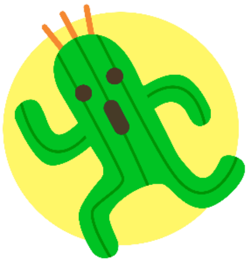
Estadística sin espinas

(5) Intervalos de confianza



v0.1 24_04_22

Aprende sin espinas
con @carlos_cactus

Sócrates se equivocaba. El conocimiento no es lo único que crece al compartirse: La alegría también.

A la inspiración del bucle_infinito,
al Cibergrupo y al tHash_A, por su amistad,
y sobre todo, a quienes dicen "pero quiero"
cuando sienten "no puedo".

¡Un saludo sin espinas!
@carlos_cactus :D

Y si quieres saber más:
¡Encuéntrame en Telegram como @carlos_cactus o habla con Espinito, el bot Sin Espinas, en @GestionSinEspinasBot.

Únete a la comunidad de Telegram Sin Espinas y no te pierdas nada!

Deja de preocuparte por aprobar y ¡Aprende sin Espinas!

I. INTERVALOS DE CONFIANZA

1.1. Conceptos previos

1.1.1. INFERENCIA

La inferencia estadística proporciona métodos para obtener conclusiones a partir de conjuntos de datos, fundamentadas en la teoría de la probabilidad, muchas veces, sin posibilidad de alcanzar una certeza absoluta.

1.1.1. ESTIMADOR

Un estimador es un **estadístico** que permite aproximarse al conocimiento de un parámetro de la población.

Po ejemplo, la media poblacional μ se puede aproximar mediante la media muestral $\bar{x}$ y la desviación típica σ se puede estima mediante la cuasivarianza muestral $\hat{s}^2$.

El estimador es el CENTRO del intervalo, en torno al cual se define el margen de error.

1.1.2. CONCEPTO DE INTERVALO DE CONFIANZA

Son constructos matemáticos que aprovechan la teoría de la probabilidad para estimar el nivel de confianza con que se infieren propiedades de la población mediante su extrapolación a partir de propiedades de una muestra estudiada.

Es decir, permiten conocer la población a partir de la exploración de la muestra.

Esa extrapolación conlleva un error asociado que se puede calcular y se puede expresar mediante intervalos de confianza.

Adoptan la forma:

$$I_{\underbrace{1-\alpha}_{\substack{NIVEL\ DE\\ CONFIANZA}}}\overbrace{(\mu)}^{\substack{Parámetro\\ estimado}} = \left(\overbrace{\underbrace{\bar{x}}_{\substack{Estimador\\ muestral}} - \underbrace{z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{\substack{Margen\\ de\ Error}}}^{\substack{extremo\\ inferior}}, \overbrace{\bar{x} + \underbrace{z_{\frac{\alpha}{2}}}_{\substack{valor\\ crítico}} \cdot \underbrace{\frac{\sigma}{\sqrt{n}}}_{\substack{Error\ estándard\\ de\ la\ media}}}^{\substack{extremo\\ superior}}\right)$$

El intervalo de confianza, por tanto:

- Está centrado en el estadístico estimador muestral del parámetro poblacional estimado.

- Los extremos del intervalo dependen de la muestra concreta a partir de la cual se construye. Es decir, muestras distintas de la misa población dan lugar a intervalos distintos para el mismo parámetro.

- **NO HAY FORMA DE CONOCER CON ABSOLUTA CERTEZA SI EL VALOR REAL ESTÁ DENTRO DEL INTERVALO CONSTRUIDO, sino que se garantiza un nivel de error máximo, es decir, un porcentaje de observaciones en las cuales no estará.**

- Evidencia la confianza de la estimación, es decir, DEL MÉTODO: el porcentaje de veces con que el intervalo contiene el valor real.

1.1.3. NIVEL DE CONFIANZA $1-\alpha$ Y NIVEL DE SIGNIFICACIÓN α

El nivel de confianza se denota como: $1-\alpha$

Donde $\boldsymbol{\alpha}$ **es el nivel de significación**, es decir, el error al cual se expone el intervalo.

Para un nivel de significación de $\alpha = 0.1$ el nivel de confianza que le corresponde es de 1 – 0.1 = 0.9, es decir, del 90%.

Para un nivel de significación de $\alpha = 0.05$ el nivel de confianza que le corresponde es de 1 – 0.05 = 0.95, es decir, del 95%.

Para un nivel de significación de $\alpha = 0.01$ el nivel de confianza que le corresponde es de 1 – 0.01 = 0.99, es decir, del 99%.

El nivel de confianza expresa el umbral de seguridad, o sea, la probabilidad con la cual se puede afirmar que el valor REAL del parámetro poblacional estimado está DENTRO del intervalo de confianza construido a partir de la observación de la muestra.

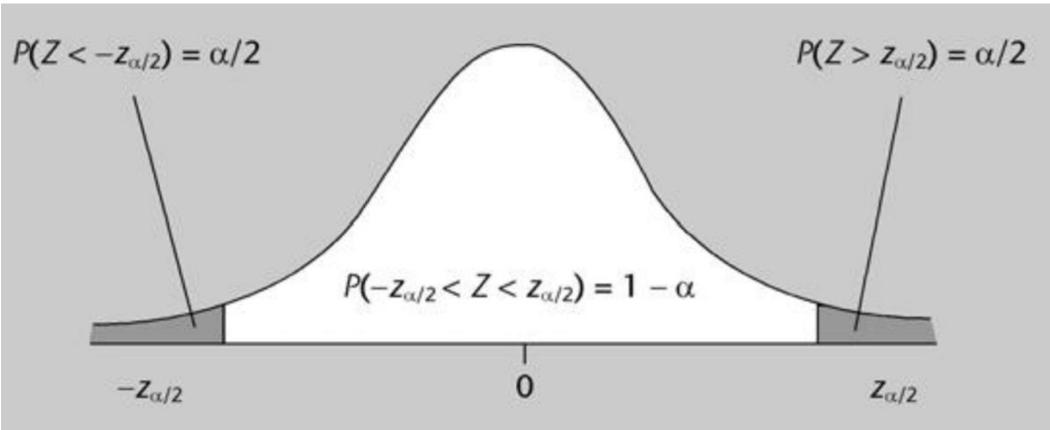
1.1.4. VALOR CRÍTICO

El valor crítico, denotado en el caso de la normal z por $z_{\frac{\alpha}{2}}$ es aquel valor de abscisa en una distribución de probabilidad que verifica:

$$P\left(Z > z_{\frac{\alpha}{2}}\right) = P\left(Z < -z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

Es decir, el valor crítico es aquel valor de abscisa que excluye un área, o sea, una probabilidad de $\frac{\alpha}{2}$ ya sea como cola hacia la derecha o hacia la izquierda.

Nótese que $\frac{\alpha}{2}$ es la PROBABILIDAD que cada cola excluye del intervalo.



La abscisa $z_{\frac{\alpha}{2}}$ delimita una región central cuya área es equivalente al nivel de confianza $1-\alpha$:

$$\acute{A}rea\ central = 1 - \left(\underbrace{P\left(Z < -z_{\frac{\alpha}{2}}\right)}_{\substack{Cola\\ izquierda}} + \underbrace{P\left(Z > z_{\frac{\alpha}{2}}\right)}_{\substack{Cola\\ derecha}}\right) = 1 - \left(\frac{\alpha}{2} + \frac{\alpha}{2}\right) = 1 - \alpha = P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right)$$

Cuanto mayor sea el nivel de confianza $1-\alpha$ más área quedará incluida en esa región central dentro del intervalo $(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$.

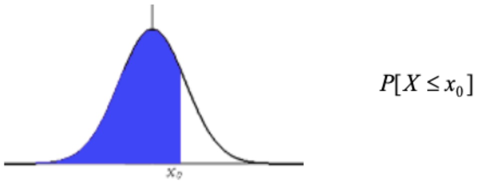
El cálculo del valor crítico se realiza mediante el uso de la tabla Z de Distribución Normal, o mediante la tabla T Student en su caso.

Los valores críticos más frecuentes para la distribución Z son:

Nivel de confianza $1-\alpha$	% confianza	Nivel de significación α	Probabilidad excluida por la cola $\alpha/2$	Valor crítico (estadístico z en la tabla Z) $z_{\alpha/2}$
0.9	90%	0.1	0.05	1.645
0.95	95%	0.05	0.025	1.96
0.99	99%	0.01	0.005	2.575

EJEMPLO

Se desea identificar el estadístico z que corresponde al nivel de confianza del 90%.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7704	0,7793	0,7823	0,7652
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8364	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8930
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9235	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9485	$\boxed{0,9495}$	$\boxed{0,9505}$	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	0,9767
2,0	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9934	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9865	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	$\boxed{0,9948}$	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9975	0,9975	0,9976	0,9977	0,9978	0,9978	0,9979	0,9980	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9990	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	$\boxed{0,9992}$	$\boxed{0,9993}$	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0.9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Se parte del nivel de confianza:

$$1 - \alpha = 0.9$$

Se aísla el nivel de significación α:

$$\alpha = 1 - 0.9 = 0.1$$

Se obtiene el valor de probabilidad $\frac{\alpha}{2}$ contenido en cada cola:

$$\frac{\alpha}{2} = \frac{0.1}{2} = 0.05$$

Y se calcula qué abscisa z corresponde a una probabilidad $\frac{\alpha}{2}$, es decir, $z_{\frac{\alpha}{2}}$, sabiendo que la cola a la derecha de $z_{\frac{\alpha}{2}}$ (así como la cola a la izquierda de $-z_{\frac{\alpha}{2}}$) contiene una probabilidad $\frac{\alpha}{2}$.

En este caso $\frac{\alpha}{2} = 0.05$. Entonces:

$$P\left(Z < -z_{\frac{\alpha}{2}}\right) = P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.05$$
$$\rightarrow P(Z < -z_{0.05})$$
$$= P(Z > z_{0.05})$$
$$= 0.05$$

Por complementariedad, se tiene:

$$\underbrace{P(Z < -z_{0.05})}_{\substack{Probabilidad \\ excluida\ por \\ la\ IZQUIERDA}} = \underbrace{P(Z > z_{0.05})}_{\substack{Probabilidad \\ excluida\ por \\ la\ DERECHA}} = 0.05 \rightarrow \underbrace{P(Z < z_{0.05})}_{\substack{Probabilidad \\ contenida \\ en\ intervalo}} = 1 - \underbrace{P(Z > z_{0.05})}_{0.05} = 0.95$$

De lo cual, ya se puede leer en la tabla qué FILA y qué COLUMNA corresponde a la entrada de probabilidad 0.95… **PERO NO APARECE.**
El estadístico $z_{0.05}$ se lee INDIRECTAMENTE DENTRO de la tabla.

Está entre 2 valores:

- El punto de abscisa 1,64 tiene 0.9495.
- El punto de abscisa 1,65 tiene 0.9505.

Exactamente entre ellos se encuentra z =1.645 que tiene asociada una probabilidad APROXIMADA de:

$$\frac{0.9495 + 0.9505}{2} = 0.95$$

Se observa que la abscisa $z_{\frac{\alpha}{2}} = 1.645$ excluye hacia la DERECHA un 5% de probabilidad.

Ahora bien, el intervalo de confianza es SIMÉTRICO (centrado en la media muestral en este caso) y se prolonga hacia AMBOS LADOS, excluyendo más allá de sus límites 2 colas (situadas a 1.645 veces la desviación estándar respecto la media en la distribución), una cola por la izquierda y otra cola por la derecha, cada una por valor de un 5% de probabilidad.

Si el estadístico $z_{\frac{\alpha}{2}} = 1.645$ EXCLUYE 2 colas de 5%, es que en $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$ hay CONTENIDA una región de 90%, es decir, el estadístico $z_{\frac{\alpha}{2}} = 1.645$ corresponde a un nivel de confianza del 90%.

NO CONFUNDIR CON 95%, que es la probabilidad que deja a su derecha el estadístico z = 1.645, es decir, la que excluye 1 única cola.

Análogamente, para el 95%:

$$1 - \alpha = 0.95 \to \alpha = 1 - 0.95 = 0.05 \to \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \to P\left(Z < -z_{\frac{\alpha}{2}}\right) = P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.025$$

Entonces:

$$P\left(Z > z_{\frac{0.05}{2}}\right) = 0.025 \to P\left(Z < z_{\frac{0.05}{2}}\right) = 1 - 0.025 = 0.975$$

En la tabla:

$$\underbrace{0.975}_{\substack{fila\ 1.9\\ col\ 0.06}} \to z_{\frac{0.05}{2}} = 1{,}96\ (cada\ cola\ exlcuye\ 2{,}5\%)$$

Para el 99% de confianza:

$$1 - \alpha = 0.99 \to \alpha = 1 - 0.99 = 0.01 \to \frac{\alpha}{2} = \frac{0.01}{2} = 0.005 \to P\left(Z < -z_{\frac{\alpha}{2}}\right) = P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.005$$

Es decir, cada cola contiene 0.5%.

Entonces:

$P\left(Z > z_{\frac{0.01}{2}}\right) = 0.005 \to P\left(Z < z_{\frac{0.01}{2}}\right) = 1 - 0.005 = \mathbf{0.995}$ **No está en tabla.**

Se promedia: $Entre\ \underbrace{0.9949}_{\substack{fila\ 2.5\\ col\ 0.07}}\ y\ \underbrace{0.9951}_{\substack{fila\ 2.5\\ col\ 0.08}} \to z_{\frac{0.005}{2}} = \underbrace{2.575}_{\substack{Entre\ col\ 0.07\\ y\ col\ 0.08}}$

1.1.5. ERROR ESTÁNDAR

El error estándar de la media es la desviación típica de la variable que modela la media muestral $\bar{X}$.

Las 2 expresiones que más se usan en este curso para calcularlo, según si se dispone de la desviación típica poblacional σ o no, son:

- Si SE CONOCE σ El error estándar de la media muestral es:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Si NO SE CONOCE σ El error estándar de la media muestral es:
-

$$s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}}$$

Donde $\hat{s}$ es la cuasivarianza muestral.

En general, el error estándar es una medida de la VARIABILIDAD del estimador (del parámetro estadístico) al cual hace referencia.

La expresión del error estándar depende de qué parámetro poblacional se analiza, de los estimadores muestrales empleados y de la distribución que siguen (así que hay mucha más variedad que estas 2 expresiones).

1.1.6. MARGEN DE ERROR

El margen de error ME define los extremos del intervalo de confianza que, bajo una distribución NORMAL (solo se trata este caso en este curso), están CENTRADOS en la media muestral.

El margen de error es el producto de un factor asociado al nivel de confianza, el estadístico denominado VALOR CRÍTICO por el error estándar de la media:

$$ME = valor\ crítico \cdot error\ estándar$$

Según si se dispone de la desviación típica poblacional σ, se calcula como:

- Si SE CONOCE σ:

$$ME = z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Donde $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$, luego $Z \sim N(0,1)$ y $P\left(Z < z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ como **cola a la izquierda**

- Si NO SE CONOCE σ:

$$ME = t_{\frac{\alpha}{2},n-1} \cdot s_{\bar{x}} = t_{\frac{\alpha}{2},n-1} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Donde la variable $\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ sigue una distribución t de Student con n − 1 grados de libertad denotada por t_{n-1} y $P\left(t_{n-1} > t_{\frac{\alpha}{2},n-1}\right) = \frac{\alpha}{2}$.

Nótese que que es una cola hacia la derecha, es decir, probabilidad más allá del valor crítico $t_{\frac{\alpha}{2},n-1}$.

1.1.7. AMPLITUD DE ERROR

La AMPLITUD DEL ERROR, denotada por A, es 2 veces la longitud del margen de error:

$$A = 2 \cdot ME$$

1.1.8. TAMAÑO DE LA MUESTRA Y MARGEN DE ERROR

1.1.8.1. Margen de error para la media muestral

Una aplicación estratégica del concepto de margen de error es averiguar qué tamaño de muestra mínima es necesario para conseguir un cierto nivel de confianza o un margen de error dado.
A partir de la expresión margen de error y un valor ME dado:

$$ME = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Se aísla el tamaño n de la muestra:

$$n \geq \left(z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\text{ME}}\right)^2$$

EJEMPLO

Se estudia la variable X que se puede considerar NORMAL. Se sabe que su desviación típica es $\sigma = 10$. Se mide la variable X en una muestra de tamaño n = 150 individuos y se obtiene una media muestral $\bar{x} = 120$.

Se desea conocer el tamaño n de la muestra que permite obtener un intervalo de confianza para la MEDIA de X al 95% de confianza, con un margen de error ME del 5%.

En primer lugar, se caracteriza la variable $\bar{X}$:

$$X \sim N(\mu, \sigma) \rightarrow \bar{X} \sim (\mu, \sigma_{\bar{x}})$$

Donde $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{150}}$. Es decir:

$$\bar{X} \sim \left(\mu, \frac{10}{\sqrt{150}}\right)$$

Por definición, la desviación típica de $\bar{X}$ es el ERROR ESTÁNDAR.

Entonces, para n = 150 y con un valor crítico $z_{\frac{\alpha}{2}} = 1.96$ (correspondiente al 95% de confianza), se tendría un ME:

$$ME = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow ME = 1.96 \cdot \frac{10}{\sqrt{150}}$$

Pero en esta ocasión, se desea identificar qué n corresponde a ME = 0.05. Es decir:

$$ME = 0.05 = 1.96 \cdot \frac{10}{\sqrt{n}}$$

Entonces, aislando n:

$$n \geq \left(1.96 \cdot \frac{10}{0.05}\right)^2 \rightarrow n \geq 153664$$

Para n ≥ 153664 observaciones, se alcanzaría un margen de error del 5% para un nivel de confianza del 95%.

1.1.8.2. Margen de error para la proporción

En el caso de la proporción, se observa:

$$ME = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Se aísla el tamaño n de la muestra:

$$n = \left(z_{\frac{\alpha}{2}}\right)^2 \cdot \frac{\hat{p} \cdot (1 - \hat{p})}{(ME)^2}$$

Obsérvese que a(1-a) es siempre menor o igual a 0.25 y que se maximiza para a = 0.5.

Entonces, si:

- Se impone un ME deseado.
- Se fija un valor crítico correspondiente al nivel de confianza deseado.

Entonces, con $\hat{p}$ = 0.5 (es decir, $\hat{p} \cdot (1 - \hat{p}) = \frac{1}{4}$, o sea, su valor máximo) se puede calcular el tamaño de muestra máximo para cada ME deseado según:

$$n \geq \frac{1}{4}\left(\frac{z_{\frac{\alpha}{2}}}{ME}\right)^2$$

***ME para no informada

1.2. Expresiones de intervalos de confianza

1.2.1. ESTRUCTURA DE UN INTERVALO DE CONFIANZA

La forma de un intervalo de confianza depende de qué distribución sigue la variable y del parámetro que se estima.

Se debe considerar:

Media poblacional: μ

Media de la muestra: $\bar{x}$

Varianza poblacional σ^2

Cuasivarianza muestral: $\hat{s}^2$

Tamaño de la muestra: n

Valor crítico z: $z_{\frac{\alpha}{2}}$

Valor crítico t: $t_{\frac{\alpha}{2},n-1}$

Margen de error: $ME = valor\ crítico \cdot error\ estándar$

Error estándar de la media $\sigma_{\bar{x}}$ $(\sigma\ conocida)$: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Error estándar de la media $s_{\bar{x}}$ $(\sigma\ desconocida)$: $s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}}$

Proporción real de la población: p

Proporción muestral observada: $\hat{p}$

¡CUIDADO! No confundir:

$s_{\bar{x}}$ = error estándar de la media con desviación poblacional desconocida.

$s_{\bar{d}}$ = CUASIDESVIACIÓN de la diferencia de medias de DATOS APAREJADOS (2 poblaciones DEPENDIENTES)

$e_{\bar{d}}$ = Error estándar de la diferencia de medias de DATOS APAREJADOS (2 poblaciones DEPENDIENTES)

$e_{\bar{x}_1-\bar{x}_2}$= Error estándar para la diferencia de medias de datos INDEPENDIENTES de 1 sola población.

Y las condiciones de aplicación de la Normal estándar o de la T de Student **PARA ESTE CURSO** son:

→ Normal estándar Z si se cumplen 2 requisitos:

- Asume normalidad de la población
- σ CONOCIDA
- **CON INDEPENDENCIA DEL TAMAÑO DE LA MUESTRA**

→ t-Student **SOLO** si se cumplen 3 requisitos:

- Asume normalidad de la población
- n < 30
- σ DESCONOCIDA

Lo anterior NO ES FORMALMENTE CORRECTO, pero quien corrija lo quiere así T_T

Conviene consensuar con el profesor en particular a qué supuestos acogerse, ya que dependiendo del tipo de estudio que se realice, las condiciones de aplicación y el cálculo del error estándar varían ligeramente.

Formalmente, se pueden encontrar 8 situaciones, en función de los datos poblacionales conocidos, el tipo de análisis practicado y las características concretas de las distribuciones que siguen las poblaciones.

Normalidad	n ≥ 30	σ conocida	Intervalo
Sí	Sí	Sí	$\bar{x} \mp z\left(\frac{\sigma}{\sqrt{n}}\right)$
Sí	Sí	No	$\bar{x} \mp z\left(\frac{\hat{s}}{\sqrt{n}}\right)$
Sí	No	Sí	$\bar{x} \mp t_{n-1}\left(\frac{\sigma}{\sqrt{n}}\right)$
Sí	No	No	$\bar{x} \mp t_{n-1}\left(\frac{\hat{s}}{\sqrt{n}}\right)$
No	Sí	Sí	$\bar{x} \mp z\left(\frac{\sigma}{\sqrt{n}}\right)(por\ TLC)$
No	Sí	No	$\bar{x} \mp z\left(\frac{\hat{s}}{\sqrt{n}}\right)(por\ TLC)$
No	No	Sí	**FUERA DE ESTE CURSO**
No	No	No	**FUERA DE ESTE CURSO**

Reemplazado en este curso por:
$\bar{x} \mp z\left(\frac{\sigma}{\sqrt{n}}\right)$

En síntesis, en este curso, asumiendo NORMALIDAD:

- **Si se conoce la desviación se sigue usando Z aunque n < 30.**

- **Solo se aplica t-Student si se desconoce la desviación poblacional y la muestra es pequeña n < 30.**

1.2.2. INTERVALO DE CONFIANZA PARA LA MEDIA M DE UNA VARIABLE NORMAL $N(\mu, \sigma)$ CON VARIANZA POBLACIONAL CONOCIDA

Nótese que se considera una población NORMAL.

Si el tamaño de la muestra es n > 30, se escribe como:

$$I_{1-\alpha}\ \overbrace{(\mu)}^{\substack{\text{Parámetro}\\ \text{estimado}}} = \left(\overbrace{\underbrace{\bar{x}}_{\substack{\text{Estimador}\\ \text{muestral}}} - \underbrace{z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{\substack{\text{Margen}\\ \text{de Error}}}}^{\substack{\text{extremo}\\ \text{superior}}},\ \overbrace{\bar{x} + z_{\frac{\alpha}{2}} \cdot \underbrace{\frac{\sigma}{\sqrt{n}}}_{\substack{\text{Error estándard}\\ \text{de la media}}}}^{\substack{\text{extremo}\\ \text{inferior}}}\right)$$

Como se observa, el intervalo está definido por 2 extremos. Cada extremo se expresa como el mejor estimador de que se dispone, la media muestral (centro del intervalo), en torno al cual se suma o se resta un término que representa el ERROR máximo cometido, es decir, el MARGEN de error, que es el producto del valor crítico por el error estándar de la media.

Formalmente, en los casos en que **n ≤ 30, se aplica la t de Student** (incluso con **con σ conocida**) ya que el comportamiento de muestras tan pequeñas no se ajusta tan bien a la normal Z como a una t. No obstante, **en este curso, si se conoce la desviación σ, SE APLICARÁ normal Z**.

1.2.3. INTERVALO DE CONFIANZA PARA LA MEDIA μ DE UNA VARIABLE NORMAL $N(\mu,\sigma)$ CON VARIANZA POBLACIONAL DESCONOCIDA

Nótese que se considera una población **NORMAL**.
Cuando se desconoce la varianza poblacional, se toma su estimador más fiable disponible, es decir, la CUASIVARIANZA MUESTRAL s.

Hay 2 situaciones, en función del tamaño de la muestra:

- Si n > 30 la distribución de la media sigue una Normal (estadístico z).
- Si n ≤ 30 la distribución de la media sigue una t-Student (estadístico t).

En ambos casos, ante el desconocimiento de σ, el error estándar se estima mediante s calculando $s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}}$. Es decir:

- Si las muestras tienen un tamaño SUPERIOR a 30 ($n > 30$), se aplica el estadístico Z y se estima el error estándar mediante s:

$$I_{1-\alpha}(\mu) = \left(\bar{x} - \underbrace{\overbrace{z_{\frac{\alpha}{2}}}^{Valor\ crítico} \cdot \frac{\hat{s}}{\sqrt{n}}}_{ME\,=\,v.c.\,\cdot\, s_{\bar{x}}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\overbrace{\hat{s}}^{\substack{Cuasidesviación \\ típica \\ MUESTRAL}}}{\sqrt{n}} \right)$$

A falta de la varianza poblacional σ^2, se usa la CUASIDESVIACIÓN TÍPICA MUESTRAL $\hat{s}$ para estimar el error estándar $s_{\bar{x}}$.

- Si las muestras tienen un tamaño INFERIOR a 30 ($n \leq 30$) se aplica T de Student y se estima el error estándar mediante s:

$$I_{1-\alpha}(\mu) = \left(\bar{x} - \underbrace{\overbrace{t_{n-1,\frac{\alpha}{2}}}^{Valor\ crítico} \cdot \frac{\hat{s}}{\sqrt{n}}}_{ME}, \bar{x} + \underbrace{t_{n-1,\frac{\alpha}{2}}}_{\substack{n-1\ df \\ grados \\ de\ libertad}} \cdot \frac{\hat{s}}{\sqrt{n}} \right)$$

Para muestras pequeñas, es decir, **con n ≤ 30, cuando se desconoce $\boldsymbol{\sigma}$**, se usa una distribución distinta de la Normal (0,1) para el cálculo del intervalo, se trata de la **t-Student.**

La t de Student tiene su propia tabla T que relaciona grados de libertad (filas) y nivel de significación α (columnas) asociado a la probabilidad.
Nótese que a diferencia de la tabla Z, la intersección de filas y columnas en la tabla T son estadísticos t, mientras que en la Z son valores de probabilidad.

Los grados de libertad ("degrees of freeedom") equivalen al tamaño de la muestra n menos 1:

$$g.l. = df = n - 1$$

La probabilidad tabulada **es UNA COLA A LA DERECHA** de la abscisa t_0, es decir, se representa $P(t_{n-1}>t_0)$ y no como en la tabla Z en que se lee $P(Z\leq z_0)$, que es la probabilidad acumulada como cola a la izquierda.

1.2.4. INTERVALO DE CONFIANZA PARA LA VARIANZA σ^2 DE UNA VARIABLE $N(\mu, \sigma)$

No se usa en este curso.

$$I_{1-\alpha}(\sigma^2) = \left(\frac{(n-1) \cdot \overbrace{s^2}^{\substack{varianza \\ muestral}}}{\underbrace{\chi^2_{\frac{\alpha}{2}, n-1}}_{Valor\ crítico}}, \frac{(n-1) \cdot \overbrace{s^2}^{\substack{varianza \\ muestral}}}{\underbrace{\chi^2_{1-\frac{\alpha}{2}, n-1}}_{\substack{Valor\ crítico \\ COMPLEMENTARIO}}} \right)$$

Se emplea la tabla de la Chi Cuadrado χ^2 para modelizar el intervalo. Nótese que los extremos del intervalo tienen estadísticos de contraste distintos:

$$\chi^2_{\frac{\alpha}{2}, n-1} \rightarrow Para\ el\ extremo\ inferior\ con\ \frac{\alpha}{2}$$
$$\chi^2_{1-\frac{\alpha}{2}, n-1} \rightarrow Para\ el\ extremo\ superior\ con\ 1-\frac{\alpha}{2}$$

1.2.5. INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS ENTRE 2 POBLACIONES $N(\mu_1, \sigma_1)$ Y $N(\mu_2, \sigma_2)$

$$I_{1-\alpha} \underbrace{(\mu_1 - \mu_2)}_{\substack{diferencia \\ de\ medias \\ poblaciones}} = \left(\underbrace{(\bar{x} - \bar{y})}_{\substack{diferencia \\ de\ medias \\ MUESTRALES}} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x} - \bar{y}) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

1.2.6. INTERVALO DE CONFIANZA PARA LA PROPORCIÓN POBLACIONAL EN $B(n,p)$

La probabilidad de éxito p se asimila aquí a la proporción poblacional. Se tiene:
p = Proporción poblacional A ESTIMAR.
$\hat{p}$ = Proporción muestral (el ESTIMADOR).
$\hat{P}$ = Variable que modela todos los valores que adopta la proporción muestral $\hat{p}$.

Se cumple, para muestras suficientemente grandes, que la variable $\hat{P}$ que modela la proporción muestral de la variable X cumple:

$$\frac{\hat{P}-p}{\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}}\sim Z(0,1)$$

Donde el error muestral de la proporción $s_{\hat{p}}$ se define como:

$$s_{\hat{p}}=\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$$

La estructura del intervalo de confianza para la proporción es:

$$I_{1-\alpha}\underbrace{(p)}_{\substack{proporción\\POBLACIONAL}}=\left(\underbrace{\hat{p}}_{\substack{proporción\\MUESTRAL}}-\underbrace{z_{\frac{\alpha}{2}}\cdot\underbrace{\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}}_{error\ estándar\ (s_{\hat{p}})}}\ ,\hat{p}+\underbrace{z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}}_{Margen\ de\ Error\ (ME)}\right)$$

Donde $s_{\hat{p}}$ es el error estándar de la proporción muestral.

1.2.7. INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE PROPORCIONES ENTRE $B(n_1,p_1)$ Y $B(n_2,p_2)$

$$I_{1-\alpha}(p_1-p_2)=\left((\hat{p}_1-\hat{p}_2)-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}_1\cdot(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2\cdot(1-\hat{p}_2)}{n_2}},(\hat{p}_1-\hat{p}_2)-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}_1\cdot(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2\cdot(1-\hat{p}_2)}{n_2}}\right)$$

Alternativamente, se puede escribir, para CUALQUIER intervalo de confianza:

$$I_{1-\alpha}(p_1-p_2)=\left((\hat{p}_1-\hat{p}_2)\mp z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}_1\cdot(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2\cdot(1-\hat{p}_2)}{n_2}}\right)$$

El símbolo $\mp$ indica que el extremo inferior está asociado a la RESTA y el superior a la SUMA.

EJEMPLO

Se tiene $X \sim N(\mu, 5)$ para los n = 16 valores de la muestra: 506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506,502, 509, 496. Se desea calcular:

a) Intervalo de confianza para la media poblacional μ con 90%, 95% y 99% de confianza.

Se tiene:

$$I_{1-\alpha}(\mu) = \left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

1. Se dispone de n = 16 y desviación típica poblacional σ = 5

$$I_{90\%}(\mu) = (\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{5}{\sqrt{16}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{5}{\sqrt{16}})$$

2. Se calcula la media muestral $\bar{x}$:

$$\bar{x} = \frac{1}{16} \cdot (506 + 508 + 499 + 503 + 504 + 510 + 497 + 512 + 514 + 505 + 493 + 496 + 506 + 502 + 509 + 496) = \frac{8060}{16} = 503.75$$

Se obtiene:

$$I_{90\%}(\mu) = \left(503.75 - z_{0.05} \cdot \frac{5}{\sqrt{16}}, \bar{x} + 503.75 \cdot \frac{5}{\sqrt{16}}\right)$$

3. Se calcula el valor crítico a partir del nivel de confianza y la tabla Z:

→ Para 90%

$$1 - \alpha = 0.9 \rightarrow \alpha = 0.1 \rightarrow \frac{\alpha}{2} = \frac{0.1}{2} = 0.05 \rightarrow P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.05 \rightarrow P\left(Z < z_{\frac{\alpha}{2}}\right) = 1 - 0.05 = 0.95 \rightarrow z_{\frac{\alpha}{2}} = 1.645$$

El valor crítico que corresponde a una probabilidad de 0.975 es 1.645 (fila 1.6 entre columnas 0.04 y 0.05, que alojan 0.9495 y 0.9505 respectivamente).

→ Para 95%

$$1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \rightarrow P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.025 \rightarrow P\left(Z < z_{\frac{\alpha}{2}}\right) = 1 - 0.025 = 0.975 \rightarrow z_{\frac{\alpha}{2}} = 1.96$$

El valor crítico que corresponde a una probabilidad de 0.95 es 1.96 (fila 1.9 columna 0.06 que aloja el valor de probabilidad buscado 0.975).

→ Para 99%

$$1-\alpha=0.99 \rightarrow \alpha=0.01 \rightarrow \frac{\alpha}{2}=\frac{0.01}{2}=0.005 \rightarrow P\left(Z>z_{\frac{\alpha}{2}}\right)=0.05$$

$$\rightarrow P\left(Z<z_{\frac{\alpha}{2}}\right)=1-0.005=0.995 \rightarrow z_{\frac{\alpha}{2}}=1.645$$

El valor crítico para una probabilidad de 0.995 es 1.645 (fila 2.5 entre columnas 0.07 y 0.08, que alojan 0.9949 y 0.9951 respectivamente).

4. Se escriben los 3 intervalo y se calculan sus extremos:

$$I_{90\%}(\mu)=\left(503.75-1.645\cdot\frac{5}{4}, 503.75+1.645\cdot\frac{5}{4}\right)=(501.69375,505.80625)$$
$$I_{95\%}(\mu)=\left(503.75-1.96\cdot\frac{5}{4}, 503.75+1.96\cdot\frac{5}{4}\right)=(501.3,506.2)$$
$$I_{99\%}(\mu)=\left(503.75-2.575\cdot\frac{5}{4}, 503.75+2.575\cdot\frac{5}{4}\right)=(500.53125,506.96875)$$

Nótese que incrementar el nivel de confianza exige ampliar los límites del intervalo, mientras que rebajar el nivel de confianza los estrecha.

b) Determinar el tamaño de la muestra que permite que con un 95% de confianza el intervalo tenga longitud 2.

El error es el término que se añade o se resta del estimador en torno al cual se centra el intervalo:

$$Error=z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$$

Se llama AMPLITUD al doble del error:

$$A=2\cdot z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$$

y se desea que la amplitud sea 2:

$$A=2\cdot z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}=2$$

Además, se sabe:

$$I_{95\%}(\mu)\rightarrow z_{\frac{\alpha}{2}}=1.96$$

$$\sigma=5$$

Entonces:

$$A=2\cdot 1.96\cdot\frac{5}{\sqrt{n}}=2$$

De lo cual:

$$n=(5\cdot 1.96)^2=96.04\rightarrow Por\ exceso: n=97$$

c) Calcular intervalos de confianza para la media poblacional μ con 90%, 95% y 99% de confianza asumiendo DESCONOCIDA la desviación típica poblacional.

Se tiene σ desconocida.
Se desea calcular $I_{90\%}$, $I_{95\%}$ y $I_{99\%}$.
Como se cumple:

1. Se desconoce σ la desviación típica poblacional
2. Se tiene $n \leq 30$

Se debe recurrir a la T-Student:

$$I_{1-\alpha}(\mu) = \left(\bar{x} - \underbrace{\overbrace{t_{n-1,\frac{\alpha}{2}}}^{Valor\ crítico} \cdot \frac{\hat{s}}{\sqrt{n}}}_{Error}, \bar{x} + t_{\frac{\alpha}{2}, \underbrace{n-1}_{\substack{\text{grados de} \\ \text{librtad} \\ \text{en tabla } T}}} \cdot \frac{\overbrace{\hat{s}}^{\substack{Cuasidesviación \\ típica\ MUESTRAL}}}{\sqrt{n}} \right)$$

1. En este caso:

$$I_{1-\alpha}(\mu) = \left(503.75 - t_{\frac{\alpha}{2},n-1} \cdot \frac{\hat{s}}{4}, 503.75 + t_{\frac{\alpha}{2},n-1} \cdot \frac{\hat{s}}{4}\right)$$

2. Se calcula la cuasidesviación típica muestral s aplicando:

$$\hat{s} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Se obtiene $\hat{s} = 6.2021$

3. Se calcula el valor crítico para cada intervalo a partir del nivel de confianza y la tabla T:

→ Los grados de libertad coinciden con el tamaño de la muestra menos 1:

$$g.l. = n - 1 \rightarrow 16 - 1 = 15$$

→ Para 90%

$$1 - \alpha = 0.9 \rightarrow \alpha = 0.1 \rightarrow \frac{\alpha}{2} = \frac{0.1}{2} = 0.05 \rightarrow t_{15,0.05} = \underbrace{1.7531}_{fila\ 15\ /\ col\ 0.05}$$

→ Para 95%

$$1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \rightarrow t_{15,0.025} = \underbrace{2.1315}_{fila\ 15\ /\ col\ 0.025}$$

→ Para 99%

$$1 - \alpha = 0.99 \rightarrow \alpha = 0.01 \rightarrow \frac{\alpha}{2} = \frac{0.01}{2} = 0.005 \rightarrow t_{15,0.005} = \underbrace{2.6025}_{fila\ 15\ /\ col\ 0.005}$$

4. Se escriben los 3 intervalo y se calculan los extremos:

$$I_{90\%}(\mu) = \left(503.75 - 1.7531 \cdot \frac{6.2021}{4}, 503.75 + 1.7531 \cdot \frac{6.2021}{4}\right)$$

$$I_{95\%}(\mu) = \left(503.75 - 2.1315 \cdot \frac{6.2021}{4}, 503.75 + 2.1315 \cdot \frac{6.2021}{4}\right)$$

$$I_{99\%}(\mu) = \left(503.75 - 2.6025 \cdot \frac{6.2021}{4}, 503.75 + 2.6025 \cdot \frac{6.2021}{4}\right)$$

EJEMPLO

De una muestra de 600 individuos, 240 cumplen cierta condición. Se pide calcular el intervalo de confianza del 95% para la proporción dada.

Para una proporción, se tiene:

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

En este caso:

$$n = 600$$
$$\hat{p} = \frac{240}{600} = 0.4$$

El valor crítico para el 95% es:

$$1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \rightarrow P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.025$$
$$\rightarrow P\left(Z < z_{\frac{\alpha}{2}}\right) = 1 - 0.025 = 0.975 \rightarrow z_{\frac{\alpha}{2}} = 1.96$$

Por tanto:

$$I_{0.95}(p) = \left(0.4 - 1.96 \cdot \sqrt{\frac{0.4 \cdot (1-0.4)}{600}}, 0.4 + 1.96 \cdot \sqrt{\frac{0.4 \cdot (1-0.4)}{600}}\right)$$

$$I_{0.95}(p) = (0.3608, 0.4392)$$

ENTREGABLE 4 – Cuestionario

PREGUNTA 1

Relacionad cada expresión con lo que corresponda:

Si queremos estudiar el % de estudiantes.
Usaremos un intervalo para la proporción.

Si quiero investigar la media y sabemos la desviación típica poblacional.
Usaremos la Normal.

Si quiero investigar la media y NO sabemos la desviación típica poblacional.
Usaremos la t-Student.

PREGUNTA 2

Sabemos que los gastos **trimestrales** de una familia en cine siguen una distribución normal con media 165 euros y desviación típica 30 euros. Por el Teorema del Limite Central, sabemos que los gastos **anuales** en cine de esta familia siguen una distribución también normal con media μ y desviación típica σ. Encontrad estos valores μ y σ.

Se tiene:
X = "gasto trimestral"
Que se distribuye según:

$$X \sim N(\mu, \sigma)$$

Por el Teorema del Limite Central se cumple:

$$X_1 \sim N(\mu_1, \sigma_1) + X_2 \sim N(\mu_2, \sigma_2) + \cdots + X_n \sim N(\mu_n, \sigma_n) = Y \sim N\left(\sum_{i=1}^{n} \mu_i \, , \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$$

En este caso, con media $\mu = 165$ y $\sigma = 30$ y asumiendo los 4 trimestres idénticos del año idénticos:

$$X_{trimestral} \sim N(165{,}30) \rightarrow X_{anual} \sim N\left(4 \cdot 165, \sqrt{4 \cdot 30^2}\right) = X_{anual} \sim N(660{,}60)$$

Se obtiene μ = 660 y σ = 60.

PREGUNTA 3
Una empresa fabrica CDs. Suponiendo normalidad y que la desviación típica poblacional es $\sigma = 36\ Mb$, ¿cuál tiene que ser el tamaño de la muestra mínima con un margen de error de 8 Mb, para tener un 95% de confianza?

Se tiene:

$$X \sim N(\mu, \sigma)$$

Se conoce la desviación típica poblacional $\sigma = 36$. Es decir, se conoce la varianza poblacional. Por tanto, la expresión del intervalo de confianza para la media es:

$$I_{1-\alpha} \overbrace{(\mu)}^{\substack{Parámetro \\ estimado}} = \left(\overbrace{\underbrace{\bar{x}}_{\substack{Estimador \\ muestral}} - \underbrace{z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{\substack{Margen \\ de\ error}}}^{\substack{extremo \\ superior}}, \overbrace{\bar{x} + \underbrace{z_{\frac{\alpha}{2}}}_{Valor\ crítico} \cdot \underbrace{\frac{\sigma}{\sqrt{n}}}_{\substack{Error\ estándard \\ de\ la\ media}}}^{\substack{extremo \\ inferior}} \right)$$

Donde el margen de error es:

$$Margen\ de\ error = valor\ crítico \cdot e_{\bar{x}} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 8\ Mb$$

Para el 95% de confianza, el valor crítico es 1.96 ya que:

$$1 - \alpha = 0.95 \rightarrow \alpha = 1 - 0.95 = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \rightarrow P\left(Z < -z_{\frac{\alpha}{2}}\right)$$
$$= P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.025$$

Entonces: $P\left(Z > z_{\frac{0.05}{2}}\right) = 0.025 \rightarrow P\left(Z < z_{\frac{0.05}{2}}\right) = 1 - 0.025 = 0.975$

En la tabla Z: $\underbrace{0.975}_{fila\ 1{,}9\ columna\ 0.06} \rightarrow z_{\frac{0.05}{2}} = 1{,}96\ (cada\ cola\ exlcuye\ 2{,}5\%)$

Se desea un margen de error máximo de ME = 8 Mb y se conoce $\sigma = 36$:

$$\mathrm{ME} = z_{\frac{0.05}{2}} \cdot \mathrm{e}_{\bar{\mathrm{x}}} = 1{,}96 \cdot \frac{36}{\sqrt{n}} = 8$$

Aislando n:

$$\left(1{,}96 \cdot \frac{36}{8}\right)^2 = 77.79 = n \underbrace{\rightarrow}_{\substack{Se\ aproxima\ por\ exceso: \\ n\ debe\ ser\ natural}} n\ mínimo = 78\ unidades$$

PREGUNTA 4

Estamos interesados en medir el tiempo medio de respuesta de un servidor a las peticiones de los usuarios. Para hacer esto medimos en horarios diferentes el tiempo, en segundos, que se tarda en establecer una conexión con nuestro servidor. Los datos obtenidos fueron: 1, 1, 2, 2, 2, 1, 0, 0.

Suponiendo que el tiempo de respuesta sigue una distribución normal, calculad el intervalo de confianza para la media al nivel de confianza 95%.

Para encontrar el intervalo de confianza, en la forma (a,b), calcularemos lo siguiente:

a) Error estándar de la media $s_{\bar{x}}$
b) Valor crítico positivo
c) Extremo inferior del intervalo
d) Extremo superior del intervalo

Se tiene que:

1. La VARIANZA POBLACIONAL ES DESCONOCIDA
2. La muestra tiene $n \leq 30$ datos.

Por tanto, se recurre a la t-Student.

Se tiene el intervalo de confianza de la forma:

$$I_{1-\alpha}(\mu) = \left(\underbrace{\bar{x} - \overbrace{t_{n-1,\frac{\alpha}{2}},}^{Valor\ crítico} \cdot \frac{\hat{s}}{\sqrt{n}}}_{Margen\ de\ error}, \bar{x} + t_{\underbrace{n-1}_{\substack{grados\\ de\ libertad\\ en\ la\ tabla\ T}},\frac{\alpha}{2}} \cdot \frac{\overbrace{\hat{s}}^{\substack{Cuasidesviación\\ típica\ MUESTRAL}}}{\sqrt{n}} \right)$$

Aquí se usa la **CUASI** desviación típica MUESTRAL $\hat{s}$ ya que se ajusta mejor a la T-Student, que difiere de la desviación típica poblacional en el tamaño de muestra: n – 1 en lugar de n.

Ahora:

1. Se calcula la media:

$$\bar{x} = \frac{0 \cdot 2 + 1 \cdot 3 + 2 \cdot 3}{8} = \frac{9}{8} = 1.125$$

2. Se calcula la **CUASI** desviación típica MUESTRAL $\hat{s}$ (con n – 1):

$$\hat{s} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$\hat{s} = \sqrt{\frac{2 \cdot (0 - 1.125)^2 + 3 \cdot (1 - 1.125)^2 + 3 \cdot (2 - 1.125)^2}{8-1}} = 0.8345$$

3. Ahora ya se puede calcular el error estándar $s_{\bar{x}}$:

$$s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}} = \frac{0.8345}{\sqrt{8}} = 0.295$$

4. Se calcula el valor crítico a partir de la tabla T-Student, el nivel de confianza 0.95 y n – 1 = 7 grados de libertad:

$$1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow t_{0.025,8-1} = 2.3646\ (fila\ 7\ columna\ 0.025)$$

5. Se calcula el margen de error:

$$ME = t_{\frac{\alpha}{2},n-1} \cdot s_{\bar{x}} = 2.3646 \cdot 0.295 = 0.6976$$

6. Se calculan los extremos del intervalo aplicando el margen de error:

$$a = \bar{x} - ME = 1.125 - 0.6976 = 0.4274$$
$$b = \bar{x} + ME = 1.125 + 0.6976 = 1.8226$$

Es decir, el intervalo de confianza es:

$$I_{0.95}(\mu) = \left(1.125 \mp \underbrace{\overbrace{2.3646}^{\substack{Valor\ crítico \\ 8-1\ g.l. \\ \frac{\alpha}{2}=0.025}} \cdot \frac{\overbrace{0.8345}^{\substack{CUASI\ desviación \\ típica\ MUESTRAL}}}{\sqrt{8}}}_{Margen\ de\ error}, \right)$$

PREGUNTA 5

Una empresa de ordenadores quiere saber la capacidad real de los discos duros de sus portátiles con 500Gb. Con este propósito cogió 10 al azar. Los datos obtenidos, en Gb, fueron los siguientes: 489, 457, 487, 466, 459, 472, 450, 472, 463, 454.

Suponiendo normalidad y que la desviación típica de la población es $\sigma = 29$, calculad el intervalo de confianza para la media de la capacidad de los discos duros, al nivel de confianza 95%. Se desea calcular:

a) Error estándar de la media $e_{\bar{x}}$
b) Valor critico positivo
c) Extremo inferior del intervalo
d) Extremo superior del intervalo

Se tiene que la VARIANZA POBLACIONAL ES CONOCIDA ($\sigma^2 = 29^2$). Por tanto, se recurre a la tabla Z Normal.

Se tiene el intervalo de confianza de la forma:

$$I_{1-\alpha} \overbrace{(\mu)}^{\substack{Parámetro \\ estimado}} = \left(\overbrace{\underbrace{\bar{x}}_{\substack{Estimador \\ muestral}} - \underbrace{z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{\substack{Margen \\ de\ error}}}^{\substack{extremo \\ superior}}, \overbrace{\bar{x} + \underbrace{z_{\frac{\alpha}{2}}}_{Valor\ crítico} \cdot \underbrace{\frac{\sigma}{\sqrt{n}}}_{\substack{Error\ estándard \\ de\ la\ media}}}^{\substack{extremo \\ inferior}} \right)$$

1. Se calcula la media:

$$\bar{x} = \frac{489 + 457 + 487 + 466 + 459 + 472 + 450 + 472 + 463 + 454}{10} = 466.9$$

2. Se calcula el error estándar $\sigma_{\bar{x}}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{10}} = 9.1706$$

3. Se calcula el valor crítico $z_{\frac{\alpha}{2}}$ a partir de la tabla Z y el nivel de confianza de 95%:

$$1 - \alpha = 0.95 \rightarrow \alpha = 1 - 0.95 = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \rightarrow P\left(Z < -z_{\frac{\alpha}{2}}\right)$$
$$= P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.025$$

Entonces: $P\left(Z > z_{\frac{0.05}{2}}\right) = 0.025 \rightarrow P\left(Z < z_{\frac{0.05}{2}}\right) = 1 - 0.025 = 0.975$

En la tabla Z: $\underbrace{0.975}_{fila\ 1{,}9\ columna\ 0.06} \rightarrow z_{\frac{0.05}{2}} = 1{,}96\ (cada\ cola\ exlcuye\ 2{,}5\%)$

4. Se calcula el margen de error:

$$ME = z_{\frac{\alpha}{2}} \cdot e_{\bar{x}} = 1.96 \cdot 9.1706 = 17.9743$$

5. Se calculan los extremos del intervalo aplicando el margen de error:

$$a = \bar{x} - ME = 466.9 - 17.974 = 448.9256$$
$$b = \bar{x} + ME = 466.9 + 17.974 = 484.8743$$

Es decir, el intervalo de confianza es:

$$I_{0.95}(\mu) = \left(466.9 \mp 1.96 \cdot \frac{29}{\sqrt{10}}\right)$$

FORMALMENTE, EN ESTE CASO SE DEBERÍA HABER APLICADO t EN LUGAR DE z, POR AJUSTARSE t MEJOR QUE z A MUESTRAS CON n < 30. PERO EN ESTE CURSO SE RESTRINGE EL USO DE t A AQUELLOS CASOS QUE CUMPLEN 2 REQUISITOS:

- **n < 30**
- **σ DESCONOCIDA.**

EN ESTE CASO, σ SE CONOCE → SE APLICA z SIN ATENDER AL TAMAÑO DE n.

PREGUNTA 6
El tiempo de espera, desde que arranca un ordenador hasta poder trabajar con él, es una variable aleatoria distribuida **uniformemente** entre los 1 y los 3 segundos. Se escoge una muestra de 50 tiempos de espera.

a) ¿Con qué probabilidad **LA MEDIA del tiempo** de espera de esta muestra será superior a 2.05 segundos?

b) Se pide calcular los valores de los parámetros $(\mu, \sigma_{\bar{x}})$ de la distribución aproximada que sigue esta media muestral.

Se desea conocer una probabilidad de su MEDIA, del tiempo PROMEDIO, es decir, sobre $\bar{X}$ y no propiamente sobre X.

Se tiene: X= "Tiempo de espera desde el arranque"

Que sigue: $X{\sim}U(a,b) \rightarrow X{\sim}U(1,3)$

1. La esperanza de la variable es:

$$E(x) = \frac{a+b}{2} \rightarrow E(x) = \frac{1+3}{2} = \frac{4}{2} = 2$$

2. La desviación típica σ se calcula a partir de la varianza:

La varianza de la variable es:

$$Var(x) = \frac{(b-a)^2}{12} \rightarrow Var(x) = \frac{(3-1)^2}{12} = \frac{4}{12} = \frac{1}{3} = \sigma^2$$

Por tanto:

$$\sigma = \sqrt{\frac{1}{3}} = \frac{\sqrt{3}}{3}$$

3. Para el cálculo de la probabilidad deseada, se aproxima la variable X a una distribución Normal según:

$$X{\sim}U(a,b) \approx \bar{X}{\sim}N(\mu,\sigma)$$

donde μ es la esperanza E(X) y $\sigma = \sqrt{Var(X)}$. Una vez aproximada la variable X que sigue una distribución uniforme a una disitrbución normal, se cumple que la distribución de SU MEDIA $\bar{X}$ verifica:

$$\bar{X}{\sim}N(\mu,\sigma) \rightarrow \bar{X}{\sim}N\left(\mu, \underbrace{\frac{\sigma}{\sqrt{n}}}_{e_{\bar{x}}}\right)$$

En este caso, con $\mu = 2, \sigma = \frac{\sqrt{3}}{3} \, y \, n = 50$:

$$\bar{X} \sim N\left(2, \frac{\frac{\sqrt{3}}{3}}{\sqrt{50}}\right) = \bar{X} \sim N\left(\underbrace{2}_{\mu}, \underbrace{\frac{\sqrt{6}}{30}}_{e_{\bar{x}}}\right)$$

Entonces, la probabilidad para EL TIEMPO DE ESPERA **PROMEDIO** es:

$$P(\bar{X} > x_0) \underbrace{=}_{Tipificando} P\left(\overbrace{\frac{\bar{X} - \mu}{\sigma}}^{Z\ estándar} > \frac{x_0 - \mu}{\sigma}\right) = P\left(Z > \frac{x_0 - \mu}{\sigma}\right)$$

En este caso, $x_0 = 2.05$:

$$P(\bar{X} > 2.05) = P\left(Z > \frac{2.05 - 2}{\frac{\sqrt{6}}{30}}\right) = \underbrace{P(Z > 0.6123)}_{\substack{Cola\ a\ la\ derecha \\ No\ está\ en\ tabla\ Z}} = 1 - \underbrace{P(Z < 0.6123)}_{\substack{Cola\ hacia\ la\ izquierda \\ Sí\ está\ en\ tabla\ Z}}$$

De lo cual:

$$P(\bar{X} > 2.05) = 1 - \underbrace{P(Z < 0.6123)}_{\substack{Cola\ hacia\ la\ izquierda \\ Sí\ está\ en\ tabla\ Z}} = 1 - \underbrace{0.7291}_{\substack{fila\ 0.6 \\ Columna\ 0.01}} = 0.2709$$

PREGUNTA 7

De los 138 estudiantes de la UOC elegidos aleatoriamente, 26 estudian el Grado de informática. Determinad un intervalo de confianza para la proporción de alumnos de la UOC que estudian el Grado de informática, al nivel de confianza 90%. Para encontrar el intervalo de confianza, en la forma (a,b), calcularemos lo siguiente:

a) Error estándar de la proporción $\sigma_{\hat{p}}$.
b) Valor crítico positivo.
c) a = extremo inferior del intervalo
d) b = extremo superior del intervalo
e) ¿Se puede decir que la mitad de los estudiantes de la UOC, estudian el grado de informática? h = 1 si SÍ se puede, h = 0 si NO SE PUEDE.

Se pide un intervalo de confianza para una proporción poblacional p desconocida a partir de una proporción muestral $\hat{p}$ conocida, cuya forma es:

$$I_{1-\alpha}(\underbrace{p}_{\substack{\textit{proporción}\\ \textit{poblacinal}}}) = \left(\overbrace{\hat{p}}^{\substack{\textit{proporción}\\ \textit{muestral}}} - \underbrace{z_{\frac{\alpha}{2}}}_{\substack{\textit{valor}\\ \textit{crítico}}} \cdot \overbrace{\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}}^{\textit{Error estándar}}, \hat{p} + \overbrace{z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}}^{\textit{Margen de error}}\right)$$

1. Se calcula la proporción $\hat{p}$ mediante la regla de LaPlace:

$$\hat{p} = \frac{26}{138}$$

2. Se calcula el valor crítico a partir de la tabla z y el nivel de confianza dado del 90%:

$$1-\alpha = 0.9 \rightarrow \alpha = 0.1 \rightarrow \frac{\alpha}{2} = \frac{0.1}{2} = 0.05 \rightarrow P\left(Z > z_{\frac{\alpha}{2}}\right) = 0.05 \rightarrow P\left(Z < z_{\frac{\alpha}{2}}\right)$$
$$= 1 - 0.05 = 0.95 \rightarrow z_{\frac{\alpha}{2}} = 1.645$$

3. Se calcula el error estándar:

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{\frac{26}{138}\cdot\left(1-\frac{26}{138}\right)}{138}} = 0.03328$$

4. Se escribe el intervalo y se calculan los extremos

$$I_{90}(p) = \left(\frac{26}{138} - 1.645 \cdot 0.03328, \frac{26}{138} + 1.645 \cdot 0.03328\right) = (0.1336{,}0.2431)$$

5. Se pretende refutar o confirmar que la proporción dada del 50% de estudiantes está dentro del intervalo calculado (0.1336,0.2431). Se observa $0.5 \notin (0.1336{,}0.2431)$, entonces NO se puede afirmar que la mitad de estudiantes cursan Informática (h = 0).

ENTREGABLE 4 – Práctica de R

Importación de los datos

Se recurre a la instrucción:

```
datos<-read.table("C:/Users/Tete/Downloads/AP.csv.csv", header=TRUE,
sep=";",na.strings="NA",
fileEncoding = "UTF-8", quote = "\"", colClasses=c(rep("character",4),rep("numeric",2),
rep("character",2)))
```

Para verificar la correcta importación de los datos, se ejecuta:

```
head(datos,3)
```

Que devuelve:

```
Cod     Country Citycode              City
1 AFG Afghanistan    40003             Herat
2 AFG Afghanistan    40001             KABUL
3 AFG Afghanistan    40002 Kandahar (Quandahar)
  Population2000 PM10Concentration1999     Region
1        323741                 46 South Asia
2       2457496                 46 South Asia
3        411752                 51 South Asia
  IncomeGroup
1  Low income
2  Low income
3  Low income
```

PREGUNTA 1

Hallar un intervalo de confianza para la concentración de partículas PM10Concentration1999 con un nivel de confianza del 90% para las ciudades de España.

Se recurre a la instrucción t.test():

```
t.test(datos$PM10Concentration1999[datos$Country==“Spain”],conf.level=0.9)
```

Esta instrucción permite calcular el intervalo de confianza de la muestra de la variable seleccionada (PM10Concentration1999) y filtrada para la región Spain según la distribución T-Student con un nivel de confianza del 90%.

Se obtiene:

```
One sample t-test
data: datos$PM10Concentration1999[datos$Country == “Spain”]
t = 33.529, df = 54, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
   38.33163  42.35927
sample estimates:
mean of x    40.34545
```

PREGUNTA 2

Hallar un intervalo de confianza para la concentración de partículas PM10Concentration1999 con un nivel de confianza del 95% según los ingresos del país al que pertenecen las ciudades, es decir, para cada grupo de ciudades que pertenecen a un determinado nivel de ingresos, tenéis que calcular un intervalo de confianza para la concentración de partículas PM10Concentration1999 con un nivel de confianza del 95%. Comentad los resultados. ¿Creéis que los ingresos del país donde pertenecen las ciudades influye en la concentración de partículas PM10Concentration1999?

Para calcular el intervalo de confianza de la muestra para la variable PM10Concentration1999 según el grupo de ingresos, se recurre a la instrucción tapply():

tapply(datos$PM10Concentration1999,datos$IncomeGroup,conf.level=0.95,t.test)

Esta instrucción permite realizar un test de contraste tipo t.test (último parámetro de la instrucción) con un nivel de confianza del 95% de forma **INDEPENDIENTE** para la variable seleccionada (PM10Concentration1999) **EN CADA UNO de los grupos de ingreso** que se forman en la muestra según la variable categórica IncomeGroup. Se obtiene, para cada valor de la variable categórica. Se obtiene:

```
$`High income`
      One sample t-test
data: X[[i]]
t = 63.708, df = 1094, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to
0
95 percent confidence interval:
  28.31483  30.11439
sample estimates:
mean of x
  29.21461
```

```
$`Low income`
      One sample t-test
data: X[[i]]
t = 17.179, df = 98, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to
0
95 percent confidence interval:
  59.50163  75.04382
sample estimates:
mean of x
  67.27273
```

```
$`Lower middle income`
      One sample t-test
data: X[[i]]
t = 46.107, df = 775, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to
0
95 percent confidence interval:
  77.12829  83.98812
sample estimates:
mean of x
  80.5582
```

```
$`Upper middle income`
      One sample t-test
data: X[[i]]
t = 56.217, df = 1267, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to
0
95 percent confidence interval:
  49.38995  52.96178
sample estimates:
mean of x
  51.17587
```

Se observan grandes diferencias entre los valores medios de cada categoría y entre los máximos y mínimos entre categoría (desde 28.315 en el grupo de Ingreso Alto hasta 83.988 en el grupo de Ingreso Medio Bajo):

	Media	Rango
○ Bajo	67.27273	[59.50163,75.04382]
○ Medio Bajo	80.5582	[77.12829,**83.98812**]
○ Medio Alto	51.17587	[49.38995,52.96178]
○ Alto	29.21461	[**28.31483**,30.11439]

Se puede concluir que hay una correlación negativa entre el número de partes por millón del contaminante y el nivel de ingreso del país.

Nótese una aparente paradoja: los máximos de la distribución no se observan en los países del grupo de ingreso Bajo, sino en el de Medio Bajo. Esto tiene una explicación sencilla. Existe un umbral de mínimo de nivel de ingreso que permite el acceso generalizado a los vehículos de motor, principales fuentes de contaminantes, que no se alcanza en estos países, y lo mismo puede argumentarse para el grado de industrialización general del país.

PREGUNTA 3

Hallar un intervalo de confianza para la proporción de ciudades que están a Europa o en Asia Central al 99% de confianza. Indicación: Tenéis que calcular primero la tabla de frecuencias de la variable Region de cara a calcular la proporción de ciudades que están en Europa o en Asia Central.

Este cálculo no se puede realizar a partir de la representación de datos dada.

Se desea calcular el intervalo de confianza:

$$I_{1-\alpha}(p) = \left[\hat{p} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}, \hat{p} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right]$$

Donde:

- $\hat{p}$ es la proporción de la cual se desea calcular el intervalo de confianza.
- α es el nivel de significación del intervalo.
- $Z_{\frac{\alpha}{2}}$ es la abscisa de la distribución de probabilidad que deja a la izquierda $\frac{\alpha}{2}$.
- n es el tamaño de la muestra.

Por tanto, para calcular el intervalo se debe calcular:

1. La proporción $\hat{p}$.
2. El nivel de significación α.
3. La abscisa $Z_{\frac{\alpha}{2}}$

En primer lugar, la proporción $\hat{p}$ es:

$$\hat{p} = \frac{Ciudades\ de\ Europa\ Central\ \&\ Asia}{Total\ de\ observaciones}$$

Para calcularla, se obtiene la tabla de frecuencias por categoría Region para la variable PM10Concentration1999. Se recurre a la instrucción table():

table(datos$Region)

Se obtiene:

```
East Asia & Pacific          Europe & Central Asia
839                          871
Latin America & Caribbean Middle East & North Africa
467                          191
North America                     South Asia
```

256 402
Sub-Saharian Africa
192

Después, se almacena en una nueva variable (p) la proporción para la cual se desea calcular el intervalo de confianza, es decir, el cociente entre las observaciones seleccionadas de la Región Europa Central y Asia (s) con la muestra total (n), es decir, Se escribe:

s<-length(datos$Region[datos$Region=="Europe & Central Asia"])
n<-length(datos$Region)
p<-s/n

En segundo lugar, para obtener el valor de significación α basta con considerar el 99% de confianza deseado para el intervalo. Es decir:

$$1-\alpha = 0.99 \rightarrow \alpha = 1-0.99 = 0.01$$

En tercer lugar, se debe obtener el valor $Z_{\frac{\alpha}{2}}$. Se recurre a la instrucción qnorm($1-\frac{\alpha}{2}$), que devuelve la abscisa de una distribución normal asociada a una probabilidad dada. En este caso, se desea el 99% de confianza, es decir:

$$\alpha = 0.01 \rightarrow \frac{\alpha}{2} = 0.005$$

Por tanto, se busca $Z_{0.05}$. La probabilidad que se desea excluir por la izquierda es de 0.005. Se escribe:

z<-qnorm(0.995)

Se obtiene:

[1] 2.575829

Ahora ya se puede calcular el intervalo deseado. Se escribe:

*p-z*sqrt(p*(1-p)/n);p+z*sqrt(p*(1-p)/n)*

Se obtienen los 2 extremos:

[1] 0.2504905
[1] 0.2908395

De modo que se puede escribir:

$$I_{0.99}(p) = \left[\frac{871}{3218} - Z_{\frac{0.005}{2}} \cdot \sqrt{\frac{\frac{871}{3218}\cdot\left(1-\frac{871}{3218}\right)}{3218}}, \frac{871}{3218} + Z_{\frac{0.005}{2}} \cdot \sqrt{\frac{\frac{871}{3218}\cdot\left(1-\frac{871}{3218}\right)}{3218}}\right]$$

$$I_{0.99}(p) = [0.2504905, 0.2908395]$$