Ejercicios T4 (a) INTERVALOS DE CONFIANZA

1. Se quiere obtener un intervalo de confianza para el valor de las ventas medias por hora que se producen en un kiosco. Para ello realizamos una muestra consistente en elegir al azar las ventas que se realizaron durante 1000 horas distintas; muestra cuyos resultados fueron: ventas medias por hora 4000 pts, y varianza de dicha muestra 4000 pts al cuadrado. Obtener dicho intervalo con un nivel de confianza del 95.5 %.

Queremos construir un intervalo para la media con las siguientes características:

tamaño muestral = n =1000 , muestreo aleatorio simple

la población no es normal ni conocemos su varianza,
el resultado de la muestra es:

$$\bar{x} = 4000; s^2 = 4000$$

si bien se trata de un intervalo para la media con varianza desconocida y población no normal, dado que el tamaño muestral es grande podemos suponer normalidad y tomar como varianza poblacional a la muestral así:
(ir a script de solución)

$$P\left(\bar{x} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

dado que para nivel de confianza del 95,5% las valores de
(ir a script de la normal) son según tablas 2 ,-2 tendremos el intervalo:

$$P\left(4000 - 2\frac{63,24}{\sqrt{1000}} \le \mu \le 4000 + 2\frac{63,24}{\sqrt{1000}}\right) = 0.955$$

luego el intevalo será

$$\mu \in [3996;4004] \text{ con nivel de confianza del } 95,5\%$$

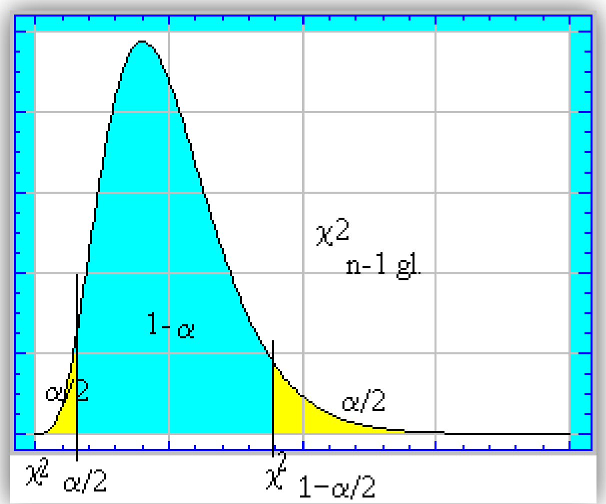
2. Resolver el ejercicio anterior si consideramos los mismos resultados muestrales pero la muestra ha sido de 26 horas.

Al disminuir el tamaño muestral de 1000 a 26, no podemos trabajar con supuesta normalidad, dado que la población no es Normal no podemos construir dicho intervalo.

3. Obtener el intervalo de confianza para la varianza de una población normal con muestreo aleatorio simple, y nivel de confianza 1-α .

Queremos construir un intervalo para la varianza de una población normal conocemos que con normalidad

$$\frac{nS^2}{\sigma^2} \Rightarrow \chi^2 \text{ con } (n-1) \text{ g.l.}$$



de manera que para un nivel de confianza dado,
un tamaño muestral dado (n) y realizada (conocida) la varianza muestral:

podemos construir un intervalo

$$P\left(\chi^2_{\alpha/2} \leq \frac{nS^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}\right) = 1-\alpha$$

establecido según se aprecia en el dibujo.

Dado que el intervalo que queremos construir lo es para la varianza poblacional, que se encuentra en el denominador;

invertimos el intervalo y así:

$$P\left(\frac{1}{\chi^2_{\alpha/2}} \geq \frac{\sigma^2}{nS^2} \geq \frac{1}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

despejando:

$$P\left(\frac{nS^2}{\chi^2_{\alpha/2}} \geq \sigma^2 \geq \frac{nS^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

por lo que el intervalo quedaría como:

$$P\left(\frac{nS^2}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq \frac{nS^2}{\chi^2_{\alpha/2}}\right) = 1-\alpha$$

4. Se desea determinar un intervalo de confianza con nivel de confianza del 99% para la proporción de amas de casa que compran sólo una vez a la semana. Si se sabe que en una muestra aleatoria simple de 400 amas de casa sólo 180 de afirmaron comprar una vez a la semana.

Se trata de un intervalo para la proporción de una característica
P = proporción de amas de casa que compran una sola vez en la semana.

Conocemos que: el nivel de confianza es 0,99 , el tamaño muestral n=400, y,

la proporción muestral resultante , el muestreo es aleatorio simple.

(ir a script de solución)

luego el intervalo:

$$P\left(\hat{p} - \lambda_{\alpha/2}\sqrt{\frac{p \cdot q}{n}} \leq p \leq \hat{p} + \lambda_{\alpha/2}\sqrt{\frac{p \cdot q}{n}}\right) = 1 - \alpha$$

para nivel de confianza 0,99 los valores de la tabla de la normal [0,1] serán:
(ir a script de la normal)

dado que no tenemos información de p, nos pondremos en el caso más desfavorable con varianza máxima **p=q=0,5**. Quedando el intervalo:

$$P\left(0{,}45 - 2{,}576\sqrt{\frac{0{,}5 \cdot 0{,}5}{400}} \leq p \leq 0{,}45 + 2{,}576\sqrt{\frac{0{,}5 \cdot 0{,}5}{400}}\right) = 0{,}99$$

luego :

$$p \in [0{,}3856; 0{,}5144]$$

con nivel de confianza de 0,99

---

5. Estimar el porcentaje de individuos que no lee ningún periódico al día en un pueblo de 1000 habitantes y con un nivel de significación del 1%. Para ello llevamos a cabo una muestra de tamaño 100 a personas distintas del pueblo, resultando que de éstas 80 no leen el periódico.

Se trata de un intervalo para la p de una característica; la proporción p de personas que no leen el periódico.

Conocemos que el nivel de significación es 0,01, el tamaño de la población es

N=1000,
el tamaño muestral es n=100 y la proporción muestral resultante es

.

Dado que conocemos el tamaño de la población y que la podemos considerar "finita" es conveniente que planteemos muestreo irrestricto, y apliquemos el factor corrector de poblaciones finitas; por lo que el intervalo quedaría:
(ir script de solución)

$$P\left(\hat{p} - \lambda_{\alpha/2}\sqrt{\frac{p \bullet q}{n}}\sqrt{\frac{N-n}{N-1}} \leq p \leq \hat{p} + \lambda_{\alpha/2}\sqrt{\frac{p \bullet q}{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$

los valores de serán según tabla (ir a script de la normal)

y , dado que no tenemos más información sobre p, tomamos el caso más desfavorable p=q=0,5

$$p\left(0{,}8 - 2{,}576\sqrt{\frac{0{,}5 \bullet 0{,}5}{100}}\sqrt{\frac{1000-100}{1000-1}} \leq p \leq 0{,}8 + 2{,}576\sqrt{\frac{0{,}5 \bullet 0{,}5}{100}}\sqrt{\frac{1000-100}{1000-1}}\right) = 0{,}99$$

resultando :

$$p \in [0{,}6777;0{,}9222]$$

con nivel de confianza del 99 %

---

6. En una empresa de 5000 trabajadores desea conocerse si ha variado mucho la valoración positiva de la gestión de la dirección, que el año pasado se concluyo feacientemente que era del 80% de los trabajadores. Para ello se realiza una muestra de tamaño 200 resultando que la valoración positiva era considerada por el 55% de los trabajadores encuestados. ¿Podemos afirmar que la valoración ha variado con probabilidad de equivocarnos del 1%?

El nivel de significación es la probabilidad de cometer un error al afirmar que el verdadero valor no se encuentra en el intervalo cuando realmente lo está, tendremos que construir, por tanto el intervalo para un nivel de significación del 0,01 y comprobar que en él se encuentra o no el valor de la proporción que se dio anteriormente.

Así el intervalo será :

$$P\left(\hat{p} - \lambda_{\alpha/2}\sqrt{\frac{p \bullet q}{n}}\sqrt{\frac{N-n}{N-1}} \leq p \leq \hat{p} + \lambda_{\alpha/2}\sqrt{\frac{p \bullet q}{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$

Conocemos que alfa es 0,01 , luego los valores de serán según tabla

el tamaño de la población es N=5000 ,(es conveniente utilizar muestreo irrestricto), el tamaño muestral es n=200 ,
la proporción muestral es , y dado que conocemos "fehacientemente" que p=0,8 es la que debemos usar; así.
(ir a script de la normal)

$$p\left(0,55-2,576\sqrt{\frac{0,8\bullet 0,2}{200}}\sqrt{\frac{5000-200}{5000-1}}\leq p\leq 0,55+2,576\sqrt{\frac{0,8\bullet 0,2}{200}}\sqrt{\frac{5000-200}{5000-1}}\right)=0,99$$

resultando:

$$p\in[0,47;0,62]$$

con nivel de confianza del 99 %,

por lo que afirmamos que la proporción ha cambiado pues el 0,8 (anterior) no se encuentra en el intervalo, y,

lo afirmamos con probabilidad de cometer un error del 1%

---

7. El ratio de productividad anual de nuestra empresa es una variable aleatoria de comportamiento desconocido si bien conocemos que su dispersión relativa es de 2 unidades de medida, desconociendo la media de dicho ratio. Dar un intervalo con confianza mínima del 90 %, para dicha media, si escogidos 40 días, resultó que la productividad media se situó en el valor 6.

Es un intervalo para la media con población desconocida, varianza poblacional conocida, tamaño muestral pequeño y dado que se nos habla de probabilidad "mínima" el intervalo responde a la acotación de Chebyshev, así:

$$P\left(\bar{x}-\frac{\sigma}{\sqrt{n\alpha}}\leq\mu\leq\bar{x}+\frac{\sigma}{\sqrt{n\alpha}}\right)\geq 1-\alpha$$

sustituyendo tendremos: (ir a script de solución)

$$P\left(6-\frac{2}{\sqrt{40\bullet 0,1}}\leq\mu\leq 6+\frac{2}{\sqrt{40\bullet 0,1}}\right)\geq 1-\alpha$$

siendo

luego el intervalo quedaría

$$\mu\in[5,5;6,5]$$

con, por lo menos, 0,9 de nivel de confianza

---

8. El número de errores diarios que se cometen al intentar conectar con una determinada red informática se distribuye normalmente con media desconocida. Para intentar conocer dicha media se realiza un M.A.S. de tamaño 10 días; resultando: 2,3,4,5,4,3,5 , -1.98, 1.98,1 errores.

Obtener un intervalo de confianza para la media de errores cometidos diariamente con un nivel de significación del 1%

El tamaño muestral es 10 luego pequeño, la población es normal, con varianza desconocida y muestra pequeña; la media muestral es 2,7 y la varianza muestral 4; con esta información el intervalo será:

$$P\left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n-1}}\right) = 1-\alpha$$

(ir a script de solución)

los valores de la tabla de la t de Student con n-1 grados de libertad, por tanto, 9 y para alfa 0,01 serán :(ir script de la t de student)

$$-t_{\alpha/2} = -3{,}250 \quad \text{y} \quad t_{\alpha/2} = 3{,}250$$

luego el intervalo quedará:

$$P\left(2{,}7 - 3{,}250\frac{2}{\sqrt{10-1}} \leq \mu \leq 2{,}7 + 3{,}250\frac{2}{\sqrt{10-1}}\right) = 0{,}99$$

donde

$$\mu \in [0{,}53;4{,}86]$$

con nivel de confianza el 99%

---

9. Para la estimación de la proporción de familias con ingresos superiores a 80000 Euros al año, se han realizado dos muestreos distintos, en ambos el tamaño muestral es el mismo, así como la forma de muestrear; en ambos, también, el nivel de confianza es idéntico (95,5%). En la ficha técnica del muestreo A se nos indica que p=q=0,5. En el muestreo B se nos indica que se utiliza como p la proporción de familias con ingresos superiores a 80000 euros que se obtuvo en un sondeo anterior. Nos preguntamos por: ¿Cuál de los dos muestreo nos dará un intervalo para dicha proporción de familias con menor amplitud? ¿Por qué? ¿Cuál de los dos muestreos es más riguroso?

El muestreo A trabaja con p=q=0,5 por lo que la varianza será máxima generando un intervalo con mayor amplitud, dado que la amplitud es mayor será un intervalo menos preciso y, por tanto, menos riguroso.

---

10. Para llevar a cabo un control de calidad sobre el peso que pueden resistir los 300 forjados (suelos) de una construcción, realizamos 12 pruebas resultando la resistencia media hasta la rotura de 350kg/cm2 con desviación típica de 20. Si trabajamos con nivel de confianza de 0,9.

a) ¿Ante que tipo de muestreo nos encontramos? ¿Por qué ?
b) ¿Entre que valores oscila la resistencia media de los 300 forjados, si por experiencias anteriores sabemos que dicha resistencia se distribuye normalmente?

a) Los ensayos son destructivos luego no es posible la "reposición" por lo que el muestreo es irrestricto y dado que la población es pequeña-finita (N=300) será de aplicación el factor corrector de poblaciones finitas. La población es pequeña-finita (N=300) será de aplicación el factor corrector de poblaciones finitas.

b) Se nos pide un intervalo para la media de la población, las características son: **población normal, varianza desconocida, muestras pequeñas, y muestreo irrestricto** luego el intervalo será;

$$P\left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n-1}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n-1}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$

(ir a script de solución)

para el nivel de significación de 0,1 y una t de Student de n-1=12-1=11 g.l. los valores de la tabla serán:

(ir script de la t de student)

quedando, por tanto el intervalo como:

$$P\left(350 - 1{,}796\frac{20}{\sqrt{12-1}}\sqrt{\frac{300-12}{300-1}} \leq \mu \leq 350 + 1{,}796\frac{20}{\sqrt{12-1}}\sqrt{\frac{300-12}{300-1}}\right) = 1-\alpha$$

Luego

$$\mu \in [339{,}37;360{,}62]$$

con nivel de confianza el 90%

---

11. Intentamos conocer el porcentaje con el que se da una determinada característica en una muy amplia población, para ello decidimos realizar un muestreo aleatorio simple. Cada encuesta (muestra) que realizamos tiene un coste de 1000 u.m y disponemos de 1000000 de u.m. . Si se pretende trabajar con un error del 8 % ¿Cuál será el nivel de confianza con el que trabajaremos, si

conocemos que dicha característica a estudiar es imposible que se de en más del 35% de la población?

Si el coste de cada encuesta es de 1000 u.m. y disponemos de 1.000.000 de u.m. está claro que el número de encuestas a realizar será de 1000000/1000=1000 luego tamaño muestral n=1000.

se trata de un intervalo para la proporción de una característica luego:

el error será :

el caso más desfavorable será aquí p=0,35 q=0,65.

Así :

este valor de la tabla corresponde a un (ir a script de la normal)

luego el nivel de confianza con el que trabajaremos será de 0,992

---

12. De que depende y en que sentido la amplitud de un intervalo de confianza.

Evidentemente existe un factor que influye en la amplitud de un intervalo pero que no es susceptible de ser modificado por el investigador, este factor es la varianza, a mayor varianza tendremos más variabilidad relativa por lo que el intervalo será mayor (más amplio).

El investigador puede modificar el tamaño muestral, a mayor tamaño muestral el intervalo se hace más preciso y por tanto menos amplio. Es lógico dado que a mayor información (muestra) más precisión en la estimación.

El investigador puede, también, modificar el nivel de confianza, a mayor nivel de confianza, mayor amplitud del intervalo y viceversa; lógico si pensamos que para "confiar" más en lo que hemos estimado hemos de ser necesariamente menos preciso luego el intervalo (amplitud) aumenta.

Si el muestreo es irrestricto, y es de aplicación el factor corrector de poblaciones finitas, dado que éste es un valor que se encuentra entre 0 y 1, hará que disminuya el error, haciendo que el intervalo sea menos amplio; lógico si pensamos que con muestreo irrestricto se tiene más información, luego hay más precisión.