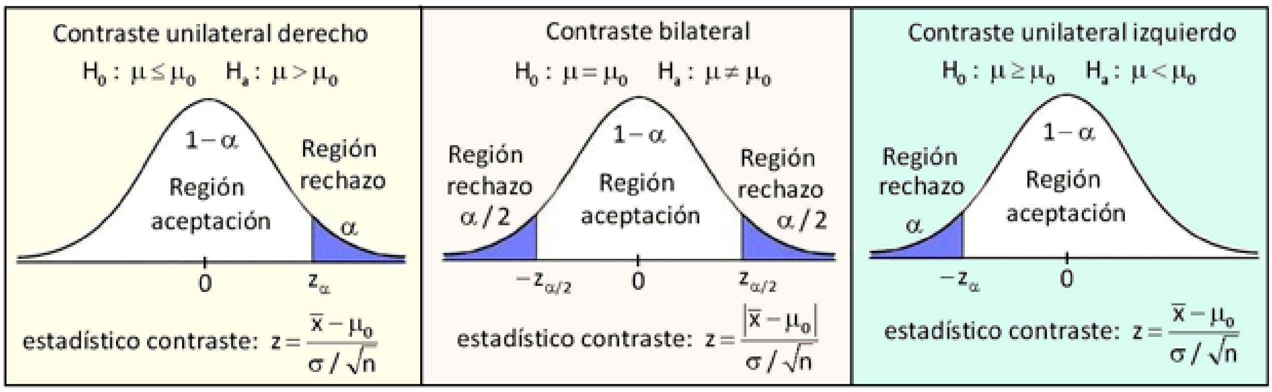
Aproximación de Welch

$$f = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}} - 2$$

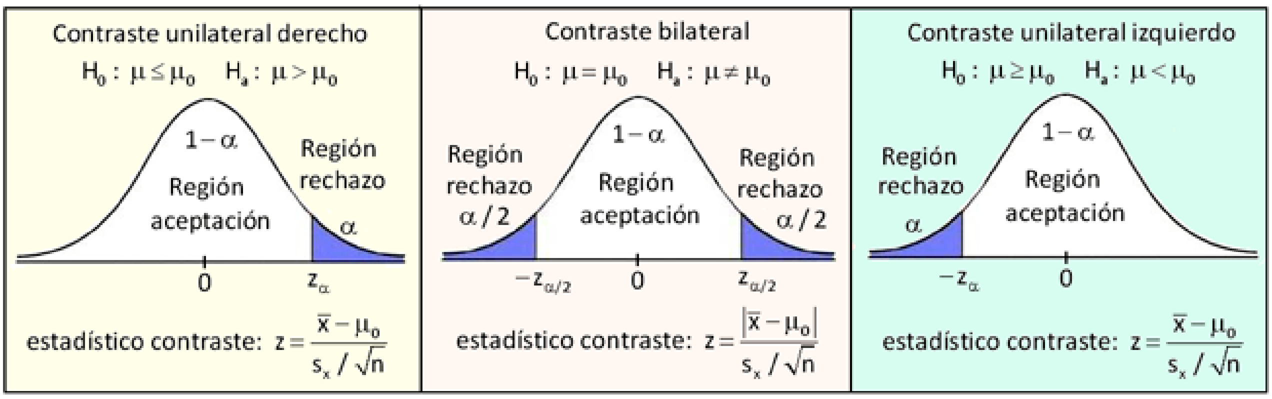
Intervalo de Confianza para la proporción de una distribución binomial de parámetros n , p , B(n, p)

$$I_{1-\alpha}(p) = \left[\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\,.\,\hat{q}}{n}}\right]$$

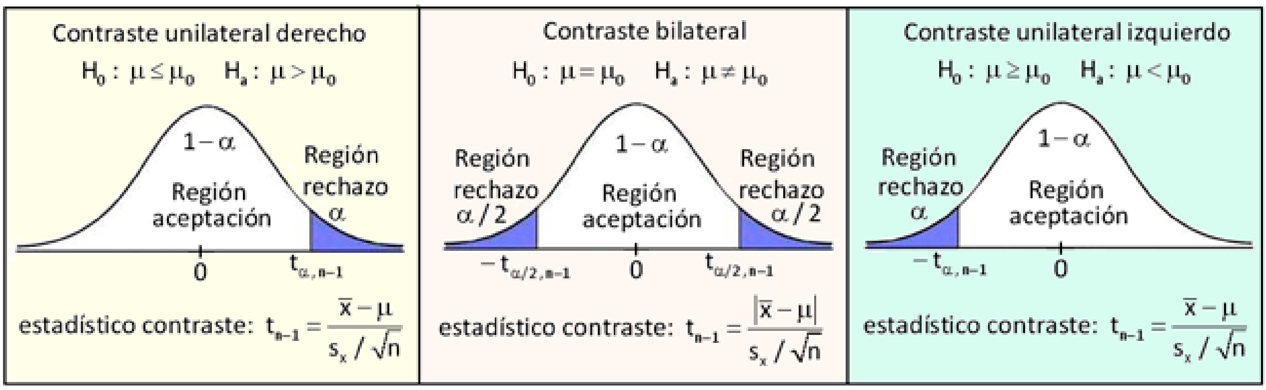
Contraste de la media de una población normal con varianza conocida



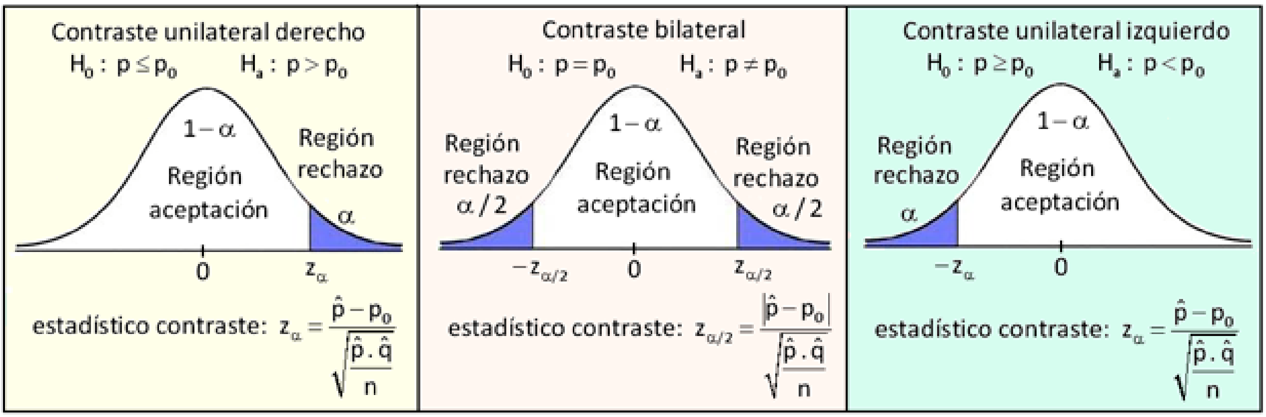
Contraste de la media de una población normal con varianza desconocida y muestras grandes n > 30



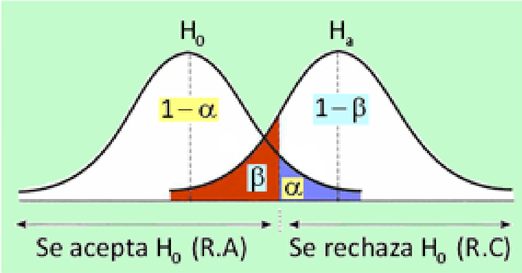
Contraste de la media de una población normal con varianza desconocida y muestras pequeñas n ≤ 30



Contraste para la proporción de una distribución binomial



TIPOS DE ERROR. POTENCIA DEL CONTRASTE



	H_0 verdadera	H_0 falsa
Se acepta H_0	Decisión correcta Potencia $P = 1-\alpha$	Error Tipo II Potencia $P = \beta$
Se rechaza H_0	Error Tipo I Potencia $P = \alpha$	Decisión correcta Potencia $P = 1-\beta$

$$\alpha = P(\text{ET I}) = P\left[\text{Rechazar } H_0 \,/\, H_0 \text{ es verdadera}\right]$$

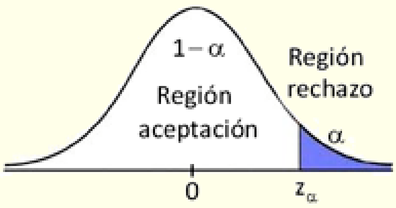
$$\beta = P(\text{ET II}) = P\left[\text{Aceptar } H_0 \,/\, H_0 \text{ falsa}\right]$$

$$\text{Potencia} = P\left[\text{Aceptar } H_0 \,/\, H_0 \text{ verdadera}\right] = 1-\alpha \qquad \text{Potencia} = P\left[\text{Rechazar } H_0 \,/\, H_0 \text{ falsa}\right] = 1-\beta$$

Contraste para la diferencia de medias de dos distribuciones normales $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$ con varianzas poblacionales σ_1^2 y σ_2^2 conocidas, donde $X - Y \sim N\left(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$

Contraste unilateral derecho

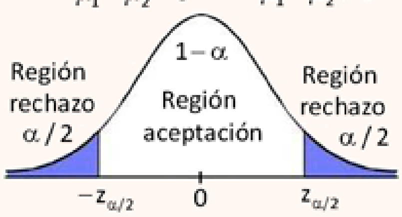
$H_0: \mu_1 \le \mu_2 \qquad H_a: \mu_1 > \mu_2$



estadístico contraste: $z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Contraste bilateral

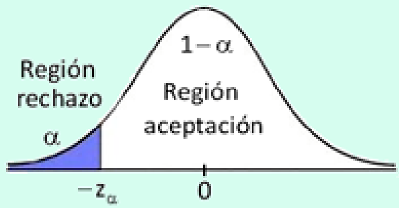
$H_0: \mu_1 = \mu_2 \quad \mu_1 - \mu_2 = 0 \qquad H_a: \mu_1 \ne \mu_2 \quad \mu_1 - \mu_2 \ne 0$



estadístico contraste: $z = \dfrac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Contraste unilateral izquierdo

$H_0: \mu_1 \ge \mu_2 \qquad H_a: \mu_1 < \mu_2$

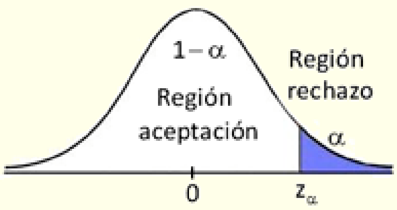


estadístico contraste: $z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Contraste para la diferencia de medias de dos distribuciones normales $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$ con varianzas poblacionales desconocidas, y muestras grandes $n_1 + n_2 > 30$, $n_1 \simeq n_2$

Contraste unilateral derecho

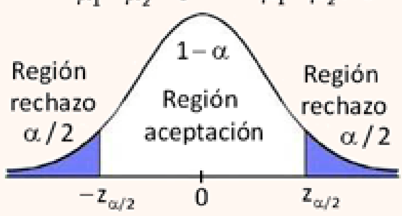
$H_0: \mu_1 \le \mu_2 \qquad H_a: \mu_1 > \mu_2$



estadístico contraste: $z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Contraste bilateral

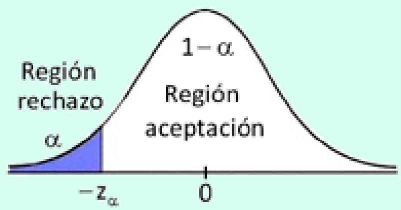
$H_0: \mu_1 = \mu_2 \quad \mu_1 - \mu_2 = 0 \qquad H_a: \mu_1 \ne \mu_2 \quad \mu_1 - \mu_2 \ne 0$



estadístico contraste: $z = \dfrac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Contraste unilateral izquierdo

$H_0: \mu_1 \ge \mu_2 \qquad H_a: \mu_1 < \mu_2$

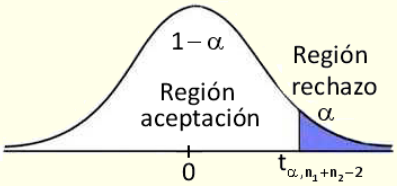


estadístico contraste: $z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Contraste para la diferencia de medias de dos distribuciones normales $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$ con varianzas poblacionales desconocidas pero iguales $\sigma_1^2 = \sigma_2^2$, y muestras pequeñas $n_1 + n_2 \leq 30$

Contraste unilateral derecho

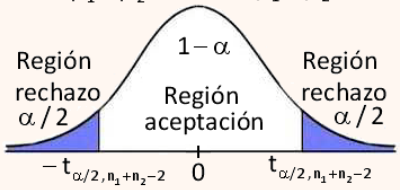
$H_0: \mu_1 \leq \mu_2 \qquad H_a: \mu_1 > \mu_2$



estadístico contraste: $t_{n_1+n_2-2} = \dfrac{\bar{x} - \bar{y}}{s_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

Contraste bilateral

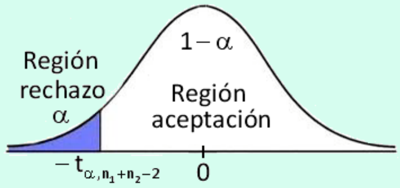
$H_0: \mu_1 = \mu_2 \quad \mu_1 - \mu_2 = 0 \qquad H_a: \mu_1 \neq \mu_2 \quad \mu_1 - \mu_2 \neq 0$



estadístico contraste: $t_{n_1+n_2-2} = \dfrac{|\bar{x} - \bar{y}|}{s_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

Contraste unilateral izquierdo

$H_0: \mu_1 \geq \mu_2 \qquad H_a: \mu_1 < \mu_2$



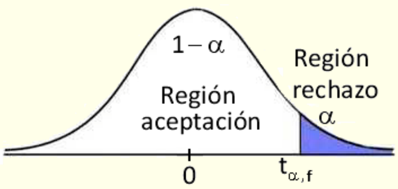
estadístico contraste: $t_{n_1+n_2-2} = \dfrac{\bar{x} - \bar{y}}{s_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

media ponderada de las cuasivarianzas muestrales $s_p^2 = \dfrac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}$

Contraste para la diferencia de medias de dos distribuciones normales $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$ con varianzas poblacionales desconocidas y distintas $\sigma_1^2 \neq \sigma_2^2$

Contraste unilateral derecho

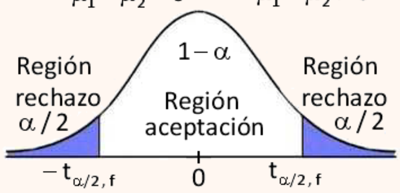
$H_0: \mu_1 \leq \mu_2 \qquad H_a: \mu_1 > \mu_2$



estadístico contraste: $t_{n_1+n_2-2} = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Contraste bilateral

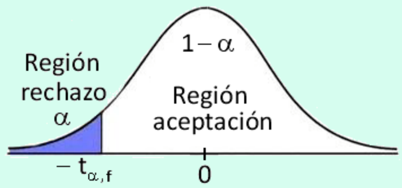
$H_0: \mu_1 = \mu_2 \quad \mu_1 - \mu_2 = 0 \qquad H_a: \mu_1 \neq \mu_2 \quad \mu_1 - \mu_2 \neq 0$



estadístico contraste: $t_{n_1+n_2-2} = \dfrac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Contraste unilateral izquierdo

$H_0: \mu_1 \geq \mu_2 \qquad H_a: \mu_1 < \mu_2$



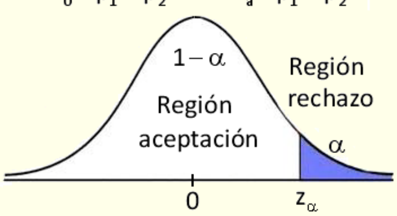
estadístico contraste: $t_{n_1+n_2-2} = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

aproximación de Welch $f = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 + 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 + 1}} - 2$

Contraste para la igualdad de proporciones de dos distribuciones binomiales $B_1(n_1, p_1)$ y $B_2(n_2, p_2)$ en muestras grandes

Contraste unilateral derecho

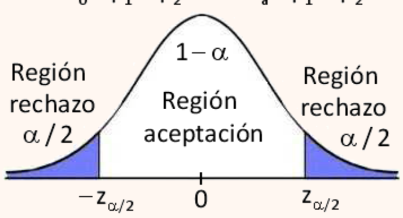
$H_0: p_1 \leq p_2 \qquad H_a: p_1 > p_2$



estadístico contraste: $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}}$

Contraste bilateral

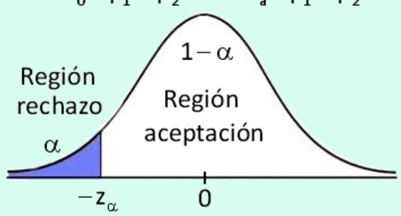
$H_0: p_1 = p_2 \qquad H_a: p_1 \neq p_2$



estadístico contraste: $z = \dfrac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\dfrac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}}$

Contraste unilateral izquierdo

$H_0: p_1 \geq p_2 \qquad H_a: p_1 < p_2$

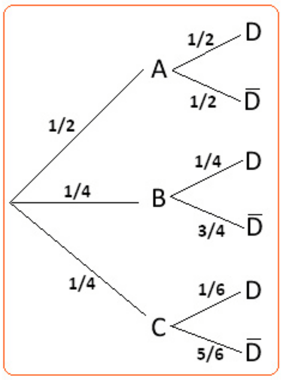


estadístico contraste: $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}}$

PAU: ESTADÍSTICA Y PROBABILIDAD

1. Una fábrica de coches tiene tres cadenas distintas de producción que fabrican, respectivamente, 1/2 , 1/4, 1/4 del total de los coches producidos. La probabilidad de que un coche sea defectuoso es: para la primera cadena 1/2, para la segunda 1/4 y para la tercera 1/6.
a) ¿Cuál es la probabilidad de que un coche sea defectuoso?
b) Si un coche es defectuoso, ¿cuál es la probabilidad de que sea de la primera cadena?

Solución:



a) $P(D) = P(A).P(D/A) + P(B).P(D/B) + P(C).P(D/C) =$

$$= \frac{1}{2}.\frac{1}{2} + \frac{1}{4}.\frac{1}{4} + \frac{1}{4}.\frac{1}{6} = \frac{1}{4} + \frac{1}{16} + \frac{1}{24} = \frac{17}{48}$$

$$P(\overline{D}) = 1 - P(D) = 1 - \frac{17}{48} = \frac{31}{48}$$

$$P(\overline{D}/A) = \frac{1}{2} \qquad P(\overline{D}/B) = \frac{3}{4} \qquad P(\overline{D}/C) = \frac{5}{6}$$

b) $$P(A/\overline{D}) = \frac{P(A).P(\overline{D}/A)}{P(A).P(\overline{D}/A) + P(B).P(\overline{D}/B) + P(C).P(\overline{D}/C)} =$$

$$= \frac{1/2.1/2}{1/2.1/2 + 1/4.3/4 + 1/4.5/6} = \frac{1/4}{1/4 + 3/16 + 5/24} = \frac{12}{31}$$

2. Un examen de opción múltiple está compuesto por 8 preguntas, con cuatro respuestas posibles cada una, de las cuales sólo una es correcta. Supóngase que uno de los estudiantes que realiza el examen responde al azar. ¿Cuál es la probabilidad de que conteste correctamente a 5 o más preguntas?. ¿Cuál es la probabilidad de que no acierte ninguna?

Solución:

Se trata de una distribución binomial $B(8;\ 0,25)$, donde el número de preguntas $n = 8$ y la probabilidad de acierto $p = 1/4 = 0,25$, por tanto $q = 0,75$

Sea $X =$ 'número de respuestas acertadas'

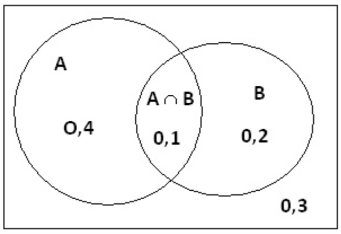
$$P(X \geq 5) = \binom{8}{5}.0,25^5.0,75^3 + \binom{8}{6}.0,25^6.0,75^2 + \binom{8}{7}.0,25^7.0,75 + \binom{8}{8}.0,25^8 = 0,027$$

$$P(\text{no acertar ninguna}) = 0,75^8 = 0,1$$

3. Sean A y B dos sucesos tales que $P(A)=0,5$, $P(B)=0,3$ y $P(A\cap B)=0,1$
Calcular las siguientes probabilidades:

$P(A/B)$ $\quad$ $P(A/A\cap B)$ $\quad$ $P(A\cap B/A\cup B)$ $\quad$ $P(A/A\cup B)$

Solución:



$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{0,1}{0,3}=\frac{1}{3}$$

$$P(A/A\cap B)=\frac{P(A\cap A\cap B)}{P(A\cap B)}=\frac{P(A\cap B)}{P(A\cap B)}=1$$

$$P(A\cap B/A\cup B)=\frac{P\left[(A\cap B)\cap(A\cup B)\right]}{P(A\cup B)}=\frac{P(A\cap B)}{P(A\cup B)}=\frac{0,1}{0,7}=\frac{1}{7}$$

$$P(A/A\cup B)=\frac{P\left[A\cap(A\cup B)\right]}{P(A\cup B)}=\frac{P(A)}{P(A\cup B)}=\frac{0,5}{0,7}=\frac{5}{7}$$

4. Considérense los siguientes datos: 9, 11, 7, 12, 11
a) Calcular la media, varianza y desviación típica.
b) Considérese también el conjunto de datos obtenidos sumando 20 a cada uno de los datos iniciales. Razonar, sin efectuar nuevos cálculos, cuál de los dos conjuntos está más disperso respecto de su media.

Solución:

a) Media: $\overline{x}=\dfrac{\sum_{i=1}^{5}x_i}{N}=\dfrac{x_1+x_2+x_3+x_4+x_5}{N}=10$

Varianza: $\sigma^2=\dfrac{\sum_{i=1}^{5}(x_i-\overline{x})^2}{N}=3,2$

Desviación típica: $\sigma=\sqrt{\sigma^2}=\sqrt{3,2}=1,79$

b) Al sumar o restar la misma constante a todos los datos, la *varianza permanece invariante*, mientras que la media y el coeficiente de variación de Pearson varían.

- $E(x)=\overline{x} \quad \mapsto \quad E(k+x)=E(k)+E(x)=k+\overline{x}$
- $Var(x)=\sigma^2 \quad \mapsto \quad \sigma^2_{k+x}=Var(k+x)=Var(k)+Var(x)=0+\sigma^2_x$

- $\sigma_{k+x} = \sqrt{\sigma^2_{k+x}} = \sqrt{\sigma^2_x} = \sigma_x$

- Coeficiente variación Pearson:

$$C.V_x = \frac{\sigma_x}{\overline{x}} \qquad C.V_{k+x} = \frac{\sigma_{k+x}}{\overline{x}} = \frac{\sigma_x}{\overline{x}+20} \quad \mapsto \quad C.V_x > C.V_{k+x}$$

En este sentido,

$$E(20+x) = 20 + \overline{x} \qquad \sigma^2_{20+x} = \sigma^2_x = 3,2 \qquad \sigma_{20+x} = \sigma_x = 1,79$$

$$C.V_x = \frac{\sigma_x}{\overline{x}} = \frac{1,79}{10} = 0,179 \qquad C.V_{20+x} = \frac{\sigma_x}{\overline{x}+20} = \frac{1,79}{30} = 0,059$$

Es más disperso el conjunto de datos inicial.

5. Los pesos en kg de 20 alumnos de cierto centro son: 51, 47, 55, 53, 49, 47, 48, 50, 43, 60, 45, 54, 62, 57, 46, 49, 52, 42, 38, 61.

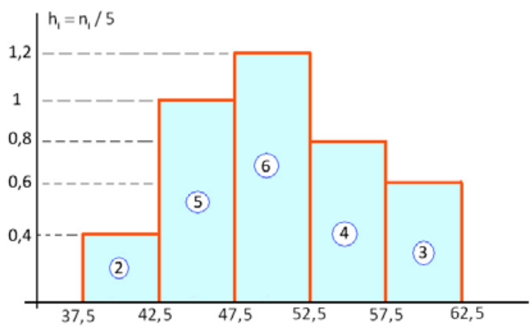
a) Agrupar los datos en clases de amplitud 5, siendo el extremo inferior del primer intervalo 37,5. Dibujar el correspondiente histograma y calcular la media de los datos agrupados.

b) Comparar la proporción de observaciones en el primer intervalo con la que cabría esperar bajo una distribución normal de media 50 y desviación típica 6,4.

Solución:

a)

Intervalo	x_i	n_i	$x_i \,.\, n_i$	$h_i = n_i / 5$
37,5 – 42,5	40	2	80	0,4
42,5 – 47,5	45	5	225	1
47,5 – 52,5	50	6	300	1,2
52,5 – 57,5	55	4	220	0,8
57,5 – 62,5	60	3	180	0,6
			$\sum_{i=1}^{5} x_i \,.\, n_i = 1005$	



$$\overline{x} = \frac{\sum_{i=1}^{5} x_i n_i}{N} = \frac{1005}{20} = 50,25$$

b) La proporción en el intervalo $[37,5-42,5)$ es $2/20=0,1$

En una distribución normal $N(50;\,6,4)$ la proporción sería:

$$P(37,5\le x\le 42,5)=P\left(\frac{37,5-50}{6,4}\le\frac{x-50}{6,4}\le\frac{42,5-50}{6,4}\right)=P\big(-1,95\le z\le -1,17\big)=$$

$$=P\big(1,17\le z\le 1,95\big)=P(z\ge 1,17)-P(z\ge 1,95)=0,1210-0,0256=0,0954$$

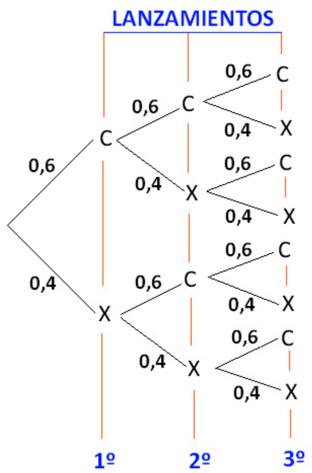
6. Se considera el experimento de lanzar una moneda tres veces. Se pide:

a) Construir el espacio muestral

b) Suponiendo que la moneda esté cargada y que la probabilidad de cara sea de 0,6. ¿cuáles son las probabilidades de los sucesos elementales?.

Solución:

a) El espacio muestral con el diagrama de árbol:



$$E=\big\{(c,\,c,\,c)\,,(c,\,c,\,x)\,,(c,\,x,\,c)\,,(c,\,x,\,x)\,,\;(x,\,c,\,c)\,,(x,\,c,\,x)\,,(x,\,x,\,c)\,,(x,\,x,\,x)\big\}$$

b) $P(c)=0,6 \qquad P(x)=0,4$

$P(c,\,c,\,c)=0,6^3 \qquad P(c,\,c,\,x)=0,6^2\,.\,0,4 \qquad P(c,\,x,\,c)=0,6^2\,.\,0,4 \qquad P(c,\,x,\,x)=0,6\,.\,0,4^2$

$P(x,\,c,\,c)=0,6^2\,.\,0,4 \qquad P(x,\,c,\,x)=0,6\,.\,0,4^2 \qquad P(x,\,x,\,c)=0,6\,.\,0,4^2 \qquad P(x,\,x,\,x)=0,4^3$

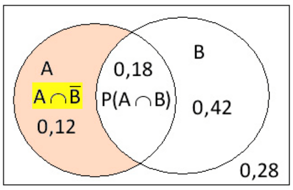
7. Sean A y B dos sucesos independientes. La probabilidad de que suceda el suceso B es de 0,6. Sabemos que $P(A/B)=0,3$

a) Calcula la probabilidad de que suceda al menos uno de los dos sucesos.

b) Calcula la probabilidad de que suceda el suceso A pero no el B.

Solución:

Por ser independientes $\Rightarrow \begin{cases} P(A/B)=P(A)=0,3 \\ P(A\cap B)=P(A)\,.\,P(B)=0,3\,.\,0,6=0,18 \end{cases}$



a) $P(A\cup B)=P(A)+P(B)-P(A\cap B)=0,3+0,6-0,18=0,72$

o también, $P(A\cup B)=0,12+0,18+0,42=0,72$

b) Si A y B son independientes, también lo son A y $\overline{B}$, con lo cual:

$$P(A\cap\overline{B})=P(A)\,.\,P(\overline{B})=0,3\,.\,0,4=0,12$$

8. En un juego se sortea cada día un premio utilizando papeletas con tres cifras, numeradas del 000 al 999.
a) Calcula la probabilidad de que el número premiado acabe en 5.
b) Calcula la probabilidad de que el número premiado acabe en 55.
c) Sabiendo que ayer salió un número acabado en 5, calcula la probabilidad de que el número de hoy acabe también en 5.

Solución:

a) En el juego hay 1.000 papeletas, entre ellas hay un total de $VR_{10,2}=10^2=100$ que acaban en 5.

$$P(\text{acabe en }5)=\frac{100}{1000}=0,1$$

b) Hay 10 papeletas que acaban en 55, con lo que:

$$P(\text{acabe en }55)=\frac{10}{1000}=0,01$$

c) Los números que salgan en el sorteo son independientes del día.

$$P(\text{acabe en }5)=\frac{100}{1000}=0,1$$

9. Una muestra aleatoria de 9 tarrinas de helado proporciona los siguientes pesos en gramos: 88 , 90 , 90 , 86 , 87 , 88 , 91 , 92 , 89
Halla un intervalo de confianza al 95% para la media poblacional, sabiendo que el peso de las tarrinas tiene una distribución normal con desviación típica de 1,8 gramos.

Solución:

Un intervalo de confianza $(1-\alpha)$ para un parámetro poblacional se define:

$$I_{1-\alpha}(\text{parámetro poblacional})=\left[\text{parámetro muestral}\ \pm\ \text{error muestral}\right]$$

En este caso, se trata de construir un intervalo de confianza para la media poblacional μ de varianza $\sigma^2 = 1,8^2$ conocida, es decir:

$$I_{1-\alpha} = \left[\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \qquad 1-\alpha = 0,95 \qquad z_{\alpha/2} = z_{0,025} = 1,96$$

$$\overline{x} = \frac{88+80+90+90+86+87+88+91+92+89}{9} = 89 \text{ gr}$$

$$I_{0,95} = \left[89 \pm 1,96 \frac{1,8}{\sqrt{9}} \right] = \left[87,824 \ ; \ 90,176 \right]$$

Error muestral: $\varepsilon = 1,96 \dfrac{1,8}{\sqrt{9}} = 1,176$

Longitud Intervalo: $L = 2 . \varepsilon = 2 \ . \ 1,176 = 2,352$

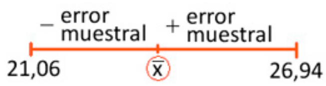
10. La longitud de la ballena azul se distribuye según una ley normal con desviación típica de 7,5 m. En un estudio estadístico realizado a 25 ejemplares se ha obtenido el intervalo de confianza (21,06 ; 26,94) para la longitud media.

a) Calcula la longitud media de los 25 ejemplares de la muestra.

b) Calcula el nivel de confianza con el que se ha construido dicho intervalo.

Solución:

a) $I_{1-\alpha}(\mu) = [\overline{x} \pm \text{ error muestral}]$



$$\overline{x} = \frac{21,06 + 26,94}{2} = 24 \text{ m}$$

b) El intervalo de confianza para la media poblacional μ de varianza σ^2 conocida, viene dado por la relación: $I_{1-\alpha}(\mu) = \left[\overline{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \right]$

En esta línea, con $n = 25$, $\overline{x} = 24$, $\sigma = 7,5$

$$I_{1-\alpha}(\mu) = \left[24 \pm z_{\alpha/2} \frac{7,5}{\sqrt{25}} \right] = \left[24 - z_{\alpha/2} \frac{7,5}{\sqrt{25}} \ ; \ 24 + z_{\alpha/2} \frac{7,5}{\sqrt{25}} \right] = (21,06 \ ; \ 26,94)$$

Longitud Intervalo: $L = 2 . \varepsilon = 26,94 - 21,06 = 5,88$

Error muestral: $\varepsilon = \dfrac{L}{2} = 2,94 = z_{\alpha/2} \dfrac{7,5}{\sqrt{25}} \quad \mapsto \quad z_{\alpha/2} = \dfrac{2,94 \ . \ 5}{7,5} = 1,96$

Observando la tabla de la normal N(0, 1) se tiene:

$z_{\alpha/2} = 1,96 \quad \mapsto \quad \alpha/2 = 0,0250 \quad \Rightarrow \quad \alpha = 0,05$

El nivel de confianza $1 - \alpha = 0,95$ (95%)

11. Un supervisor somete una muestra de 16 fusibles a cierta sobrecarga. Los tiempos que tardaron en fundirse dieron una media de 10,63 minutos. Considerando que la variable 'tiempo que tarda en fundirse un fusible sometido a esa sobrecarga es normal' con desviación típica de 2,48 minutos.

a) Construye un intervalo de confianza para la media poblacional con un nivel de confianza del 95%.

b) ¿Cuál debe ser el tamaño de la muestra para que el error de estimación de la media sea inferior a 1 minuto, con un nivel de confianza del 95%?

Solución:

a) El intervalo de confianza para la media poblacional μ de varianza σ^2 conocida, viene dado por la relación: $I_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}\right]$

Siendo: $n = 16$, $\bar{x} = 10,63$, $\sigma = 2,48$, $1 - \alpha = 0,95$, $z_{\alpha/2} = z_{0,025} = 1,96$

$$I_{1-\alpha}(\mu) = \left[10,63 \pm 1,96 \ \frac{2,48}{\sqrt{16}}\right] = [9,41 \ ; \ 11,84]$$

b) Error muestral: $\varepsilon = z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \quad \mapsto \quad 1 = 1,96 \ \dfrac{2,48}{\sqrt{n}}$

$n = (1,96 \ . \ 2,48)^2 = 23,62 \approx 24$

Las muestras deben de ser de tamaño 24 o mayor.

12. En una cadena de centros comerciales trabajan 150 personas en el departamento de personal, 450 en el de ventas, 200 en el de contabilidad y 100 en el de atención al cliente. Con objeto de realizar una encuesta laboral se quiere seleccionar una muestra de 180 trabajadores.

a) ¿Qué tipo de muestreo se debe emplear para que en la muestra estén representados todos los departamentos?

b) ¿Qué número de trabajadores de cada departamento habrá en la muestra?

Solución:

a) Se debe utilizar un muestreo aleatorio estratificado, ya que la población se encuentra dividida en cuatro estratos (cuatro departamentos), y además se quiere que haya elementos en la muestra de todos los estratos.

b) La población está formada por $150+450+200+100=900$ elementos.

Los integrantes de la muestra de cada departamento se repartirán de forma proporcional a los trabajadores de cada departamento.

$$\frac{180}{900}=0,2=\frac{n_1}{150}=\frac{n_2}{450}=\frac{n_3}{200}=\frac{n_4}{100} \Rightarrow \begin{cases} n_1=0,2\,.\,150=30 & \text{Personal} \\ n_2=0,2\,.\,450=90 & \text{Ventas} \\ n_3=0,2\,.\,200=40 & \text{Contabilidad} \\ n_4=0,2\,.\,100=20 & \text{Atención Cliente} \end{cases}$$

13. En una ciudad se seleccionó al azar una muestra de 225 familias. A cada familia seleccionada se le preguntó si tenía contratatado algún seguro de incendios. Se obtuvo como resultado que 75 familias tenían contratado dicho seguro. A partir de esa información, determina:

a) El intervalo de confianza al 95% para la proporción de familias de esa ciudad que tienen contratado algún seguro de incendios.

b) El error máximo que cometeríamos, con una confianza del 95%, si diéramos como estimación de dicha proporción el cociente $\frac{75}{225}$

Solución:

a) La proporción muestral $\hat{p}=\frac{75}{225}=\frac{1}{3}$, $\hat{q}=\frac{2}{3}$, donde $n=225$

$$1-\alpha=0,95 \quad \mapsto \quad z_{\alpha/2}=z_{0,025}=1,96$$

El intervalo de confianza para la proporción p de una población viene dado por la expresión: $I_{1-\alpha}(p)=\left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\,.\,\hat{q}}{n}}\right]$

$$I_{0,95}(p)=\left[\frac{1}{3} \pm 1,96\sqrt{\frac{\frac{1}{3}\,.\,\frac{2}{3}}{225}}\right]=\left(0,272\ ;\ 0,395\right)$$

b) Error muestral: $\varepsilon=z_{\alpha/2}\sqrt{\frac{\hat{p}\,.\,\hat{q}}{n}}$, o bien, $L=2\,.\,\varepsilon$

$$L=2\,.\,\varepsilon=0,395-0,272=0,123 \quad \mapsto \quad \varepsilon=0,0615$$

o bien, $\varepsilon=z_{\alpha/2}\sqrt{\frac{\hat{p}\,.\,\hat{q}}{n}}=1,96\sqrt{\frac{\frac{1}{3}\,.\,\frac{2}{3}}{225}}=0,0615$

El error máximo que se cometería es del 6,15%

14. El alcalde de una ciudad prometió en su programa oponerse a la construcción de una central de tratamiento de ciertos residuos, puesto que en aquel momento sólo un 10% de los ciudadanos estaba a favor de la central.
En los últimos días se ha encuestado a 100 personas de las cuales 14 están a favor de la central. El alcalde afirma, sin embargo, que el porcentaje de los ciudadanos a favor sigue siendo del 10% o incluso ha disminuido.

a) Plantea un test para contrastar la hipótesis defendida por el alcalde a que sucedió lo contrario, como parecen indicar los datos. Si se concluye que el porcentaje ha aumentado y la hipótesis del alcalde fuera falsa, ¿cómo se llama el error cometido?
b) Explica claramente a qué conclusión se llega en el test planteado en el apartado anterior para un nivel de significación del 5%.

Solución:

a) Se trata de un contraste unilateral para la proporción.

Hipótesis nula $H_0: p \leq 0,1$ *El alcalde tiene razón*
Hipótesis alternativa $H_a: p > 0,1$ *El porcentaje ha aumentado*

Si se admitiera que el porcentaje ha aumentado (rechazando la hipótesis nula), y esto **fuera falso**, se cometería un **error TIPO I**.

Si se admitiera la hipótesis nula, ***afirmando que el alcalde lleva razón***, el porcentaje no ha cambiado, y esto **fuera falso**, se cometería un **error TIPO II**.

Es mucho más grave cometer un error TIPO II.

b) Estadístico de contraste para aceptar H_0 : $\dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} \leq z_\alpha$

$\alpha = 0,05 \quad \mapsto \quad z_{0,05} = 1,645 \qquad n = 100 \qquad \hat{p} = 0,14 \qquad \hat{q} = 0,86$

$$\frac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} = \frac{0,14-0,1}{\sqrt{\dfrac{0,14\,.\,0,86}{100}}} = 1,152 < 1,645 = z_{0,05}$$

En consecuencia, se acepta la hipótesis nula, llevaba razón el alcalde, con una fiabilidad del 95%, al afirmar que la situación no había cambiado.

15. El número de reclamaciones presentadas durante la campaña de Navidad en 9 tiendas de una empresa ha sido: 25 , 31 , 28, 30 , 32 , 20 , 22 , 34 , 30

Se acepta que estos números de reclamaciones siguen una distribución normal con desviación típica igual a 5. Se desea contrastar si el número de reclamaciones es 26, con un nivel de significación de 0,05

a) Plantéense cuáles son las hipótesis nula y alternativa en el contraste.

b) Determínese la región crítica del contraste.

c) ¿Es posible aceptar la hipótesis con el nivel de significación indicado?

Solución:

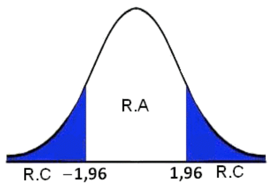
a) Se trata de un contraste bilateral para la media de una población normal con varianza conocida.

Hipótesis nula $H_0: \mu = 26$

Hipótesis alternativa $H_a: \mu \neq 26$

Estadístico de contraste para aceptar H_0: $z = \dfrac{|\bar{x} - \mu|}{\sigma / \sqrt{n}} \leq z_{\alpha/2}$

b)



$\alpha = 0,05 \quad \mapsto \quad z_{0,05/2} = z_{0,025} = 1,96$

c) $\sigma = 5$ reclamaciones $\qquad n = 9$ tiendas

$$\bar{x} = \frac{25+31+28+30+32+20+22+34+30}{9} = 28 \ \text{ reclamaciones}$$

Como $z = \dfrac{|\bar{x} - \mu|}{\sigma / \sqrt{n}} = \dfrac{|28 - 26|}{5 / \sqrt{9}} = 1,2 < 1,96 = z_{0,025}$

Por tanto, se admite la hipótesis nula, concluyendo que el número medio de reclamaciones presentadas es 26.

16. La media de una variable aleatoria X con distribución normal es 5 veces la desviación típica. Además se verifica: $P(X \leq 6)=0,8413$

Calcula la media y la desviación típica de la variable aleatoria X.

Solución:

La variable aleatoria X sigue una distribución normal $N(\mu = 5\sigma\,,\ \sigma)$

$$P(X \leq 6) = P\left(\frac{X-5\sigma}{\sigma} \leq \frac{6-5\sigma}{\sigma}\right) = P\left(z \leq \frac{6-5\sigma}{\sigma}\right) = 0,8413$$

$P\left(z \geq \frac{6-5\sigma}{\sigma}\right) = 01587$, en la tabla N(0, 1) se observa $P(z \geq 1) = 01587$

$$\frac{6-5\sigma}{\sigma} = 1 \quad \mapsto \quad \sigma = 1 \qquad \mu = 5\sigma = 5$$

17. Se sabe que dos poblaciones distintas, X e Y, se distribuyen normalmente con media 0. Además $P(X \geq 2) = P(Y \geq 3) = 0,1587$. Calcula sus respectivas varianzas.

Solución:

Sean las variables aleatorias $X \sim N(0, \sigma_1)$ e $Y \sim N(0, \sigma_2)$

$$P(X \geq 2) = P\left(\frac{X-0}{\sigma_1} \geq \frac{2-0}{\sigma_1}\right) = P\left(z \geq \frac{2}{\sigma_1}\right) = 0,1587$$

$$P(Y \geq 3) = P\left(\frac{Y-0}{\sigma_2} \geq \frac{3-0}{\sigma_2}\right) = P\left(z \geq \frac{3}{\sigma_2}\right) = 0,1587$$

En la tabla N(0, 1), se observa $P(z \geq 1) = 01587$, de donde:

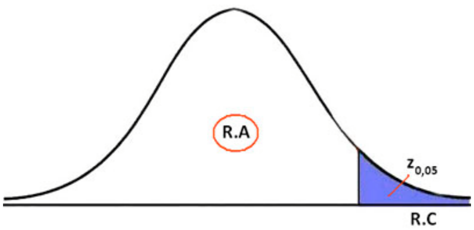
$$\frac{2}{\sigma_1} = 1 \quad \mapsto \quad \sigma_1 = 2 \quad \mapsto \quad \sigma_1^2 = 4 \qquad \frac{3}{\sigma_2} = 1 \quad \mapsto \quad \sigma_2 = 3 \quad \mapsto \quad \sigma_2^2 = 9$$

18. El consumo de carne de pollo parece haberse disparado desde que hace unos meses cundió la alarma sobre otros tipos de carne. En cierta carnicería, las ventas diarias de carne de pollo seguían hasta entonces una normal de media 19 kilos y desviación típica 3 kilos. En una muestra de 35 días posteriores a la citada alarma se obtuvo una media de 21 kilos de carne de pollo vendidos al día. Suponiendo que las ventas siguen una normal con la misma desviación típica:
a) Plantea un test para contrastar que la venta de pollo no ha aumentado, como parecen indicar los datos. ¿A qué conclusión se llega a un nivel de significación del 5%?
b) Calcula un intervalo de confianza del 95% para la venta diaria media de carne de pollo.

Solución:

a) Se trata de un contraste unilateral para la media poblacional con varianza conocida. Se establecen las hipótesis:

Hipótesis nula $H_0: \ \mu = 19$ kg
Hipótesis alternativa $H_a: \ \mu > 19$ kg



Se acepta H_0 si el estadístico de contraste

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \leq z_\alpha$$

Región Aceptación $\equiv (-\infty ; \ 1{,}645)$

$\bar{x} = 21$ kg $\quad n = 35$ días $\quad \sigma = 3$ kg $\quad \alpha = 0{,}05 \rightarrow z_{0,05} = 1{,}645$

Siendo $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{21 - 19}{3 / \sqrt{35}} = 3{,}94 > 1{,}645 = z_{0,05}$

Se rechaza la hipótesis nula, aceptando en consecuencia la hipótesis alternativa, afirmando que el consumo medio de carne de pollo ha aumentado con la alarma, con una fiabilidad del 95%.

Adviértase que $3{,}94 \notin (-\infty ; \ 1{,}645)$, es un valor que está fuera de la región de aceptación.

b) El Intervalo de confianza para la media poblacional de una distribución normal de varianza conocida: $I_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \right]$ donde $z_{\alpha/2} = z_{0,025} = 1{,}96$

con lo cual, $I_{0,95}(\mu) = \left[21 \pm 1{,}96 \dfrac{3}{\sqrt{35}} \right] = [20{,}01 ; \ 21{,}99]$

Se observa que $19 \notin [20{,}01 ; \ 21{,}99]$, se encuentra fuera del intervalo de confianza

19. Un estudiante hace dos pruebas en un mismo día. La probabilidad de que pase la primera prueba es de 0,6; la de que pase la segunda es de 0,8, y la de que pase ambas es de 0,5. Halla las siguientes probabilidades:

a) Que pase al menos una prueba.

b) Que no pase ninguna prueba.

c) ¿Son las pruebas sucesos independientes?

d) Que pase la segunda prueba en el caso de no haber superado la primera.

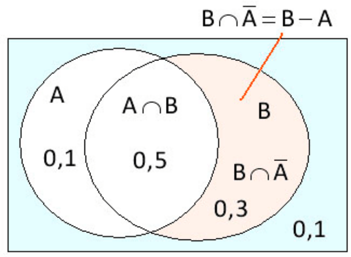
Solución:

Sean los sucesos:

$$\begin{cases} A = \text{"pasar la primera prueba"} & P(A) = 0,6 & P(\overline{A}) = 0,4 \\ B = \text{"pasar la segunda prueba"} & P(B) = 0,8 & P(\overline{B}) = 0,2 \\ A \cap B = \text{"pasar las pruebas A y B"} & P(A \cap B) = 0,5 & P(\overline{A \cap B}) = 0,5 \end{cases}$$

Leyes de Morgan: $\overline{A \cup B} = \overline{A} \cap \overline{B} \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$

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



a) $\begin{cases} P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,6 + 0,8 - 0,5 = 0,9 \\ P(A \cup B) = 0,1 + 0,5 + 0,3 = 0,9 \end{cases}$

b) $P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0,9 = 0,1$

c) A y B son independientes $\Leftrightarrow \quad P(A \cap B) = P(A) . P(B)$

$P(A \cap B) = 0,5 \ \boxed{\neq} \ 0,6 . 0,8 = 0,48 = P(A) . P(B)$

No son independientes

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

Siendo $B = (A \cap B) \cup (\overline{A} \cap B) \quad \begin{cases} P(B) = P(A \cap B) + P(\overline{A} \cap B) \\ 0,8 = 0,5 + P(\overline{A} \cap B) \quad \mapsto \quad P(\overline{A} \cap B) = 0,3 \end{cases}$

20. En una clase en la que todos practican algún deporte, el 60% de los alumnos juega al fútbol o al baloncesto, y el 10% practica ambos deportes. Además hay un 60% que no juega al fútbol. Halla la probabilidad de que, escogido al azar un alumno de la clase:

a) Juegue sólo al fútbol.

b) Juegue sólo al baloncesto.

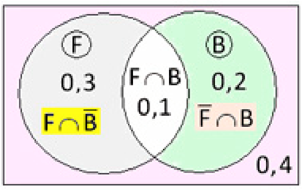
c) Practique uno solo de los deportes.

d) No juegue ni al fútbol ni al baloncesto.

Solución:

Sean los sucesos: $F = \text{"juega al fútbol"} \quad B = \text{"juega al baloncesto"}$

$P(F \cup B) = 0,6 \quad P(F \cap B) = 0,1 \quad P(\overline{F}) = 0,6 \quad \mapsto \quad P(F) = 0,4$



$P(F \cup B) = P(F) + P(B) - P(F \cap B) \quad \mapsto \quad \begin{cases} 0,6 = 0,4 + P(B) - 0,1 \\ P(B) = 0,3 \end{cases}$

a) $P(F \cap \overline{B}) = P(F) - P(F \cap B) = 0,4 - 0,1 = 0,3$

b) $P(\overline{F} \cap B) = P(B) - P(F \cap B) = 0,3 - 0,1 = 0,2$

c) $P(F \cap \overline{B}) + P(\overline{F} \cap B) = 0,3 + 0,2 = 0,5$

d) $P(\overline{F} \cap \overline{B}) = P(\overline{F \cup B}) = 1 - P(F \cup B) = 1 - 0,6 = 0,4$

21. Un dado con las caras numeradas del 1 al 6 está trucado de modo que la probabilidad de obtener un número es directamente proporcional a dicho número.

a) Halla la probabilidad de que salga 3 si se sabe que salió impar.

b) Calcula la probabilidad de que salga par si se sabe que salió mayor que 3.

Solución:

$P(1) = x \quad P(2) = 2x \quad P(3) = 3x \quad P(4) = 4x \quad P(5) = 5x \quad P(6) = 6x$

$P(\Omega) = 1 = x + 2x + 3x + 4x + 5x + 6x = 21x \quad \mapsto \quad x = \dfrac{1}{21}$

a) $P(3 / \text{salió impar}) = \dfrac{P(3)}{P(1,3,5)} = \dfrac{\dfrac{3}{21}}{\dfrac{1}{21} + \dfrac{3}{21} + \dfrac{5}{21}} = \dfrac{3}{9} = \dfrac{1}{3}$

b) $P(\text{par} / \text{mayor que } 3) = \dfrac{P(4,\ 6)}{P(4,\ 5,\ 6)} = \dfrac{\dfrac{4}{21} + \dfrac{6}{21}}{\dfrac{4}{21} + \dfrac{5}{21} + \dfrac{6}{21}} = \dfrac{10}{15} = \dfrac{2}{3}$

22. a) Halla la probabilidad de obtener al menos un seis doble en n tiradas de dos dados.

b) ¿Cuántas partidas habrá que jugar para que la probabilidad anterior sea de $\frac{1}{2}$?

Solución:

a) Denotando el suceso $A_i =$ "sacar 6 doble en la tirada i-ésima".

La probabilidad pedida es:

$$P(A) = P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) = 1 - P(\overline{A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n}) =$$

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

$$= 1 - P(\overline{A}_1) .\ P(\overline{A}_2) .\ P(\overline{A}_3) \cdots P(\overline{A}_n) = 1 - \left[P(\overline{A}_i)\right]^n = 1 - \left(\frac{35}{36}\right)^n$$

Adviértase que todos los sucesos son independientes y tienen la misma probabilidad.

$$P(\text{al menos un 6 doble en n tiradas de dos dados}) = 1 - \left(\frac{35}{36}\right)^n$$

b) $1 - \left(\frac{35}{36}\right)^n = \frac{1}{2} \quad \mapsto \quad \left(\frac{35}{36}\right)^n = \frac{1}{2} \quad \mapsto \quad Ln\left(\frac{35}{36}\right)^n = Ln\frac{1}{2} \quad \mapsto \quad Ln\left(\frac{35}{36}\right)^n = Ln\frac{1}{2}$

$$n\,(Ln35 - Ln36) = -Ln2 \quad \mapsto \quad n = \frac{Ln2}{Ln36 - Ln35} \simeq 25 \text{ partidas}$$

23. Si la probabilidad de que ocurra un suceso A es 1/5, ¿cuál es el mínimo número de veces que hay que repetir el experimento para que la probabilidad de que ocurra al menos una vez el suceso A sea mayor que 1/2?. ¿Cuál es la probabilidad de que ocurra al menos dos veces A al realizar 5 veces el experimento?.

Solución:

a) X="número de éxitos en n pruebas", $X \sim B(n,\ 0{,}2)$, $p = 0{,}2$, $q = 0{,}8$

$$P(X \geq 1) > \frac{1}{2} \quad \mapsto \quad P(X < 1) \leq \frac{1}{2} \quad \Rightarrow \quad P(X = 0) \leq \frac{1}{2}$$

$$P(X = 0) = \binom{n}{0} .\ 0{,}2^0 .\ 0{,}8^n = 0{,}8^n \leq \frac{1}{2} \xrightarrow{0{,}8^4 = 0{,}4096} n = 4$$

Para que el suceso A tenga al menos una probabilidad mayor $\frac{1}{2}$, hay que repetir el proceso un mínimo de 4 veces.

b) Se trata de una distribución binomial B(5; 0,2)

$$P(X \geq 2) = 1 - [P(X=0) + P(X=1)] = 1 - \left[\binom{5}{0} . 0,2^0 . 0,8^5 + \binom{5}{1} . 0,2^1 . 0,8^4\right] =$$

$$= 1 - (0,3277 + 0,4096) = 0,2627$$

24. Sean A y B dos sucesos independientes de un experimento aleatorio tales que $P(A) = 0,5$ y $P(\overline{B}) = 0,8$. Calcúlese:

a) $P(A \cap B)$ y $P(A \cup B)$

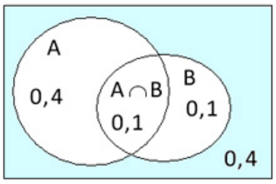
b) $P(\overline{A} / \overline{B})$

Solución:

a) $P(A) = 0,5 \;\rightarrow\; P(\overline{A}) = 0,5 \quad P(\overline{B}) = 0,8 \;\rightarrow\; P(B) = 0,2$

A y B independientes $\Rightarrow \begin{cases} P(A \cap B) = P(A) . P(B) \\ P(\overline{A} \cap \overline{B}) = P(\overline{A}) . P(\overline{B}) \end{cases}$ *$\overline{A}$ y $\overline{B}$ independientes*

$$P(A \cup B) = \begin{cases} P(A) + P(B) - P(A \cap B) = 0,5 + 0,2 - 0,1 = 0,6 \\ 0,4 + 0,1 + 0,1 = 0,6 \end{cases}$$



$P(A \cap B) = 0,5 . 0,2 = 0,1$

b) $$P(\overline{A} / \overline{B}) = \frac{P(\overline{A} \cap \overline{B})}{P(\overline{B})} = \frac{1 - P(A \cup B)}{P(\overline{B})} = \frac{0,4}{0,8} = 0,5$$

$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$

o también, $P(\overline{A} \cap \overline{B}) = P(\overline{A}) . P(\overline{B}) = 0,5 . 0,8 = 0,4$

25. La distancia diaria recorrida, en kilómetros, por un taxi en una gran ciudad puede aproximarse por una variable aleatoria con distribución normal de media μ desconocida y desviación típica $\sigma = 16$ km.

a) Se toma una muestra aleatoria simple de 81 taxis y se obtiene el intervalo (159 ; 165). Determínese el nivel de confianza con el que se obtuvo dicho intervalo.

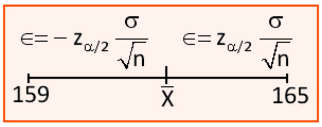
b) Si la media de la distancia recorrida fuera $\mu = 160$ km, y se toma una muestra aleatoria simple de 64 taxis, calcúlese la probabilidad de que la media de la muestra, $\overline{X}$, sea mayor que 156 km.

Solución:

a) Se trata de un inervalo de confianza para la media de una población μ de varianza conocida:

$$I_{1-\alpha}(\mu)=\left[\bar{x}\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \quad \mapsto \quad P\left[\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\leq\mu\leq\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]=1-\alpha$$

El Intervalo de confianza está centrado en la media $\bar{X}$:



$$\bar{X}=\frac{159+165}{2}=162$$

Longitud del intervalo: $L=2\in=2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}=165-159=6$

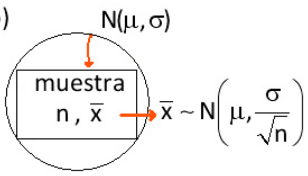
Error estimación: $\in=z_{\alpha/2}\frac{\sigma}{\sqrt{n}}=3$

siendo $\sigma=16$ km y $n=81$ taxis, se tiene: $\in=z_{\alpha/2}\frac{16}{\sqrt{81}}=3 \quad \rightarrow \quad z_{\alpha/2}=\frac{27}{16}=1,69$

Observando en la tabla N(0, 1): $z_{\alpha/2}=1,69 \quad \mapsto \quad \alpha/2=0,0455$

$\alpha=0,091 \quad \Rightarrow \quad 1-\alpha=0,91$ (91% nivel de confianza)

b)



La media muestral $\bar{x}\sim N\left(160,\frac{16}{\sqrt{64}}\right)\approx N(160,2)$

$n=64$ taxis $\quad \mu=160$ km

tipificando, se obtiene:

$$P(\bar{X}>156)=P\left(\frac{\bar{X}-160}{2}>\frac{156-160}{2}\right)=P(z>-2)=P(z<2)=0,97725$$

26. En un estudio de mercado se necesita comparar las ventas en meses de diciembre de dos años consecutivos para decidir si existen diferencias significativas entre ambos. Se dispone de datos del primer año en 52 tiendas que proporciona una media muestral de 942,30 euros y una desviación típica de 49 euros. En el segundo año se han observado las ventas en 64 tiendas obteniendo una venta media de 981,20 euros y una desviación típica de 74 euros.

a) Plantea un contraste de hipótesis para decidir si existen diferencias entre las ventas en ambos meses de diciembre.

b) Contrasta la hipótesis con un nivel de significación de 0,05.

c) ¿Cuál es el error del tipo I en este contraste?. ¿Con qué probabilidad se comete?

d) ¿En qué consiste el error de tipo II?

Solución:

a) Se trata de un contraste bilateral de igualdad de medias de dos poblaciones normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ de varianzas poblacionales desconocidas, con muestras grandes $n_1 + n_2 > 30$

Las variables aleatorias $X =$ "ventas primer año" e $Y =$ "ventas segundo año", donde $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$

Se establecen las hipótesis:

Hipótesis nula: $H_0:\ \mu_1 - \mu_2 = 0$ Hipótesis alternativa: $H_a:\ \mu_1 - \mu_2 \neq 0$

b) Si el estadístico de contraste $z = \dfrac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} \leq z_{\alpha/2}$ se acepta la hipótesis nula con un nivel de confianza $(1-\alpha)$

donde s_x^2 y s_y^2 son, respectivamente, las cuasivarianzas muestrales de las dos poblaciones X e Y. En general, se tiene la relación: $n_x \,.\, \sigma_x^2 = (n_x - 1)\,.\, s_x^2$

Se tiene $\begin{cases} \bar{x} = 942,3 & n_x = 52 & \sigma_x = 49 \\ \bar{y} = 981,2 & n_y = 64 & \sigma_y = 74 \end{cases}$

$$s_x^2 = \frac{n_x \,.\, \sigma_x^2}{(n_x - 1)} = \frac{52\,.\,49^2}{51} = 2448,08 \qquad s_y^2 = \frac{n_y \,.\, \sigma_y^2}{(n_y - 1)} = \frac{64\,.\,74^2}{63} = 5562,92$$

Con un error de significación $\alpha = 0,05 \ \mapsto\ z_{\alpha/2} = z_{0,025} = 1,96$
La Región de Aceptación será $(-1,96\ ;\ 1,96)$

En consecuencia,

$$z = \frac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} = \frac{|942,3 - 981,2|}{\sqrt{\dfrac{2448,08}{52} + \dfrac{5562,92}{64}}} = 3,36 > 1,96 = z_{\alpha/2}$$

Por lo que se rechaza la hipótesis nula, concluyendo que hay diferencias significativas entre las ventas del producto en diciembre, con una fiabilidad del 95%

Adviértase que la diferencia de medias muestrales $(\bar{x} - \bar{y})$ bajo la hipótesis nula sigue una distribución muestral

$$(\bar{x} - \bar{y}) \sim N\left[(\mu_x - \mu_y);\ \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right] \equiv N\left[0;\ \sqrt{\frac{2448,08}{52} + \frac{5562,92}{64}}\right] = N(0;\ 11,58)$$

En el caso de haber construido un intervalo de confianza:

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}-\bar{y})\pm z_{\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}\right]=\left[(942,3-981,2)\pm 1,96\sqrt{\frac{2448,08}{52}+\frac{5562,92}{64}}\right]$$

$$I_{0,95}(\mu_1-\mu_2)=\left[-61,59\,;-16,21\right]$$

Como el intervalo no cubre el 0 hay diferencias significativas entre las medias poblacionales con un nivel de confianza del 95%.

c) $\alpha=P\left[ET\,I\right]=P\left[\text{Rechazar } H_0\middle|H_0 \text{ cierta}\right]=P\left[\bar{x}-\bar{y}\neq 0\,/\,H_0:\mu_1-\mu_2=0\right]=0,05$

El error Tipo I consiste en afirmar que las ventas son distintas en el mes de diciembre de ambos años cuando en realidad son iguales. La probabilidad de cometer un error tipo I es de 0,05

d) $\beta=P\left[ET\,II\right]=P\left[\text{Aceptar } H_0\middle|H_0 \text{ falsa}\right]=P\left[\bar{x}-\bar{y}=0\,/\,H_a:\mu_1-\mu_2\neq 0\right]$

El error Tipo II consiste en afirmar que las ventas son iguales en el mes de diciembre de ambos años cuando en realidad son distintas.

27. Se ha realizado una encuesta sobre una población de escasa cultura, de la que solo un 15% ha leído más de tres libros. Elegida al azar una muestra de 50 personas, calcula:

a) La probabilidad de que haya más de cinco personas que han leído más de tres libros.

b) La probabilidad de que como máximo haya seis personas que han leído más de tres libros.

Solución:

Sea la variable aleatoria X="número de personas que han leído mas de tres libros", donde $X\sim B(50\,;\,0,15)$, donde:

$n=50 \quad p=0,15 \quad q=0,85 \quad \mu=n.p=50.0,15=7,5>5 \quad n.q=50.0,85=42,5>5$

Como $\mu=n.\,p>5$ y $n.\,q>5$, la distribución binomial $B(50\,;\,0,15)$ se puede aproximar a una distribución normal $N(\mu=n.\,p\,;\,\sigma=\sqrt{n.\,p.\,q})$, con lo cual:

$X\sim N(50.\,0,15\,;\,\sigma=\sqrt{50.\,0,15.\,0,85})$, es decir, $X\sim N(7,5\,;\,\sigma=2,525)$

Cuando la variable aleatoria discreta X toma valores $\{0,\ 1,\ 2,\ 3,\ ...\}$ al considerar la variable como continua, en un intervalo se establece el factor de corrección:

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

a) $P(X>5)=P(X'>5,5)=P\left(\frac{X'-7,5}{2,525}>\frac{5,5-7,5}{2,525}\right)=P(z>-0,79)=P(z<0,79)=$

$$=1-P(z\geq 0,79)=1-0,2148=0,7852$$

b) $P(X \le 6) = P(X' \le 6,5) = P\left(\frac{X'-7,5}{2,525} \le \frac{6,5-7,5}{2,525}\right) = P(z \le -0,39) = P(z \ge 0,39) = 0,3483$

28. Una compañía de autobuses realiza un estudio sobre el número de veces que semanalmente utilizan el autobús los usuarios. Se sabe que los datos se distribuyen en una normal N(10, 3). Calcula la probabilidad de que un usuario utilce el autobús:

a) Más de 11 veces.

b) Menos de 8 veces.

Solución:

La variable aleatoria X="número de veces que un individuo toma el autobús en una semana" es discreta, teniendo que considerar cada valor de la variable discreta como marca de clase, estableciendo el factor de corrección:

$$P(X = a) = P(a \le X \le a) = P(a - 0,5 \le X' \le a + 0,5)$$

a) $P(X > 11) = P(X' > 11,5) = P\left(\frac{X'-10}{3} > \frac{11,5-10}{3}\right) = P(z > 0,5) = 0,3085$

b) $P(X \le 8) = P(X' \le 7,5) = P\left(\frac{X'-10}{3} \le \frac{7,5-10}{3}\right) = P(z \le -0,83) = P(z \ge 0,83) = 0,2033$

29. Supón que en cierta población pediátrica, la presión sistólica de la sangre en reposo se distribuye normalmente con media de 115 mm Hg y desviación típica de 15.

a) Halla la probabilidad de que un niño elegido al azar en esta población tenga presión sistólica superior a 145 mm Hg.

b) ¿Por debajo de qué valor de presión sistólica estará el 75% de los niños?

Solución:

La variable aleatoria X = "presión sistólica medida en mm Hg", $X \sim N(115; 15)$

a) $P(X > 145) = P\left(\frac{X-115}{15} > \frac{145-115}{15}\right) = P(z > 2) = 0,0228$

b) $P(X \le k) = 0,75 \quad \mapsto \quad P\left(\frac{X-115}{15} \le \frac{k-115}{15}\right) = P\left(z \le \frac{k-115}{15}\right) = 0,75$

$P\left(z \le \frac{k-115}{15}\right) = 0,75 \quad \to \quad P\left(z \ge \frac{k-115}{15}\right) = 0,25$

Observando en la tabla de la N(0, 1) se tiene que $P(z \ge 0,67) = 0,2546$ y $P(z \ge 0,68) = 0,2483$. Puede tomarse el valor 0,67 por estar más próximo, o bien se puede interpolar.

Eligiendo 0,67, queda: $\dfrac{k-115}{15}=0,67 \quad \mapsto \quad k=115+15.\,0,67\simeq 125$ mm Hg

30. Se ha comprobado que el tiempo de espera (en minutos) hasta ser atendido, en cierto servicio de urgencias, sigue un modelo normal de probabilidad. A partir de una muestra de 100 personas que fueron atendidas en dicho servicio, se ha calculado un tiempo medio de espera de 14,25 minutos y una desviación típica de 2,5 minutos.
a) ¿Podríamos afirmar, con un nivel de significación del 5%, que el tiempo medio de espera, en ese servicio de urgencias, no es de 15 minutos?
b) ¿Qué podríamos concluir si el nivel de significación hubiese sido del 0,1%?
c) ¿Existe contradicción en ambas situaciones?

Solución:

a) Se trata de un contraste bilateral de la media de una población normal con varianza poblacional desconocida, con muestras grandes $n=100>30$.
Se establecen las hipótesis:

Hipótesis nula: $H_0:\ \mu=15$ minutos   Hipótesis alternativa: $H_a:\ \mu\neq 15$ minutos

Con el estadístico de contraste $z=\dfrac{|\bar{x}-\mu|}{s_x/\sqrt{n}}\leq z_{\alpha/2}$ se acepta H_0

$\alpha=0,05 \quad \mapsto \quad z_{\alpha/2}=z_{0,025}=1,96$

Región Aceptación $R.A=\{z:-z_{\alpha/2}\leq z\leq z_{\alpha/2}\} \quad \equiv \quad R.A\sim(-1,96\,;\,1,96)$

$$\bar{x}=14,25 \quad n=100 \quad \sigma_x=2,5 \quad (n-1)s_x^2=n\sigma_x^2 \ \Rightarrow \begin{cases} s_x^2=\dfrac{n\sigma_x^2}{n-1}=\dfrac{100.\,2,5^2}{99}=6,31 \\ s_x=\sqrt{6,31}=2,51 \end{cases}$$

La cuasivarianza muestral $s_x^2=\dfrac{\sum_{i=1}^{k}(x_i-\bar{x})^2.n_i}{n}$

Estadístico de contraste $z=\dfrac{|\bar{x}-\mu|}{s_x/\sqrt{n}}=\dfrac{|14,25-15|}{2,51/\sqrt{100}}=2,99>1,96=z_{0,025}$

Con lo que se rechaza la hipótesis nula , no pudiendo afirmar, con una fiabilidad del 95%, que el tiempo medio de espera en el servicio de urgencias sea de 15 minutos.

b) $\alpha=0,001 \quad \mapsto \quad z_{\alpha/2}=z_{0,005}=3,27 \quad \mapsto \quad R.A\sim(-3,27\,;\,3,27)$

Siendo $z = 2{,}99 < 3{,}27 = z_{0{,}0005}$ se aceptaría la hipótesis nula, pudiendo afirmar con una fiabilidad del 99,9%, que el tiempo medio de espera en el servicio de urgencias es de 15 minutos.

c) No existe ninguna contradicción. A mayor nivel de confianza $(1-\alpha)$, la longitud del intervalo de confianza es mayor porque el error muestral $\in = z_{\alpha/2}\dfrac{s_x}{\sqrt{n}}$ es más grande. Adviértase que la longitud del intervalo: $L = 2\in$

31. Una encuesta, realizada a 64 empleados de una fábrica, concluyó que el tiempo medio de duración de un empleo en la misma era de 6,5 años, con una desviación típica de 4.
¿Sirve esta información para aceptar, con un nivel de significación del 5%, que el tiempo medio de empleado en esa fábrica es menor o igual que 6?. Justifica adecuadamente la respuesta.

Solución:

Se trata de un contraste unilateral de la media de una población normal con varianza poblacional desconocida, con muestras grandes $n = 64 > 30$.
Se establecen las hipótesis:

Hipótesis nula: $H_0 :\ \mu \le 6$ años   Hipótesis alternativa: $H_a :\ \mu > 6$ años

Con el estadístico de contraste $z = \dfrac{\overline{x}-\mu}{s_x/\sqrt{n}} \le z_\alpha$ se acepta H_0 ; en caso contrario se rechaza, aceptándose la hipótesis alternativa H_a

$\alpha = 0{,}05 \quad \mapsto \quad z_\alpha = z_{0{,}05} = 1{,}645$

Región Aceptación $R.A = \{z : z \le z_\alpha\} \quad \equiv \quad R.A \sim (-\infty\, ;\, 1{,}96]$

$$\overline{x} = 6{,}5 \quad n = 64 \quad \sigma_x = 4 \quad (n-1)s_x^2 = n\sigma_x^2 \ \Rightarrow \begin{cases} s_x^2 = \dfrac{n\sigma_x^2}{n-1} = \dfrac{64\,.\,4^2}{63} = 16{,}25 \\ s_x = \sqrt{16{,}25} = 4{,}03 \end{cases}$$

La cuasivarianza muestral $s_x^2 = \dfrac{\sum_{i=1}^{k}(x_i-\overline{x})^2.n_i}{n}$

Estadístico de contraste $z = \dfrac{\overline{x}-\mu}{s_x/\sqrt{n}} = \dfrac{6{,}5-6}{4{,}03/\sqrt{64}} = 0{,}99 < 1{,}645 = z_{0{,}05}$

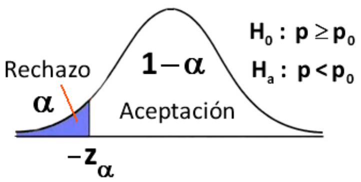
Con lo que se acepta la hipótesis nula , pudiendo afirmar con una fiabilidad del 95%, que el tiempo medio de empleo en la fábrica es menor o igual que 6 años.

32. Se afirma que, en una determinada ciudad, al menos el 30% de las familias poseen ordenador. Se toma una muestra aleatoria de 200 familias de la ciudad y resulta que 50 poseen ordenador.
A un nivel de significación de 0,05, ¿hay suficiente evidencia para refutar la afirmación?

Solución:

Se trata de un contraste unilateral para el parámetro p de una distribución binomial. Se establecen las hipótesis:

Hipótesis nula: $H_0:\ p \geq 0,3$ Hipótesis alternativa: $H_a:\ p < 0,3$



$\alpha = 0,05 \ \mapsto \ z_\alpha = z_{0,05} = 1,645$

Región Aceptación $R.A \equiv (-1,645\,;\ \infty)$

Si el estadístico de contraste $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} > -z_\alpha$ se acepta H_0

$\hat{p} = \dfrac{50}{200} = 0,25 \quad \hat{q} = 0,75 \quad n = 200$

Estadístico de contraste $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} = \dfrac{0,25 - 0,3}{\sqrt{\dfrac{0,25\,.\,0,75}{200}}} = -1,63 > -1,645 = z_{0,05}$

Con lo que se acepta la hipótesis nula , pudiendo afirmar que al menos el 30% de las familias poseen ordenador, con una fiabilidad del 95%.

33. Se trabaja con la hipótesis se que uno de cada diez varones manifiesta algún tipo de daltonismo.

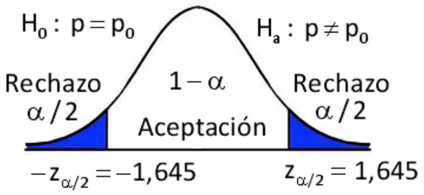
a) Elegidos 400 varones, se detectan 50 daltónicos. Con un nivel de significación del 10%, ¿se puede aceptar la hipótesis de partida?

b) Sobre la muestra estudiada en (a), ¿se obtendría la misma conclusión si $\alpha = 0{,}02$?

Solución:

a) Se trata de un contraste bilateral para el parámetro p de una distribución binomial. Se establecen las hipótesis:

Hipótesis nula: $H_0: \ p = 0{,}1$ Hipótesis alternativa: $H_a: \ p \neq 0{,}1$



$\alpha = 0{,}10 \quad \mapsto \quad z_{\alpha/2} = z_{0,05} = 1{,}645$

Región Aceptación $R.A \equiv (-1{,}645\,; \ 1{,}645)$

Si el estadístico de contraste $z = \dfrac{|\hat{p} - p|}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} \leq z_{\alpha/2}$ se acepta H_0

$\hat{p} = \dfrac{50}{400} = 0{,}125 \quad \hat{q} = 0{,}875 \quad n = 400$

Estadístico de contraste $z = \dfrac{|\hat{p} - p|}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} = \dfrac{|0{,}125 - 0{,}1|}{\sqrt{\dfrac{0{,}125\,.\,0{,}875}{400}}} = 1{,}512 \leq 1{,}645 = z_{0,05}$

Con lo que se acepta la hipótesis nula , pudiendo afirmar que cada uno de diez varones manifiesta algún tipo de daltonismo, con un nivel de confianza del 90%.

b) Si $\alpha = 0{,}02 \quad \mapsto \quad z_{\alpha/2} = z_{0,01} = 2{,}33$.

Siendo la región de aceptación $(-2{,}33\,; \ 2{,}33)$

Como $z = 1{,}512 \leq 2{,}33 = z_{0,01}$ se acepta la hipótesis nula , pudiendo afirmar que cada uno de diez varones manifiesta algún tipo de daltonismo, con una fiabilidad del 98%.

Adviértase que a mayor nivel de confianza, la región de aceptación es mayor.

34. El 42% de los escolares de un cierto país suelen perder un día de clase a causa de gripes y catarros. Sin embargo, un estudio sobre 1000 escolares revela que en el último curso hubo 450 en tales circunstancias.
Las autoridades sanitarias defienden que el porcentaje del 42% para toda la población de escolares se ha mantenido.

a) Contrasta, con un nivel de significación del 5%, la hipótesis defendida por las autoridades sanitarias, frente a que el porcentaje ha aumentado como parecen indicar los datos, explicando claramente a qué conclusión se llega.

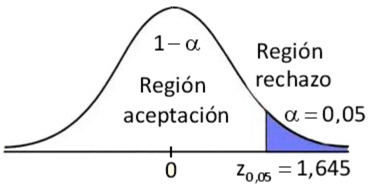
b) ¿Cómo se llama la probabilidad de concluir erróneamente que el tanto por ciento se ha mantenido?

Solución:

a) Se trata de un contraste unilateral para la proporción de una distribución binomial. Se establecen las hipótesis:

Hipótesis nula $H_0:\ p \leq 0{,}42$ Hipótesis alternativa $H_a:\ p > 0{,}42$

Estadístico de contraste para aceptar H_0: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} \leq z_\alpha$



$\alpha = 0{,}05 \quad \mapsto \quad z_\alpha = z_{0,05} = 1{,}645$

La región de aceptación se determina con el nivel de significación, así Región Aceptación: $(-\infty\,,\ 1{,}645)$

En la muestra: n=1000 $\hat{p} = \dfrac{450}{1000} = 0{,}45 \quad \hat{q} = 1 - 0{,}45 = 0{,}55$

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}}} = \frac{0{,}45 - 0{,}42}{\sqrt{\dfrac{0{,}45\,.\,0{,}55}{1000}}} = 1{,}91 > 1{,}645 = z_{0,05}$$

Con lo que se rechaza la hipótesis nula, aceptando con un nivel de confianza del 95% que la proporción ha aumentado

b) La probabilidad de concluir erróneamente que el tanto por ciento se ha mantenido, es aceptar la hipótesis nula H_0 cuando es falsa. Es un error Tipo II.