Inferencia Estadística

Dep. de Matemáticas

IES PEDRO CERRADA

Finalidad:

Obtener conclusiones válidas para toda la población a partir del estudio de una muestra.

Ejemplo: He preguntado la nota de matemáticas a 3 alumnos y la media de las notas es 6,4. ¿Podemos extraer alguna conclusión sobre la nota media de la clase?¿Con qué grado de confianza?

¿Cómo?:

Mediante los m**étodos de estimación puntual** y de **intervalos de confianza**

Repaso del cálculo de probabilidades de una Normal

La mayoría de los resultados que aparecen en la estimación por intervalos están relacionados con **distribuciones Normales**

$$X \sim \mathcal{N}(\mu,\, \sigma)$$

Por esta razón es conveniente hacer un repaso de su manejo

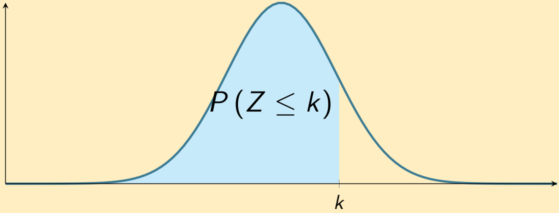
Para calcular la probabilidad **se utiliza una tabla** que ya tiene calculadas probabilidades de la $Z \sim \mathcal{N}(0, 1)$ ...

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,999
3,1	0,99903	0,99906	0,9991	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,9994	0,99942	0,99944	0,99946	0,99948	0,9995
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,9996	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,9997	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,9998	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,9999	0,9999	0,9999	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Cálculo práctico de la probabilidad de la Normal:

... y que refleja:

$$P\left(Z \leq k\right), \quad k \in \left[0, 4'09\right]$$



z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,999
3,1	0,99903	0,99906	0,9991	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,9994	0,99942	0,99944	0,99946	0,99948	0,9995
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,9996	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,9997	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,9998	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,9999	0,9999	0,9999	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Cálculo en la $Z \sim \mathcal{N}(0,\,1)$ o normal estándar

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,5
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,6
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,6
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,6
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,7
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,7
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,7
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,8
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,8
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,8
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,8
1,2	0,88493	0,88686	0,8887	0,89065	0,89251	0,89435	0,89617	0,8
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,9
1,4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

- $P\,(Z \leq 0) = 0{,}5$. Ya que en este caso $k = 0 = 0{,}00$ la suma del valor de la fila 0 con el valor de la columna 0 me da el valor de k, y la probabilidad asociada es 0,5
- $P\,(Z \leq 1{,}23) = 0{,}89065$. En este caso la probabilidad asociada a 1.23 se busca en la fila 1.2 y la columna 0.03

Cálculo en la $Z \sim \mathcal{N}(0, 1)$ o normal estándar

Las probabilidades en conjuntos de valores de la distribución que no se puedan obtener directamente de la tabla se transformarán en operaciones con probabilidades que sí estén en la tabla:

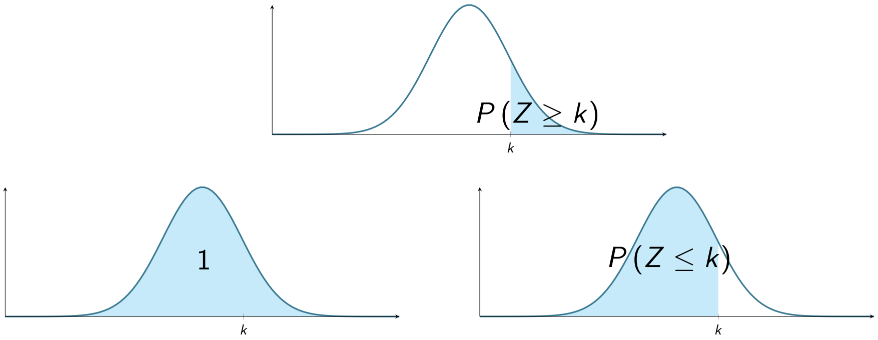
Veamos algunos ejemplos:

Cálculo en la $Z \sim \mathcal{N}(0,\,1)$ o normal estándar

Ejemplo

$P\,(Z \geq 1{,}23) = 1 - P\,(Z \leq 1{,}23) = 1 - 0{,}89065 = 0{,}10935.$

Basta fijarse en la relación que hay entre las áreas:

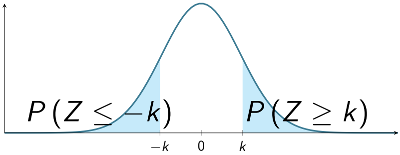


Cálculo en la $Z \sim \mathcal{N}(0, 1)$ o normal estándar

Ejemplo

$P(Z \leq -2{,}15) = P(Z \geq 2{,}15) = 1 - P(Z \leq 2{,}15) = 1 - 0{,}98422 = 0{,}01578$.

Basta fijarse en la relación que hay entre las áreas:

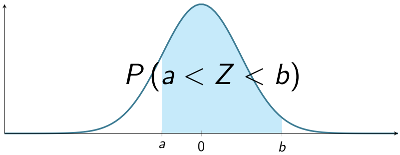


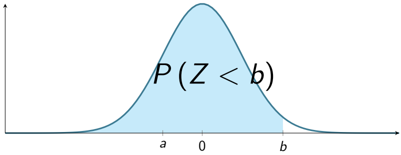
Cálculo en la $Z \sim \mathcal{N}(0,\, 1)$ o normal estándar

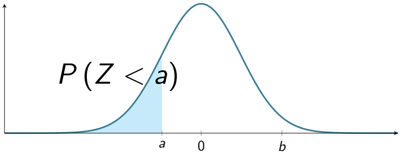
Ejemplo

$P\,(-1{,}3 < Z < 3{,}1) = P\,(Z < 3{,}1) - P\,(Z < -1{,}3) =$
$P\,(Z < 3{,}1) - [1 - P\,(Z < 1{,}3)] = 0{,}99903 - 1 + 0{,}9032 = 0{,}90223.$

Basta fijarse en la relación que hay entre las áreas:





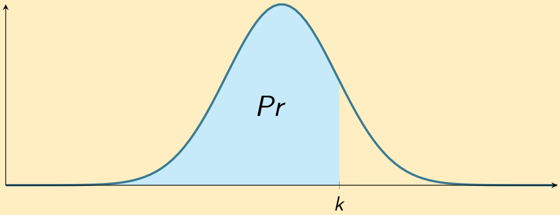


Cálculo en la $Z \sim \mathcal{N}(0, 1)$ o normal estándar

Cálculo del valor de la variable a partir de la probabilidad

El uso de la tabla normal nos permite realizar el proceso inverso. Es decir, fijada una probabilidad Pr, encontrar el valor de la variable k que cumpla:

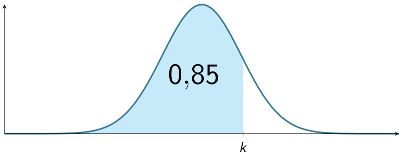
$$P(Z <= k) = Pr$$



Ejemplo

Dada $Z \sim \mathcal{N}(0, 1)$, calcula el valor de la variable sabiendo que la probabilidad de que tome un valor menor que ese es de un 85 %.

$$P(Z <= k) = 0{,}85$$



Vamos a la tabla y buscamos los dos valores seguidos de la tabla entre los que se quede el 0,85 y encontramos:

$$0{,}84849 < 0{,}85 < 0{,}85083$$

Como queda más cerca del 0,85083, me quedo con la celda correspondiente a la fila 1 y columna 0,04 $\Rightarrow k = 1{,}04$.

Cálculo en $X \sim \mathcal{N}(\mu,\, \sigma)$ - Tipificación

Para manejar $X \sim \mathcal{N}(\mu,\, \sigma)$ reduciremos los cálculos a cálculos en la $Z \sim \mathcal{N}(0,\, 1)$ a partir de la siguiente propiedad:

$$Si\ X \sim \mathcal{N}(\mu,\, \sigma) \Rightarrow X' = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,\, 1)$$

Al proceso de transformar la variable anterior a una $Z \sim \mathcal{N}(0,\, 1)$ se denomina **tipificarla**.

Ejemplos

- $X \sim \mathcal{N}(1, 2) \to P(X <= 2{,}32) = P(X' <= \frac{2{,}32-1}{2}) = P(Z <= 0{,}66) = 0{,}74537$
- $X \sim \mathcal{N}(5, 3) \to P(X <= 3{,}59) = P(X' <= \frac{3{,}59-5}{3}) = P(Z <= -0{,}47) = 1 - P(Z <= 0{,}47) = 1 - 0{,}68082 = 0{,}31918$

Estimación Puntual de la media y la varianza

Dada una población de media μ y desviación típica σ

Estimación de la media

Un buen estimador de μ es la **media muestral** $\overline{x}$:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Estimación de la varianza

Un buen estimador de σ^2 es la cuasivarianza muestral:

$$\widehat{\sigma}^2 = \frac{n-1}{n} s^2$$

siendo s^2 la varianza muestral:

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

Ejemplo

Una muestra aleatoria de 36 personas, empleadas en una gran industria, da el número medio de días al año que faltan al trabajo es $\overline{x} = 12$ con $s = 4$

- Dar una estimación puntual de μ (media poblacional)
  Un estimador es la media muestral que en este caso vale 12
- Dar una estimación puntual de σ^2 (varianza poblacional)
  Un estimador es la cuasivarianza muestral:

$$\widehat{\sigma}^2 = \frac{n-1}{n} s^2 = \frac{35}{36} \cdot 4 \approx 3{,}9$$

La estimación puntual sirve de poco mientras desconozcamos cuál es el grado de aproximación del estimador al parámetro real. Por ese motivo se procede a la **estimación mediante un intervalo de confianza**.

Veremos cómo calcular intervalos de confianza para estimar

- La media de la población
- La proporción de la población que cumple una característica

Distribuciones muestrales

A partir de los diferentes valores que puede tomar una muestra se pude construir una variable aleatoria a la que podemos asociar una probabilidad. Esto nos da una distribución a la que llamaremos **distribución muestral**

- Media La distribución de medias muestrales nos permitirá obtener intervalos de confianza de la media poblacional

$$\overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Proporción La distribución de proporciones muestrales nos permitirá obtener intervalos de confianza de la media poblacional

$$\widehat{p} \to N\left(p, \sqrt{\frac{p \cdot (1-p)}{n}}\right)$$

Ejemplo

Solución: Como $\overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, tipificando $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \to Z(0,1)$ y por tanto:
$P(\mu - 1 < \overline{X} < \mu + 1) = P(\frac{-1}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{1}{\frac{\sigma}{\sqrt{n}}}) \approx 0{,}967866646858422$

- Se ha seleccionado una muestra al azar de 50 mujeres de una población de mayores de 18 años. En la muestra se ha observado que la media de las 50 tallas es 1,60 m. Si se sabe que la desviación típica en la población es de 3,3 cm, determina la probabilidad de que la media de la población no difiere en más de 1 cm de la de la muestra.
- Como $\overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, tipificando $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \to Z(0,1)$ y por tanto:
  $P(\mu - 1 < \overline{X} < \mu + 1) = P(\frac{-1}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{1}{\frac{\sigma}{\sqrt{n}}}) \approx 0{,}967866646858422$

Estimación de la media por intervalo de confianza

A partir de una muestra de tamaño n y un grado de confianza $1-\alpha$:

Intervalo de confianza para la media

$$\left(\overline{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)$$

siendo:

- $\overline{x}$: La media de los datos de la muestra
- $z_{\alpha/2}$ o valor crítico: El valor de la distribución normal $Z \rightsquigarrow N((0,1)$ tal que $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$
- σ: La desviación típica de la distribución de la población (o si no se conoce de un estimador sesgado de la misma)
- n: El tamaño de la muestra

Ejemplo

Problema

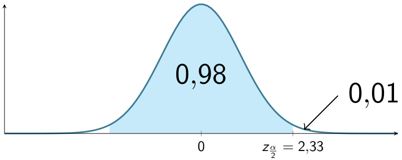
La duración de las bombillas de un fabricante es una variable aleatoria con distribución normal de desviación típica 75 horas. Decidimos tomar un tamaño de la muestra igual a 150, comprobamos la duración de cada bombilla y calculamos su promedio, que resulta ser 1053 horas. Calcular el intervalo de confianza al 98 % para la media de la duración de las bombillas del fabricante.

Ejemplo: Solución

Datos: $\overline{x} = 1053$, Confianza=98, $\sigma = 75$ y n=150

$\alpha = 1 - 0{,}98 = 0{,}02 \rightarrow \frac{\alpha}{2} = \frac{0{,}02}{2} = 0{,}01$.

Por tanto, el valor crítico será: $P\left(Z \leqslant z_{\alpha/2}\right) = 0{,}98 + 0{,}1 = 0{,}99 \rightarrow z_{\alpha/2} = 2{,}33$



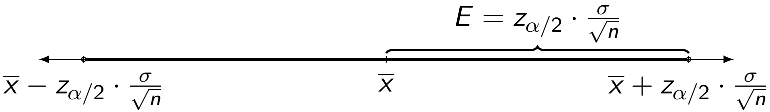
A partir de la definición de intervalo de confianza de la media:

$$\left(\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Operando nos queda:

$$\left(1053 - 2{,}33 \cdot \frac{75}{\sqrt{150}}, 1053 - 2{,}33 \cdot \frac{75}{\sqrt{150}}\right) = (1038{,}73, 1067{,}27)$$

Error máximo del intervalo de la media



Error máximo cometido:

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

Es el radio del entorno dado por el intervalo de confianza.

NOTA: Disminuye al aumentar el tamaño de la muestra y por tanto si se quiere garantizar un error determinado para un nivel de confianza habrá que tomar muestras de al menos un **tamaño determinado de la muestra**.

Ejemplo

Se desea realizar una investigación para estimar el peso medio de los recién nacidos de madres fumadoras. Se admite un error máximo de 50 gramos, con una confianza del 95 %. Si por estudios anteriores se sabe que la desviación típica del peso medio de tales recién nacidos es de 400 gramos, ¿qué tamaño mínimo de muestra se necesita en la investigación?

Ejemplo: Solución

Datos: $E = 50$, Confianza=95 y $\sigma = 400$
$\alpha = 1 - 0{,}95 = 0{,}05 \rightarrow \frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$.
Por tanto, el valor crítico será:
$P\left(Z \leqslant z_{\alpha/2}\right) = 0{,}95 + 0{,}025 = 0{,}975 \rightarrow z_{\alpha/2} = 1{,}96$
A partir de la definición de error máximo admitido:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

Luego:

$$n = \left(\frac{1{,}96 \cdot 400}{50}\right)^2 \approx 245{,}8534 \rightarrow n = 246$$

Estimación de la proporción por intervalo de confianza

A partir de una muestra de tamaño n y un grado de confianza $1 - \alpha$:

Intervalo de confianza para la proporción

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

siendo:

- $\overline{p}$: La proporción muestral
- $z_{\alpha/2}$ o valor crítico: $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$
- σ: La desviación típica de la distribución de la población
- n: El tamaño de la muestra

Llamaremos error máximo a:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\overline{p} \cdot (1 - \overline{p})}{n}}$$

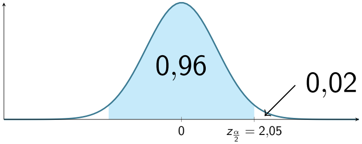
Ejemplo

Problema

Para estimar la proporción de personas con sobrepeso en una población se ha tomado una muestra aleatoria simple de tamaño 100 personas, de las cuales 21 tienen sobrepeso. Calcular el intervalo de confianza al 96 % para la proporción de personas con sobrepeso en la población.

Ejemplo: Solución

Datos: $\overline{p} = 0{,}21$, Confianza$=96$
$\alpha = 1 - 0{,}96 = 0{,}04 \rightarrow \frac{\alpha}{2} = \frac{0{,}04}{2} = 0{,}02$
.
Por tanto, el valor crítico será:
$P\left(Z \leqslant z_{\alpha/2}\right) = 0{,}96 + 0{,}02 = 0{,}98 \rightarrow$
$z_{\alpha/2} \approx 2{,}05$
A partir de la definición de error máximo admitido:



$$E = z_{\alpha/2} \cdot \sqrt{\frac{\overline{p} \cdot (1 - \overline{p})}{n}} \approx 2{,}05 \cdot \sqrt{\frac{0{,}21 \cdot 0{,}79}{100}} \approx 0{,}08$$

Luego el intervalo es:

$$(0{,}21 - 0{,}08, 0{,}21 + 0,08) = (0{,}13, 0{,}29)$$