
1. Si $\bar{X} \sim N(40,10)$, calcular $\Pr(39 \leq \bar{X} \leq 41)$ para $n=10$. ¿En qué intervalo se obtendrán el 95% de los resultados?

SOLUCIÓN:

$$\Pr(39 \leq \bar{X} \leq 41) = \Pr\left(\frac{39-40}{\sqrt{10}} \leq \frac{\bar{X}-40}{\sqrt{10}} \leq \frac{41-40}{\sqrt{10}}\right) = \Pr(-0.31623 \leq \bar{X} \leq 0.31623)$$

$$Z = \frac{\bar{X}-40}{\sqrt{10}} \rightarrow N(0,1); \Pr(39 \leq \bar{X} \leq 41) = \Pr(Z \leq 0.31623) - \Pr(Z \leq -0.31623) =$$

$$= 2 \Pr(Z \leq 0.31623)$$

$$\text{Y por tanto, } \Pr(39 \leq \bar{X} \leq 41) = 2 * 0.6241 - 1 = .02482$$

$$\Pr(\mu - \varepsilon \leq \bar{X} \leq \mu + \varepsilon) = 0.95$$

$$\Pr(\mu - \varepsilon \leq \bar{X} \leq \mu + \varepsilon) = 2 * \Pr(Z \leq \frac{\varepsilon}{\sqrt{10}}) - 1$$

$$\Pr(Z \leq \frac{\varepsilon}{\sqrt{10}}) = \frac{1+0.95}{2} = 0.975 \rightarrow Z_{0.975} \rightarrow \varepsilon = 1.96\sqrt{10} = 6.1981$$

Por tanto, el intervalo es: (33.802, 46.198)

2. Si el contenido en gr. de un determinado medicamento X sigue una distribución $N(7.5, 0.3)$, calcular la probabilidad de que para una muestra de tamaño $n=5$, se obtenga medio menor que 7, $\Pr(\bar{X} \leq 7)$.

SOLUCIÓN:

A partir de una muestra de tamaño $n=5$ de una población normal $N(\mu=7.5, \sigma=0.3)$, tenemos que:

$$\Pr(\bar{X} \leq 7) = \Pr\left(\frac{\bar{X} - 7.5}{\frac{0.3}{\sqrt{5}}} \leq \frac{7 - 7.5}{\frac{0.3}{\sqrt{5}}}\right) = \Pr(Z \leq -3.7269)$$

Donde Z tiene una distribución normal estándar, y por tanto, $\Pr(\bar{X} \leq 7) = 0.0001$

3. Si la altura de un grupo de población sigue una distribución normal N(176,12), calcular la $\Pr(S \leq 10)$ para una muestra de tamaño 8.

SOLUCIÓN:

Considerando una muestra aleatoria de tamaño n de una población normal N(μ, σ), por el teorema de Fisher tenemos que:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

En particular, para una muestra de tamaño n=8 de una población normal N(176,12), el estadístico $\frac{7}{144}S^2$ sigue una distribución χ_7^2 , y por tanto

$$\Pr(S \leq 10) = \Pr(S^2 \leq 100) = \Pr\left(\frac{7}{144}S^2 \leq \frac{700}{144}\right) = \Pr(T \leq 4.8611)$$

Donde la variable T sigue una distribución χ_7^2 , es decir,

$$\Pr(S \leq 10) = 0.3232$$

4. Un ascensor limita el peso de sus cuatro ocupantes a 300Kg. Si el peso de un individuo sigue una distribución N(71,7), calcular la probabilidad de que el peso de 4 individuos supere los 300Kg.

SOLUCIÓN:

Teniendo en cuenta que el peso de cada individuo tiene una distribución normal $N(\mu = 71, \sigma = 7)$, si seleccionamos una muestra aleatoria de 4 individuos, tenemos que:

$$\begin{aligned} \Pr\left(\sum_{i=1}^4 X_i > 300\right) &= \Pr\left(\frac{\sum_{i=1}^4 X_i}{4} > \frac{300}{4}\right) = \Pr(\bar{X} > 75) = \Pr\left(\frac{\bar{X} - 71}{\frac{7}{\sqrt{4}}} > \frac{75 - 71}{\frac{7}{\sqrt{4}}}\right) = \\ &= \Pr(Z > 1.1429) = 1 - \Pr(Z \leq 1.1429) \end{aligned}$$

donde Z tiene una distribución normal estándar, y por tanto,

$$\Pr\left(\sum_{i=1}^4 X_i > 300\right) = 1 - 0.8735 = 0.1265$$

5. Calcular la probabilidad de que la media μ se encuentre entre $\bar{X} \pm 3S$ para poblaciones normales y $n = 5$.

SOLUCIÓN:

A partir del teorema de Fisher, en el muestreo sobre poblaciones normales, tenemos que los estadísticos \bar{X} y S^2 son independientes, siendo la distribución del estadístico $T = \sqrt{n} * \frac{\bar{X} - \mu}{S}$ una t_{n-1} (t de Student de $n - 1$ grados de libertad). En particular, si consideramos una muestra aleatoria de tamaño $n = 5$, la probabilidad de que la media esté entre $\bar{X} \pm 3S$ viene dada por:

$$\Pr(\bar{X} - 3S < \mu < \bar{X} + 3S) = \Pr\left(-3 < \frac{\mu - \bar{X}}{S} < 3\right) = \Pr\left(-3\sqrt{5} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{5}}} < 3\sqrt{5}\right) = \Pr(-3\sqrt{5} < T < 3\sqrt{5})$$

donde T tiene una distribución t_4 , y por tanto:

$$\Pr(\bar{X} - 3S < \mu < \bar{X} + 3S) = \Pr(-3\sqrt{5} < T < 3\sqrt{5}) = 2 \Pr(T < 6.7082) - 1 = 2 * 0.9987 - 1 = 0.9974$$

6. Calcular un intervalo de confianza al nivel $\alpha = 0.05$ para la probabilidad de p de que un recién nacido sea niño si en una muestra de tamaño 123 se han obtenido 67 niños.

SOLUCIÓN:

Teniendo en cuenta que la proporción de varones recién nacidos puede modelizarse por una variable Bernoulli de parámetro p (probabilidad de que un recién nacido sea varón), el intervalo de confianza al nivel $\alpha = 0.05$ viene dado por:

$$\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Donde $n = 123$. $\hat{p} = \frac{67}{123}$ y $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$, es decir,

$$(0.544715 - 0.0880096, 0.544715 + 0.0880096)$$

y por tanto, el intervalo $(0.0456706, 0.632725)$ contendrá a la proporción de varones nacidos con una probabilidad del 95%.

7. Calcular un intervalo de confianza al nivel $\alpha = 0.001$ para el peso exacto mediante los resultados obtenidos con 10 básculas:

7.20, 7.01, 7.36, 6.91, 7.22, 7.03, 7.11, 7.12, 7.03, 7.05

SOLUCIÓN:

Suponiendo que las medidas del peso de las básculas sigue una distribución normal $N(\mu, \sigma^2)$ con media el peso exacto, estamos interesados en encontrar un intervalo de confianza que contenga a la media de esta distribución, que a un nivel $\alpha = 0.001$ y desviación típica desconocida, esta determinado por:

$$\left(\bar{X} - t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

donde $n = 10$, $\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = 7.1040$, $S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = 0.1286$, y utilizando la tabla de la distribución t de Student $t_{n-1; 1-\frac{\alpha}{2}} = t_{9; 0.9995} = 4.78091$. Por tanto, el intervalo de confianza al nivel 0.001 es: $(6.9096, 7.2984)$

Y representa que la media del peso estará en dicho intervalo con una probabilidad de acierto del 99.9%.

8. Calcular un intervalo de confianza al nivel $\alpha = 0.05$ para σ^2 mediante las desviaciones que se producen en un proceso de fabricación cuya distribución es $N(0, \sigma)$ a partir de la muestra

1.2, -2.2, -3.1, -0.2, 0.5, 0.6, -2.1, 2.2, 1.3

SOLUCIÓN:

Sabiendo que el proceso de fabricación sigue una distribución normal de media conocida $\mu = 0$, un intervalo de confianza para la varianza σ^2 al nivel $\alpha = 0.05$ es el siguiente:

$$\left(n \frac{T}{\chi^2_{n; 1-\frac{\alpha}{2}}}, n \frac{T}{\chi^2_{n; \frac{\alpha}{2}}} \right)$$

donde $n = 9$, $T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} = 3.05333$, y utilizando la tabla de la distribución χ^2 , se tiene $\chi^2_{n; 1-\frac{\alpha}{2}} = \chi^2_{9; 0.975} = 2.7004$, es decir, (1.44458, 10.1763)

Es el intervalo que contendrá con un 95% de acierto las desviaciones que se producen en el proceso de fabricación.

9. Calcular qué tamaño muestral debemos tomar para obtener μ con una precisión de 0.001 a partir de una muestra de una población $N(\mu, 3)$.

SOLUCIÓN:

Teniendo en cuenta que el intervalo de confianza que contiene a la media μ de una población normal con varianza conocida es de la forma

$$\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

es decir, el error que cometemos al estimar μ mediante un intervalo de confianza al nivel $\alpha = 0.05$, es

$$error = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{3}{\sqrt{n}}$$

Por tanto, si en esta estimación deseamos obtener la media con una precisión de 0.001, tenemos que calcular n tal que el error que se cometa esté acotado por esta cantidad, $error \leq 0.001$, es decir:

$$1.96 \frac{3}{\sqrt{n}} \leq 0.001 \Leftrightarrow \frac{5.88}{0.001} \leq \sqrt{n} \rightarrow n = 5880^2 = 34574400$$

10. Para estudiar la efectividad de un medicamento contra la diabetes se mide la cantidad de glucemia en sangre antes y después de la administración de dicho medicamento, obteniéndose los resultados siguientes:

Antes	7.2	7.3	6.5	4.2	3.1	5.3	5.6
Después	5.2	5.4	5.3	4.7	4.1	5.4	4.9

Estimar la reducción producida por el medicamento.

SOLUCIÓN:

Suponiendo que las cantidades de glucemia siguen una distribución normal, las muestras obtenidas mediante las medicaciones anteriores y posteriores a la administración del medicamento son apareadas, es decir, sobre cada individuo de la población se obtiene una observación de la variable y posteriormente se repite la observación una vez que el individuo ha tomado el medicamento siendo esta última observación dependiente de la primera.

Para estudiar el efecto de dicho medicamento, resulta de gran interés la variable diferencia entre ambas mediciones, puesto que nos permitirá estimar la reducción o incremento de la glucemia provocada por este medicamento.

En este caso, la diferencia de ambas observaciones sigue una distribución normal de media la diferencia de ambas y desviación típica desconocida, $N(\mu_A - \mu_D, \sigma^2)$, por lo que un intervalo de confianza al nivel $\alpha = 0.05$, viene dado por:

$$\left((\bar{X}_A - \bar{X}_D) - t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, (\bar{X}_A - \bar{X}_D) + t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

donde $n=7$, $\bar{X}_A - \bar{X}_D = \frac{\sum_{i=1}^n (\bar{X}_{Ai} - \bar{X}_{Di})}{n} = 0.60$ y

$$S = \sqrt{\frac{\sum ((\bar{X}_{Ai} - \bar{X}_{Di}) - (\bar{X}_A - \bar{X}_D))^2}{n-1}} = 1.17473$$

y utilizando la tabla de la distribución t de Student, $t_{n-1; 1-\frac{\alpha}{2}} = t_{6; 0.975} = 2.4469$, es decir, la reducción

de glucemia por el medicamento estará contenida a un nivel 0.05 en el intervalo $(-0.486448, 1.68645)$, siendo la reducción estimada 0.6.

11. Se ha hecho un estudio sobre la proporción de enfermos de cáncer de pulmón detectados en hospital que fuman, obteniéndose que de 123 enfermos 41 de ellos eran fumadores. Obtener un intervalo de confianza para dicha proporción. Estudiar si dicha proporción puede considerarse igual a la proporción de fumadores en la población si ésta es de un 29%.

SOLUCIÓN:

Como la proporción de fumadores en la población de los enfermos de cáncer de pulmón detectados en un hospital puede modelizarse por una variable Bernoulli de parámetro p, entonces:

$$\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

es un intervalo de confianza al nivel $\alpha = 0.05$, y a partir de las observaciones realizadas, $n = 123$, tenemos que $\hat{p} = \frac{41}{123}$ y $z_{1-\frac{\alpha}{2}} = 1.96$, es decir: $(0.250023, 0.416643)$

Por otra parte, si consideramos la proporción de fumadores en la población global, $p = 0.29$, observamos que esta proporción se encuentra dentro del intervalo de confianza obtenido para un nivel 0.05, es decir, como la proporción de fumadores en la población pertenece al intervalo que contiene a la proporción de fumadores entre los enfermos de cáncer con una probabilidad de

acuerdo del 95%, podemos considerar que esta proporción de fumadores en los enfermos de cáncer se corresponde con la de los fumadores en la población global.

12. Sospechamos que nuestro cromatógrafo está estropeado, y queremos determinar si los resultados que nos proporciona son lo suficientemente precisos. Para ello, realizamos una serie de 8 mediciones del contenido de una solución de referencia que, sabemos, contiene 90% de un determinado compuesto. Los resultados que obtenemos son:

93.3, 86.8, 90.4, 90.1, 94.9, 91.6, 92.3, 96.5

Construir un intervalo de confianza al nivel de 95% para la varianza poblacional. ¿Que conclusiones podemos realizar?

SOLUCIÓN:

Sea la variable aleatoria X que representa el valor medio del contenido, si modelizamos el error incontrolable de medición por una variable aleatoria ε normal de media 0, y de varianza σ^2 desconocida, tenemos que $X \sim (\text{valor real}, \sigma^2)$, donde conocemos el valor real del contenido de la solución que es igual a 90. Por lo tanto, estamos en el caso en que la medida poblacional es conocida e igual a 90, y tenemos que construir un intervalo de confianza para la varianza poblacional, por lo que usamos el estadístico:

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

Cuya distribución en el muestreo es una χ_n^2 con n grados de libertad. Así, se deduce que, un intervalo de confianza del $100(1 - \alpha^2)\%$, para la varianza de una distribución normal con media conocida, está dado por:

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n-1; 1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n-1; \frac{\alpha}{2}}^2}$$

En nuestro caso, para un nivel de confianza del 95%, observamos en la tabla de la distribución χ^2 que $\chi_{n; 1-\frac{\alpha}{2}}^2 = \chi_{8; 0.975}^2 \approx 17.53$ y $\chi_{n; \frac{\alpha}{2}}^2 = \chi_{8; 0.025}^2 \approx 2.18$, y calculamos el numerador anterior:

$$\sum_{i=1}^n (x_i - \mu)^2 = (93.3 - 90)^2 + \dots + (96.5 - 90)^2 = 95.41$$

por lo que obtenemos que el intervalo de confianza pedido es:

$$5.44 \leq \sigma^2 \leq 43.77$$

Esto representa que el intervalo para la dispersión es $2.33 \leq \sigma \leq 6.62$ (es una dispersión grande), de donde deducimos que la precisión del cromatógrafo es insuficiente.

13. En el departamento de control de calidad de una empresa, se quiere determinar si ha habido un descenso significativo de la calidad de su producto entre las producciones de dos semanas consecutivas a consecuencia de un incidente ocurrido durante el fin de semana. Deciden tomar una muestra de la producción de cada semana, si la calidad de cada artículo se mide en una escala de 100, obtienen los resultados siguientes:

Semana 1	93	86	90	90	94	91	92	96
Semana 2	93	87	97	90	88	87	84	93

Suponiendo que las varianzas de la puntuación en las dos producciones son iguales, construye un intervalo de confianza para la diferencia de medias al nivel de 95%. Interpreta los resultados obtenidos.

SOLUCIÓN:

En primer lugar, observamos que se disponen de dos poblaciones, la primera corresponde a la producción de la primera semana mientras que la segunda corresponde a la de la segunda semana. En este sentido, introducimos las dos variables X_1 que mide la puntuación de calidad de un artículo de la primera semana, y X_2 para la segunda.

Además, en el caso en el que las varianzas en las dos poblaciones son desconocidas pero iguales, X_1 y X_2 se asumen normales e independientes, utilizamos el estadístico:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde $S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$, el cuál sigue una distribución $t_{n_1 + n_2 - 2}$ de Student de

$n_1 + n_2 - 2$ grados de libertad. Así, un intervalo de confianza al $100(1 - \alpha)\%$ para la diferencia entre medias de dos distribuciones normales, con varianzas desconocidas pero iguales es:

$$\bar{x}_1 - \bar{x}_2 - cS_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + cS_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde $c = t_{n_1+n_2-2; 1-\frac{\alpha}{2}}$. En particular, para los datos de este problema, se tiene

$\bar{x}_1 = 91.5, S_1^2 = 90.1, n_1 = 8, n_2 = 8, \bar{x}_2 = 89.9, S_2^2 = 17.8$, por lo que:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{7 * 9.1 + 7 * 17.8}{14} = 13.4$$

Y $c = t_{n_1+n_2-2; 1-\frac{\alpha}{2}} = t_{14; 0.975} \approx 2.145$, obteniendo el intervalo de confianza:

$$1.6 - 2.145\sqrt{13.4} \sqrt{\frac{1}{8} + \frac{1}{8}} \leq \mu_1 - \mu_2 \leq 1.6 + 2.145\sqrt{13.4} \sqrt{\frac{1}{8} + \frac{1}{8}}$$

es decir, el intervalo de confianza al nivel del 95% para la diferencia de medias es:

$$-2.31 \leq \mu_1 - \mu_2 \leq 5.56$$

Por último, podemos concluir que, con los datos de la muestra, es posible que la diferencia de las medias poblacionales sea igual o muy próximo a cero, en consecuencia no podemos afirmar que ha habido un descenso significativo de la calidad entre las dos semanas.

14. Eres el encargado de un departamento de producción en una fábrica y recibes un lote de 2000 piezas necesarias para la fabricación de un artículo. Tienes la responsabilidad de aceptar o rechazar el lote, si estimas que la calidad de éste no es suficiente. El fabricante te asegura que, en este lote, no hay más de 100 piezas defectuosas, pero decides tomar una muestra para estimar la proporción de las mismas.

a) ¿Cuántas piezas decides examinar para que, con un nivel de confianza del 95%, el error que cometes en la estimación de la proporción poblacional de defectuosas no sea mayor que 0.05?

b) Si decides tomar una muestra de 100 artículos escogidos al azar en el lote y realizas el recuento de piezas defectuosas en esta muestra, encontrado 4 artículos defectuosos. Construye para la proporción de defectuosos en el lote, un intervalo de confianza al nivel de 95% de confianza. ¿Se debe rechazar el lote?

SOLUCIÓN:

Observar que el tamaño de la población (el lote de piezas) es de 2000, por lo que podemos considerar que dicho tamaño es lo suficientemente grande como para que la podamos tratar como una población infinita.

a) Para determinar el tamaño muestral necesario para estimar la proporción poblacional si queremos que el error cometido sea menor que 0.05 con un nivel de confianza de $100(1 - \alpha)\%$, viene dado por:

$$n \geq \left(z_{1-\frac{\alpha}{2}} \right)^2 \frac{p(1-p)}{error^2}$$

Donde no conocemos el valor de la proporción exacta p , puesto esto es precisamente lo que queremos encontrar. Por lo tanto, debemos utilizar una aproximación de p , usando una muestra piloto o eligiendo el p más pesimista posible, es decir, el p que conlleva la elección del tamaño muestral más grande, $p=0.5$. así tenemos dos posibilidades:

1) el valor que, según el vendedor, es la proporción poblacional, $p = \frac{100}{2000} = 0.05$. En este caso, el error máximo admisible es $error = 0.05$ y el nivel de confianza 95%, $\alpha = 0.05$ y utilizando las tablas de la normal estándar $z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$. Por lo que se llega:

$$n \geq 1.96^2 * \frac{0.5 * 0.95}{0.0025} \approx 72.99$$

es decir, el valor de n que convendría usar es 73.

2) el valor más pesimista, $p=0.5$, y en este caso, de forma similar a la anterior, se tiene:

$$n \geq 1.96^2 * \frac{0.5 * 0.5}{0.0025} \approx 384.16$$

Por último, observar que tendremos que escoger entre las dos opciones anteriores según el tiempo, y/o el dinero, que nos llevará examinar cada artículo de la muestra.

b) en el caso de tomar una muestra de tamaño 100, en la cuál hemos encontrado 4 piezas defectuosas, tenemos que determinar un intervalo de confianza para la proporción poblacional.

Para ello, si consideramos la variable aleatoria \hat{P} que designa el estadístico proporción muestral, y p la proporción en la población, tenemos que la distribución el siguiente estadístico:

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Es aproximadamente una distribución normal estándar. A partir de esto, deducimos que un intervalo de confianza de nivel $100(1-\alpha)\%$ para la proporción p de la población, es de la forma:

$$\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \hat{p} \leq \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Y aplicándolo a los datos del problema, $p = \frac{4}{100} = 0.04$, $\alpha = 0.05$ y $z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$, por

lo que el intervalo de confianza al 95% para la proporción es:

$$0.04 - 1.96 \sqrt{\frac{0.04 * 0.96}{100}} \leq p \leq 0.04 + 1.96 \sqrt{\frac{0.04 * 0.96}{100}}$$

es decir; $0.002 \leq p \leq 0.078$

y teniendo en cuenta que el fabricante nos aseguraba que la proporción de defectuosos en el lote era de 0.05, concluimos que no está en contradicción con los resultados que hemos encontrado en nuestra muestra, por lo que decidimos confiar en el fabricante y aceptamos el lote.

15. Los tiempos de reacción, en mili segundos, de 17 sujetos frente a una matriz de 15 estímulos fueron los siguientes:

448, 460, 514, 488, 592, 490, 507, 513, 492, 534, 523, 452, 464, 562, 584, 507, 461

Suponiendo que el tiempo de reacción se distribuye Normalmente, determine un intervalo de confianza para la media a un nivel de confianza del 95%.

SOLUCIÓN:

Mediante los cálculos básicos obtenemos que la media muestral vale 505,35 y la desviación típica 42,54.

Buscando en las tablas de la t de Student con 16 grados de libertad, obtenemos que el valor que deja por debajo una probabilidad de 0,975 es 2,12.

Sustituyendo estos valores en la expresión del intervalo de confianza de la media tenemos:

$$(505,35 - 2,12 * 42,54 / 4, 505,35 + 2,12 * 42,54 / 4)$$

Operando (482,80, 527,90)

39. Una firma comercial encuesta a 100 individuos para conocer sus opiniones sobre la elección de dos productos alternativos A y B recientemente fabricados. El resultado de la encuesta arroja que el producto A lo han elegido 55 individuos y el producto B 45. Hallar un intervalo de confianza al 95% para la proporción de individuos que eligen cada producto.

SOLUCIÓN:

Estamos ante el caso del cálculo de intervalos de confianza para proporciones, ya que este problema es ajustable al caso en que $\hat{p} = x / n$ es la proporción estimada del número de veces que aparece un suceso de Bernoulli (los encuestados eligen A o B exclusivamente) de entre n repeticiones de un experimento (x designa el número de veces que aparece el suceso, es decir el número de veces que los encuestados eligen A o B). En este caso el intervalo de confianza para la proporción se basa en el siguiente estadístico:

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-p)}{n}}} \rightarrow N(0,1)$$

lo que nos lleva al intervalo de confianza para p definido por:

$$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-p)}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-p)}{n}}$$

En nuestro caso, para el producto A tenemos el intervalo:

$$\left[\frac{55}{100} - 1,96 \sqrt{\frac{0,45 * 0,55}{100}} \quad \frac{55}{100} + 1,96 \sqrt{\frac{0,45 * 0,55}{100}} \right] = [0,45 \quad 0,65]$$

Para el producto B tenemos el intervalo:

$$\left[\frac{45}{100} - 1,96 \sqrt{\frac{0,55 * 0,45}{100}} \quad \frac{45}{100} + 1,96 \sqrt{\frac{0,55 * 0,45}{100}} \right] = [0,35 \quad 0,55]$$

Como conclusión podemos decir que hay una probabilidad del 95% entre 0,45 y 0,65 de que el producto elegido sea el A, y hay una probabilidad entre 0,35 y 0,55 de que el producto elegido sea el B.

40. En un proceso de fabricación de pilas alcalinas se sabe que su duración media es de 1100 horas y que dicha duración sigue una distribución normal. El nuevo proceso busca reducir la dispersión de la duración de las pilas por lo que se hace necesario construir

intervalos de confianza para la citada dispersión con coeficientes de confianza 90% y 98%. Construir dichos intervalos a partir de una muestra de tamaño 20 cuya dispersión es $\sqrt{2240}$ horas.

SOLUCIÓN:

Estamos ante el caso del cálculo de intervalos de confianza para la desviación típica de una distribución normal con media conocida. En este caso el intervalo de confianza para la desviación típica se basa en el siguiente estadístico:

$$\frac{n\hat{\sigma}^2}{\sigma^2} \rightarrow X_{n-1}^2$$

lo que nos lleva al intervalo de confianza para la desviación típica definido por:

$$\left(\sqrt{\frac{n\hat{\sigma}^2}{X_{\alpha/2,n}^2}}, \sqrt{\frac{n\hat{\sigma}^2}{X_{1-\alpha/2,n}^2}} \right)$$

En nuestro caso, para el nivel del 90% tenemos:

$$1 - \alpha = 0,9 \Rightarrow \alpha = 0,1 \Rightarrow \frac{\alpha}{2} = 0,05 \Rightarrow X_{0,05;20}^2 = 31,4, \quad X_{0,95;20}^2 = 10,9$$

El intervalo de confianza será entonces:

$$\left(\sqrt{\frac{20 * 2240}{31,4}}, \sqrt{\frac{20 * 2240}{10,9}} \right) = [119,42 \quad 203,2]$$

Para el nivel del 95% tenemos:

$$1 - \alpha = 0,98 \Rightarrow \alpha = 0,02 \Rightarrow \frac{\alpha}{2} = 0,01 \Rightarrow X_{0,01;20}^2 = 37,6, \quad X_{0,99;20}^2 = 8,26$$

El intervalo de confianza será entonces:

$$\left(\sqrt{\frac{20 * 2240}{37,6}}, \sqrt{\frac{20 * 2240}{8,26}} \right) = [109,2 \quad 232,88]$$

Podemos concluir entonces que hay una probabilidad del 90% de que la dispersión de la duración de las pilas en el nuevo proceso de fabricación esté entre 119,42 y 203,2 horas, y hay una probabilidad del 98% de que la dispersión de la duración de las pilas en el nuevo proceso de fabricación esté entre 109,2 y 232,88 horas.

53. El tiempo de vida en días X de los individuos de una población afectados de una nueva enfermedad es una variable aleatoria continua con función de densidad

$$f_{\phi}(x) = 2\phi^2 x^{-3} \quad \text{si } x > 0$$

Y $f_{\phi}(x) = 0$ si $x \leq \phi$ siendo $\phi > 0$ un parámetro desconocido.

Con objeto de estimar el parámetro ϕ , se extrajo una muestra aleatoria simple de dicha población, obteniéndose los siguientes tiempos de vida, en días, de los 10 individuos seleccionados, todos los cuales fallecieron por la enfermedad en estudio.

398, 356, 615, 265, 650, 325, 400, 223, 368, 680

Determinar la estimación de la máxima verosimilitud de ϕ .

SOLUCIÓN:

La función de densidad de la variable aleatoria en estudio X, tiempo de vida de los individuos de la población afectados por la enfermedad en estudio, nos indica que dichos individuos contraen la enfermedad de un momento desconocido, ϕ , (puesto que en este punto la función de distribución F(x) empieza a crecer desde cero, o lo que es lo mismo la función de supervivencia $S(x) = 1 - F(x)$ vale 1, lo que quiere decir que todos los individuos están vivos), momento a partir del cual, y por la forma de dicha función de densidad, la probabilidad de sobrevivir va disminuyendo.

Es precisamente el inicio de las enfermedades objeto de la estimación. Para ello, utilizaremos el método de la máxima verosimilitud (CB-sección 5.2 o EII-sección 2.2). La función de verosimilitud de la muestra será:

$$L(\phi) = f_{\sigma}(\chi_1, \dots, \chi_n) = \prod_{i=1}^n f_{\sigma}(\chi_i) = 2^n \phi^{2n} \left(\prod_{i=1}^n x_i \right)$$

Si $\chi_1, \dots, \chi_n > \phi$

Como siempre, el método de la máxima verosimilitud se basa en asignar a ϕ el valor que maximice la función $L(\phi)$; el problema es que ahora ϕ aparece en el recorrido de la variable, es decir, que $L(\chi_1)$ toma un valor distinto de cero si $\phi < \chi_1, \dots, \chi_n$, y si algún χ_1 , es tal que $\chi_1 \leq \phi$ será $L(\phi) = 0$. En la estimación de ϕ habrá que tener también en cuenta, por tanto, el recorrido de $L(\phi)$

La función $L(\phi) = 2^n \phi^{2n} \prod_{i=1}^n x_i^{-3}$ crece al crecer ϕ , por lo que será tanto mayor cuanto mayor sea ϕ , y esto hasta que ϕ llegue al primer χ_1 , a partir de donde, por los comentarios anteriores, $L(\phi)$ vale cero. Por tanto, el valor de ϕ que hace máxima $L(\phi)$ es el mínimo de los n valores $\{\chi_1, \dots, \chi_n\}$ el cual se suele denotar por $x_{(1)}$.

La estimación de máxima verosimilitud de ϕ será, a partir de los 10 datos de muestra, $x_{(1)}=223$.

Este problema es un ejemplo de lo importante que es la determinación del estimador de máxima verosimilitud es calcular el valor de ϕ que maximiza la función $L(\phi)$, máximo que en muchas ocasiones se podrá determinar derivando $L(\phi)$ respecto a ϕ e igualando a cero dicha derivada, pero que en otras ocasiones, como pasa en general al determinar el máximo de cualquier función, deberá utilizarse otras herramientas distintas de la derivada.

54. Una muestra de tamaño 10 de una población de mujeres presenta una altura media de 172 cm. y una muestra de 12 varones de otra población presenta una altura media de 176,7 cm. Sabiendo que ambas poblaciones son normales con varianzas 225 y 256 respectivamente, se trata de analizar si con una probabilidad del 95% se puede asegurar que los varones son más altos en media que las mujeres o viceversa.

SOLUCIÓN:

Para resolver este problema hallaremos un intervalo de confianza para la diferencia de medias al 95% y comprobaremos si dicho intervalo contiene el valor cero, en cuyo caso se puede aceptar la hipótesis de que las alturas medias son iguales con una probabilidad del 95%. Estamos entonces ante el caso del cálculo de intervalos de confianza para la diferencia de medias de dos distribuciones normales con varianzas conocidas. En esta situación el intervalo de confianza para la diferencia de medias se basa en el siguiente estadístico:

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

lo que nos lleva al intervalo de confianza para la varianza definido por:

$$\bar{x}_1 - \bar{x}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

El intervalo de confianza será entonces:

$$170,2 - 176,7 \pm 1,96 \sqrt{\frac{225^2}{10} + \frac{256^2}{12}} = [-207,56 \quad 194,56]$$

Como este intervalo contiene el valor cero, se puede aceptar con una probabilidad del 95% que $\bar{x}_1 - \bar{x}_2 = 0$, es decir, $\bar{x}_1 = \bar{x}_2$. Por tanto, aceptaremos la hipótesis de que la estatura

de los varones y las mujeres de ambas poblaciones es similar con un coeficiente de confianza del 95%.

55. Los responsables municipales de la salud miden la radiactividad en el agua de una fuente natural en una zona abundante en granito. Realizadas 12 mediciones en diferentes fechas del año se observó una media de 3,6 picocurios con una desviación típica de 0,82. determinar, al 95% y al 99%, intervalos de confianza para la radiación media y para la varianza.

SOLUCIÓN:

Tenemos los siguientes datos muestrales

$$N=12, \quad \bar{x} = 3,6 \quad y \quad s = 0,82.$$

Para determinar el intervalo de confianza para σ^2 tenemos en cuenta que

$$\frac{12S^2}{\sigma^2}$$

Es un variable χ^2 con 11 grados de libertad. Los percentiles necesarios para la determinación de los intervalos al 95% y 99% de confianza son

$$\chi_{11;0,005}^2 = 2,603 \quad \chi_{11;0,025}^2 = 3,816 \quad \chi_{11;0,975}^2 = 21,92 \quad y \quad \chi_{11;0,995}^2 = 26,8.$$

Así el intervalo de confianza al 95% es

$$\left(\frac{12 \cdot 0,82^2}{21,92}; \frac{12 \cdot 0,82^2}{3,816} \right)$$

Y al 99%

$$\left(\frac{12 \cdot 0,82^2}{26,8}; \frac{12 \cdot 0,82^2}{2,603} \right)$$

Efectuando los cálculos se obtiene, respectivamente,

$$(0,3681; 2,1145) \quad y \quad (0,3011; 3,0998).$$

En lo relativo a la media, el intervalo se determina teniendo en cuenta que

$$\frac{\bar{x} - \mu}{S / \sqrt{11}}$$

Es una variable t de Student con 11 grados de libertad. El intervalo, para un nivel de confianza $p=1-\alpha$, viene dado por

$$\bar{x} \pm t_{11; 1-\alpha/2} \frac{s}{\sqrt{11}}$$

Siendo $t_{11; 1-\alpha/2}$ el valor del percentil $100 \cdot (1-\alpha/2)$ para 11 grados de libertad. Concretamente

$$t_{11; 0,975}=2,201 \quad \text{y} \quad t_{11; 0,995}=3,1058$$

y los intervalos de confianza son, respectivamente,

$$3,6 \pm 2,201 \frac{0,82}{\sqrt{11}} \quad \text{y} \quad 3,6 \pm 3,1058 \frac{0,82}{\sqrt{11}},$$

$$\text{Obteniéndose } 3,6 \pm 0,544 \quad \text{y} \quad 3,6 \pm 0,768.$$

56. Para m.a.s. de tamaño n, de una población con funcion de densidad

$$f(x|\theta) = \frac{5x^4}{\theta^5} I(0,\theta](x) \quad \text{con} \quad \theta > 0$$

a) Hállese el estimador de máxima verosimilitud T.

b) Determínese la región de confianza de grado 0,95 para θ , de la forma $(\lambda T, +\infty)$ para λ conveniente.

SOLUCIÓN:

a) La función de densidad de la muestra es

$$f(x_1, \dots, x_n | \theta) = 5^n \frac{\prod_{i=1}^n x_i^4}{\theta^{5n}} I_{[x_{(n)}, +\infty)}(\theta)$$

por lo que el EMV es $\hat{\theta} = X_{(n)}$.

a) La región de confianza, de la forma $(\lambda X_{(n)}, +\infty)$ estará de terminada, cuando se obtenga el valor de λ tal que

$$P\{\lambda X_{(n)} \leq \theta | \theta\} = 1 - \alpha \quad \text{para todo } \theta > 0$$

para lo que ha de cumplirse que

$$P\{X_{(n)} \leq y | \theta\} = P^n\{X \leq y | \theta\} = \left(\frac{y}{\theta}\right)^{5n}$$

De donde ha de ser $\lambda^{5n} = 1 - \alpha$; o lo que es lo mismo $\lambda = (1 - \alpha)^{\frac{1}{5n}}$ y

$$RC_{1-\alpha}(\theta) = \left((1 - \alpha)^{\frac{1}{5n}} X_{(n)} + \infty \right).$$

57. Determínese, con m.a.s de tamaño n, el EMV para θ en los siguientes casos:

a) $f(x | \theta) = \frac{1}{\theta} I_{\{1,2,\dots,\theta\}}(x)$ con θ entero positivo

b) $f(x | \theta) = e^{-x+\theta} I_{[\theta,+\infty)}(x)$ con $\theta \in \mathbb{R}$

SOLUCIÓN:

a) La función de verosimilitud es

$$L(\theta | x_1, \dots, x_n) = \frac{1}{\theta^n} \prod_{i=1}^n I_{\{1,2,\dots,\theta\}}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^n I_{\{x_{(n)}, x_{(n)}+1, \dots\}}(x_i) = I_{\{x_{(n)}, x_{(n)}+1, \dots\}}(\theta)$$

a medida que θ aumenta uno a uno desde $x_{(n)}$ esta función disminuye,

luego es $\hat{\theta} = X_{(n)}$

b) La función de verosimilitud es

$$L(\theta | x_1, \dots, x_n) = e^{-\sum_{i=1}^n x_i + n\theta} \prod_{i=1}^n I_{[\theta, +\infty)}(x_i) = e^{-\sum_{i=1}^n x_i + n\theta} I_{(-\infty, x_{(1)}]}(\theta)$$

expresión que, como función de θ , es la máxima cuando θ es lo mayor posible, es decir es $\hat{\theta} = X_{(1)}$

58. En una muestra de 65 sujetos las puntuaciones en una escala de extroversión tienen una media de 32,7 puntos y una desviación típica de 12,64.

a) Calcule a partir de estos datos el correspondiente intervalo de confianza, a un nivel del 90%, para la media de la población.

b) Indique, con un nivel de confianza del 95%, cual sería el máximo error que podríamos cometer al tomar como media de la población el valor obtenido en la estimación puntual.

SOLUCIÓN:

a) Buscando en las tablas de la t de Student obtenemos que el valor que deja por debajo una probabilidad del 95% es 1,671 (aproximadamente). Sustituyendo los valores de esta muestra en la expresión del intervalo de confianza obtenemos:

$$(32,7 - 1,671 \cdot 12,64 / 8, 32,7 + 1,671 \cdot 12,64 / 8)$$

Operando

$$(30,06, 35,34)$$

b) En las tablas de la t de Student encontramos que el valor de la variable que deja por debajo una probabilidad de 0,975 es 2. En consecuencia a un nivel de confianza del 95% la media de la población puede valer

$$32,7 \pm 2 \cdot 12,64 / 8$$

luego el máximo error que se puede cometer, a este nivel de confianza, es: 3,16

59. Los tiempos de reacción, en mili segundos, de 17 sujetos frente a una matriz de 15 estímulos fueron los siguientes: 448, 460, 514, 488, 592, 490, 507, 513, 492, 534, 523, 452, 464, 562, 584, 507, 461.

Suponiendo que el tiempo de reacción se distribuye Normalmente, determine un intervalo de confianza para la media a un nivel de confianza del 95%.

SOLUCIÓN:

Mediante los cálculos básicos obtenemos que la media muestral vale 505,35 y la desviación típica 42,54.

Buscando en las tablas de la t de Student con 16 grados de libertad, obtenemos que el valor que deja por debajo una probabilidad de 0,975 es 2,12

Sustituyendo estos valores en la expresión del intervalo de confianza de la media tenemos:

$$(505,35 - 2,12 \cdot 42,54 / 4, 505,35 + 2,12 \cdot 42,54 / 4)$$

Operando

$$(482,80, 527,90)$$

60. Obtener el E.M.V. del parámetro λ de una distribución de Poisson.

SOLUCIÓN:

Para una muestra de tamaño n tendremos que la función de verosimilitud será:

$$L(x_1, x_2, \dots, x_n / \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \frac{e^{-\lambda} \lambda^{x_2}}{x_2!} \dots \frac{e^{-\lambda} \lambda^{x_n}}{x_n!}$$

maximizar L será equivalente a minimizar el numerador de L , llamémosle L'

$$L' = e^{-n\lambda} \lambda^{\sum x_i}$$

maximizar L' es equivalente a maximizar su logaritmo $\ln L'$

$$\ln L' = -n\lambda + \sum x_i \ln \lambda$$

$$\frac{d \ln L'}{d \lambda} = -n + \sum x_i \frac{1}{\lambda} = 0$$

lo que se cumple para $\lambda = \frac{\sum x_i}{n} = \bar{x}$

Por lo tanto el estimador maximo-verosimil de λ es $\hat{\lambda} = \bar{x}$

61. De una población cuya distribución se desconoce se obtiene una muestra aleatoria de 2000 valores en que la media muestral resulta ser 225 y la desviación típica muestral 10. Suponiendo que la varianza muestral coincide con la de la población, estimar un intervalo para la media de la población con un nivel de confianza del 95%

SOLUCIÓN:

Según hemos visto

$$\mu \in \left[\bar{x} - \frac{\sigma}{\sqrt{n\sigma}}, \bar{x} + \frac{\sigma}{\sqrt{n\sigma}} \right] \text{ con una confianza de } 1 - \sigma$$

En este caso:

$$1 - \sigma = 0,95$$

$$\sigma = 0,05$$

$$\bar{x} = 225$$

$$s = 10$$

$$n = 2000$$

Sustituyendo $\mu \in [224, 226]$ con el 95% de confianza

62. La longitud, en centímetros, de las piezas fabricadas por una cierta máquina se distribuye según una $N(10, 0.25)$. Para muestras de tamaño 25, calcular:

a) $P(9.68 \leq \bar{X} \leq 10.1)$

b) $P(S^2 \leq 0.19)$

SOLUCIÓN:

a) La media muestral de muestras de tamaño n procedentes de una población $N(\mu, \sigma^2)$ se distribuye según una $N(\mu, \sigma^2/n)$. En nuestro caso, la media muestral lo hará según una ley $N(10, 0.01)$. Por tanto,

$$P(9.68 \leq \bar{X} \leq 10.1) = P(-3.2 \leq Z \leq 1) = 0.8406$$

Siendo una normal típica.

b) El teorema de Fisher afirma que

$$\frac{ns^2}{\sigma^2}$$

Se distribuye según una ley chi-cuadrado con $n-1$ grados de libertad. Por tanto, se tiene que

$$P(S^2 \leq \bar{X} \leq 0.19) = P\left(\frac{25 S^2}{0.25} \leq \frac{25 * 0.19}{0.25}\right) = P(U \leq 19) = 0.25$$

Donde U se distribuye χ^2_{24}

63. De una población normal de media y desviación típica desconocida se ha obtenido una muestra de 26 elementos que tiene como media aritmética 5 y desviación típica 1.2. ¿Cuál es la probabilidad de que la media poblacional sea superior a 5.3?

SOLUCIÓN:

Sabemos que la variable

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

Se distribuye según una ley t de Student con n-1 grados de libertad, y lo que se pide es

$$P\left(Z < \frac{5 - 5.3}{1.2} \sqrt{25}\right) = P(Z < -1.25) = 0.114$$

Donde Z sigue una t de Student con 25 grados de libertad.

64. De una población normal de varianza c^2 se extraen dos muestras de tamaño n. Hallar cuál debe ser el tamaño de ambas muestras, para que la probabilidad de que las medias muestrales difieran en más de $2c$, sea 0.05.

SOLUCIÓN:

La distribución de las dos medias muestrales es, suponiendo que μ es la media de la población, $N(\mu, c^2/n)$.

$$\bar{X} - \bar{Y} : N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)$$

Como en este caso ambas muestras provienen de la misma población, la distribución de Z es:

$$Z = \bar{X} - \bar{Y} : N\left(0, \frac{2\sigma^2}{n}\right)$$

Y la probabilidad del suceso

$$|Z| = \left| \bar{X} - \bar{Y} \right| > 2c$$

Será

$$P(|Z| > 2c) = 1 - P(|Z| < 2c) = 1 - P(-2c < Z < 2c) = 0.05$$

O bien

$$P(-2c < Z < 2c) = 0.95 \rightarrow P(0 < Z < 2c) = 0.475 \rightarrow P(Z < 2c) = 0.975$$

Y pasando a una $N(0,1)$, Y , tendremos

$$P(Z < 2c) = P\left(Y < \frac{2c}{\sqrt{2/n}}\right) = P(Y < \sqrt{2n}) = 0.975$$

De donde la longitud, $\sqrt{2n} = 1.95$ esto es, $n=2$.

65. En una muestra de 100 personas de un barrio de Pamplona se ha observado una proporción de 0.18 personas que leen el periódico diariamente. ¿Puede ser que la verdadera proporción de personas que leen el periódico en ese barrio sea 0.20?

SOLUCIÓN:

$$\sigma_p = \sqrt{\frac{p_{pob} * q_{pob}}{N}} = \sqrt{\frac{0.20 * 0.8}{100}} = 0.04 ;$$

$$Z = \frac{p_{pob} * q_{pob}}{N} = \frac{0.18 - 0.20}{0.04} = -0.5 ;$$

$$Z = -0.5 > -1.645 \quad \text{ó} \quad |0.5| < |1.645|$$

Por consiguiente no puede rechazarse la hipótesis nula. Es decir, nada se opone a aceptar que la muestra procede de tal barrio donde un 20% de personas leen el periódico diariamente.

Con frecuencia p_{pob} es desconocida y, por lo tanto, también q_{pob} es desconocida. El error típico σ_p se calcula entonces, salvo excepciones, empleando la proporción p_m de la muestra y $q_m = 1 - p_m$;

$$\sigma_p = \sqrt{\frac{p_m * q_m}{N}}$$

Si en lugar de proporciones se toma la frecuencia absoluta el error típico de la frecuencia absoluta será: $\sigma_f = \sqrt{N * p_{pob} * q_{pob}}$ ya que $f = N * p$ y que $\sigma_f = N * \sigma_p$

66. Consideramos dos conjuntos de puntuaciones en un test de intuición sociológica, ambos al mismo grupo de 20 estudiantes. Llamamos \bar{X}_1 a la media de puntuaciones

obtenidas antes de que comenzase el curso y \bar{X}_2 a la media obtenida después de concluir el curso. Interesa determinar si \bar{X}_2 es significativamente diferente de \bar{X}_1 .

SOLUCIÓN:

Los datos son los siguientes:

Antes del curso	Después del curso	Diferencia
18	20	2
16	22	6
18	24	6
12	10	-2
20	25	5
17	19	2
18	20	2
20	21	1
22	23	1
20	20	0
10	10	0
8	12	4
20	22	2
12	14	2
16	12	-4
16	20	4
18	22	4
20	24	4
18	23	5
21	17	-4

$$\bar{X}_1 = 16.5 \quad \bar{X}_2 = 18.5 \quad \bar{D} = \frac{40}{20} = 2 \quad \sum (D - \bar{D})^2 = 164$$

Para resolver el problema partimos de la hipótesis nula: no hay diferencia significativa entre las dos medias; es decir, puede ser cero por azar.

1. Cálculo de la desviación típica de las diferencias:

$$S_D = \sqrt{\frac{\sum (D - \bar{D})^2}{N}} = \sqrt{\frac{164}{20}} = 2.863$$

2. Cálculo del error típico de la media de las diferencias:

$$\sigma_D = \frac{S_D}{\sqrt{N-1}} = \frac{2.863}{\sqrt{19}} = \frac{2.863}{4.3588} = 0.657$$

3. Hallamos la razón crítica t:

$$t = \frac{|\bar{D}|}{\sigma_D} = \frac{2}{0.657} = 3.044$$

4. Suponemos que estamos trabajando al nivel de confianza 1%. A ese nivel, buscamos en las tablas de distribución t de Student, la t correspondiente para $N - 1 = 20 - 1 = 19$ grados de libertad: $t = 2.54$ (se trata de prueba unilateral).

5. Como la t hallada $= 3.044 > 2.54$ rechazamos la hipótesis nula con 99% de probabilidades de acertar o lo que es lo mismo, a lo más con 1% de riesgo de equivocarnos; podemos decir que tal diferencia es significativa y, por lo tanto, que la media después de finalizar el curso es significativamente mayor que la media antes de comenzar el curso, al nivel de confianza 1%.

67. En una «reunión» una droga fue tomada por 14 personas, de las cuales 6 lo hacen por primera vez y 8 ya son habituales de ella. La droga produjo en el primer grupo sueños de duración 11,12,13,16,17 y 15 horas, mientras que en el segundo grupo 8,7,9,10,6,7,9 y 8 horas. Se pide:

- a) **Media y desviación típica de cada grupo**
- b) **Formar el estadístico que se distribuye según una t de Student de 12 grados de libertad, sabiendo que las poblaciones tienen la misma media y desviación típica.**

SOLUCIÓN:

- a) Grupo 1

$$\bar{x}_1 = \frac{11+12+13+16+17+15}{6} = 14$$

$$s_1^2 = \frac{(11-14)^2 + (12-14)^2 + (13-14)^2 + (16-14)^2 + (17-14)^2 + (15-14)^2}{6} = 4.67$$

Grupo 2

$$\bar{x}_2 = \frac{8+7+9+10+6+7+9+8}{8} = 8$$

$$s_2^2 = \frac{(8-8)^2 + (7-8)^2 + (9-8)^2 + (10-8)^2 + (6-8)^2 + (7-8)^2 + (9-8)^2 + (8-8)^2}{8} = 1.5$$

$$b) \ t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} * \frac{\sqrt{n_1 + n_2 - 2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{14 - 8}{\sqrt{6 * 4.67 + 8 * 1.5}} * \frac{\sqrt{6 + 8 - 2}}{\sqrt{\frac{1}{6} + \frac{1}{8}}} = 6.084$$

68. Según una encuesta realizada sobre una muestra de 2500 personas elegidas al azar, el 80% está decidida a votar en las últimas elecciones.

a) ¿Puede ser cierto que llegue a votar el 85% de la población?

b) Con un 99% de nivel de confianza ente qué valores estará el porcentaje de los votantes de la población.

SOLUCIÓN:

a) Se trata de una comprobación de hipótesis.

$$\sigma_p = \sqrt{\frac{0.85 * 0.15}{2500}} = 0.0071 \quad \left(\begin{array}{l} p_{pob} = 0.85 \\ q_{pob} = 1 - p_{pob} = 0.15 \end{array} \right)$$

Establecemos la hipótesis nula, la cual dice que puede ser cierto que vote el 85% de la población.

Hallamos la razón crítica

$$Z = \frac{0.85 * 8.15}{0.0071} = -7.04$$

Una $Z = |7.04| > 2.33$. Luego, de ser cierto que en la población votase el 85%, en una muestra de 2500 personas no podía haber salido un porcentaje de 80%. Por ello, rechazamos la hipótesis nula y afirmamos con 99% de probabilidades de acertar que no llegará a votar el 85% en circunstancias normales (= circunstancias similares al momento en que se hubiera realizado el muestreo).

$$\sigma_p = \sqrt{\frac{0.80 * 50.20}{2500}} = 0.008$$
$$0.80 \pm 2.58 \sigma_p = 0.80 \pm 2.58 * 0.008 \Rightarrow 0.7793$$

$$\Rightarrow 0.8220$$

69. La renta media de los habitantes de un país se distribuye uniformemente entre 4,0 millones pesetas. Y 10,0 millones ptas. Calcular la probabilidad de que al seleccionar al azar a 100 personas la suma de sus rentas supere los 725 millones ptas.

$$\mu = \frac{(4+10)}{2} = 7$$

$$\sigma^2 = \frac{(10-4)^2}{12} = 3$$

Media: $n * \mu = 100 * 7 = 700$

Varianza: $n * \sigma^2 = 100 * 3 = 300$

$$Y = \frac{725 - 700}{17,3} = 1,44$$
$$P(x > 725) = P(Y > 1,44) = 1 - P(Y < 1,44) = 1 - 0,9251 = 0,0749$$

Es decir, la probabilidad de que la suma de las rentas de 100 personas seleccionadas al azar supere los 725 millones de pesetas es tan sólo del 7,49%.

70. En una asignatura del colegio la probabilidad de que te saquen a la pizarra en cada clase es del 10%. A lo largo del año tienes 100 clases de esa asignatura. ¿Cuál es la probabilidad de tener que salir a la pizarra más de 15 veces?

SOLUCIÓN:

Salir a la pizarra es una variable independiente que sigue el modelo de distribución de Bernouilli:

"Salir a la pizarra", le damos el valor 1 y tiene una probabilidad del 0,10

"No salir a la pizarra", le damos el valor 0 y tiene una probabilidad del 0,9

La media y la varianza de cada variable independiente es:

$$\mu = 0,10$$

$$\sigma^2 = 0,10 * 0,90 = 0,09$$

Por tanto, la suma de las 100 variables se distribuye según una normal cuya media y varianza son:

$$\text{Media: } n * \mu = 100 * 0,10 = 10$$

$$\text{Varianza: } n * \sigma^2 = 100 * 0,09 = 9$$

Para calcular la probabilidad de salir a la pizarra más de 15 veces, calculamos el valor equivalente de la variable normal tipificada:

$$Y = \frac{15 - 10}{3} = 1,67$$

Luego:

$$P(X > 15) = P(Y > 1,67) = 1 - P(Y < 1,67) = 1 - 0,9525 = 0,0475$$

Es decir, la probabilidad de tener que salir más de 15 veces a la pizarra a lo largo del curso es tan sólo del 4,75%.

107. Un fabricante de refrescos sin burbujas desea sacar al mercado una variedad de su producto que tenga burbujas. Su director comercial opina que al menos el 50% de los consumidores verá con buenos ojos la innovación. Se realiza un sondeo de mercado y resulta que de 100 consumidores encuestados 40 son favorables a la innovación.

a) Contrastar la hipótesis del director comercial frente a la alternativa de que el % de aceptación es inferior, con un nivel de significación del 2,5%.

b) Si es aceptable la hipótesis de que el % de aceptación del nuevo producto es inferior o igual al 30% el fabricante decidirá no fabricarlo. Si es aceptable el criterio del director comercial entonces sí fabricarán el refresco con burbujas. Y si ninguna de las 2 hipótesis es aceptable procederán a hacer otro sondeo. Para tomar esta decisión trabajarán con un nivel de significación del 2,5% ¿ Por qué opción se decantarán?

SOLUCIÓN:

a)

$$H_0: p = 0,5$$

$$H_1: p < 0,5$$

$$\alpha = 0,025 \quad -\lambda_\alpha / \phi(-\lambda_\alpha) = 0,025 \quad -\lambda_\alpha \text{ es negativo}$$

y su opuesto es tal que $(+\lambda_\alpha) = 1 - 0,025 = 0,975$, así que $-\lambda_\alpha = -1,96$

$$\hat{p} = 0,4 \quad n = 100 \gg 30$$

$-2 < -1,96$ así es que rechazamos que $p = 0,5$

b)

$$H_0: p = 0,3$$

$$H_1: p > 0,3$$

$$\alpha = 0,025 \quad \lambda_\alpha = 1,96$$

$$\hat{p} = 0,4 \quad n = 100$$

$$\frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0,1}{0,5} = 2$$

$2 > 1,96$ así que rechazamos que $p = 0,3$

Como no se aceptan ninguna de las dos hipótesis habrá que hacer otro sondeo.

108. Contrastar la hipótesis de que dos poblaciones tienen la misma dispersión con un nivel de significación del 1% y sabiendo que la desviación típica de una muestra realizada sobre la primera población era 12 con un tamaño muestral de 25 y que en una muestra sobre la segunda de tamaño 30 la desviación típica resultó ser 7. Considérese que ambas poblaciones son normales.

SOLUCIÓN:

$$S^2x = 144$$

$$S^2y = 49 \quad \frac{S^2x}{S^2y} * \frac{n_x}{n_y} * \frac{(n_y - 1)}{(n_x - 1)} = \frac{144 * 25 * 29}{49 * 30 * 24} = 2.95918$$

$$F_{24,29} (\alpha=0.01) = 2.49$$

Como $2.95918 > 2.49$, rechazamos la hipótesis de que las varianzas sean iguales.

109. Una empresa de publicidad desea comprobar si un determinado programa de televisión es visto por el 30% de la audiencia. Para ello de escogen al azar 200 familias y resulta que de ellas 50 lo están viendo. Contrastar la hipótesis con una nivel de significación del 5%.

SOLUCIÓN:

$$H_0: p = 0.3$$

$$H_1: p \neq 0.3$$

$$\hat{p} = 50/200 = 0.25 \quad \frac{\hat{p} - p_0}{\sqrt{p_0 p_0 / n}} = -1.5430$$

$$n = 200$$

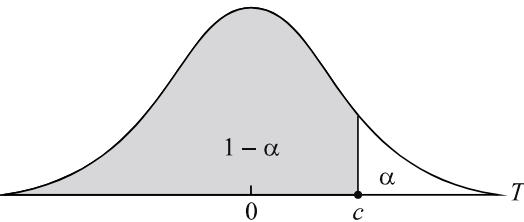
$$\alpha = 0.05 \quad \lambda_{\alpha/2} = 1.96$$

Como $-1.5430 \in [-1.96, 1.96]$ aceptamos la hipótesis de que sea visto por el 30% de la audiencia.

ANEXO

TABLA DE LA DISTRIBUCION t -Student

La tabla da áreas $1 - \alpha$ y valores $c = t_{1-\alpha, r}$, donde, $P[T \leq c] = 1 - \alpha$, y donde T tiene distribución t -Student con r grados de libertad...



$1 - \alpha$

r	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819

23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
□	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576