Tema 9. Inferencia Estadística.

Intervalos de confianza.

Indice

1. Introducción.

El problema de la inferencia estadística es el inverso a los temas anteriores, que buscábamos la probabilidad de que ocurran distintas distribuciones planteadas. Ahora se trata a partir de los datos de muestras representativas se inferirán ***resultados acerca de la población,*** como por ejemplo estimar el valor de μ (estimación puntual de μ).

Por ejemplo si queremos calcular la altura media de todos los escolares, y para ello tenemos una muestra de n=100. ¿qué valor elegimos como el más aproximado a μ?. Si la media de la muestra es de 165cm, podremos afirmar que es "aproximadamente de 165 cm". Pero no podemos decir que exactamente el valor de μ es de 165cm, pues generalmente el valor de la media muestral no es exactamente el mismo que la media poblacional. Es por esto que esta ésta estimación se dice ***estimación puntual.*** Los estimadores puntuales sólo dan una idea aproximada del verdadero valor del parámetro a estimar, sin saber como de fiable es tal aproximación.

La estimación puntual es poco útil, es mucho más interesante obtener un intervalo dentro del cual se tiene cierta confianza (fijada de antemano) de que se encuentre el parámetro que se desee aproximar. Estimar un parámetro poblacional, por ejemplo μ, mediante un intervalo [a,b] con un nivel de confianza 1-α (que se suele dar en tanto por cién) se denomina ***estimación por intervalo de confianza → $P(a\leq\mu\leq b)=1-\alpha$***

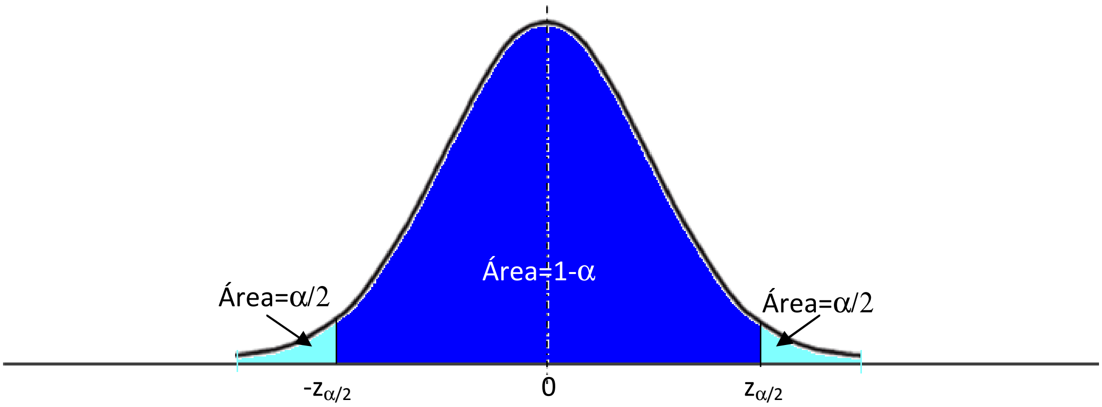
2. Intervalo de confianza para media poblacional. Tamaño de la muestra.

2.1. Intervalo de confianza

Partimos de una población formada por un gran número de elementos y de la que queremos estudiar una variable aleatoria X que sigue una distribución normal N(μ,σ) con media, μ, y desviación, σ, desconocidas.

Con el fin de estimar μ se toma una muestra aleatoria simple de tamaño n que nos proporciona una media $\overline{x_n}$, que será el estimador puntual de μ. Por el teorema central del límite (que vimos en el tema anterior) sabemos que la si la población grande, n>30, entonces las medias $\bar{x}$ siguen la ley normal N(μ,$\sigma/\sqrt{n}$), de forma que la variable tipificada será $z=\frac{\overline{x_n}-\mu}{\sigma/\sqrt{n}}$ que sigue distribución normal (N(0,1)). Si nos dicen el nivel de confianza es 1-α, el intervalo de confianza en Z será: $IC_Z=[-z_{\alpha/2}, z_{\alpha/2}]$. Siendo $z_{\alpha/2}$ el valor que cumple $P(Z\leq z_{\alpha/2})=1-\alpha/2$.

Veámoslo gráficamente:



$P(z\leq z_{\alpha/2})=1-\alpha+\alpha/2=1-\alpha/2$

Para obtener el intervalo de confianza de las medias de x, $\overline{x_n}$, y no de Z sólo tenemos que deshacer la tipificación, si $z=\frac{\overline{x_n}-\mu}{\sigma/\sqrt{n}}$ → $\overline{x_n}=\mu+z\cdot\frac{\sigma}{\sqrt{n}}$. De esta forma se cumple que el intervalo [a,b] de confianza de $\overline{x_n}$, equivalente al de z, $IC_Z=[-z_{\alpha/2}, z_{\alpha/2}]$, serán $a=\overline{x}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$, b= $\overline{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$ y x por lo que el intervalo de confianza es entonces:

$$\boxed{IC=[\ \overline{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \overline{x}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\ ]}$$

Siendo el error máximo cometido al estimar μ mediante $\overline{x_n}$ con precisión de 1-α igual a $E=z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$. Si n≥30 podemos asumir que σ=s, varianza muestral.

Los valores de $z_{\alpha/2}$ se encuentran sin problema en la tabla de distribución normal, aunque en la siguiente tabla ponemos los valores más usualmente usados:

Probabilidad	80%	90%	95%	99%
1-α	0,8	0,9	0,95	0,99
α(nivel significación)	0,2	0,1	0,05	0,01
$\alpha/2$	0,1	0,05	0,025	0,005
$z_{\alpha/2}$	1,282	1,645	1,960	2,575

Para cualquier otro valor de α, se debe utilizar la tabla de probabilidad, buscando $z_{\alpha/2}$ de forma que se cumpla $p(z\leq z_{\alpha/2})=1-\alpha$

Ejemplo: En los paquetes de arroz de cierta marca pone que el peso que contiene es de 500 gramos. Una asociación de consumidores toman 100 paquetes para los que obienen una media de 485g y desviación típica 10 g.

a) Se puede aceptar con un grado de signifiación igual a 0,05 que el fabricante está empaqutando realmente una media de 500g?
b) Calcular el intervalo de confianza al nivel de 99% para el peso de los paquetes.

Solución:

Deducimos del enunciado quela media muestral es $\bar{x} = 485$ y la desviación muestral es σ=10, para n=100, con α=0,05.

a) Para α=0,05, se cumple que $z_{\alpha/2}$=1,96, con lo que $E=z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}=1,96\cdot\frac{10}{\sqrt{100}}=1,96$,

de forma que el intervalo de confianza es entonces:

IC=(-E, $\bar{x}$+E)=(485-1.96, 485+1.96)=(483.04, 486.96), como $500\notin IC$ se puede estimar que las medias son diferentes, y por tanto no puede aceptarse que el fabricante esté empaquetando con una media de 500g.

b) En este caso $\bar{x}$=500g, y α=0,01. De esta forma, $z_{\alpha/2}$=2.575, con lo que el error máximo es $E= z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}=2.575\cdot\cdot\frac{10}{\sqrt{100}}=2.575$. El intervalo de confianza es ahora:

IC=(500-2.575, 500+2.575)=(497.425, 502.575)

2.2. Tamaño de la muestra

Recordemos que el error máximo cometido o radio del intervalo es E= $z_{\alpha/2}\cdot\sigma/\sqrt{n}$, y que por tanto el error disminuye con el tamaño de la mustra, n. Esto permite determinar el tamaño adecuado de la muestra como veremos en el siguiente ejemplo:

Ejemplo: Una variable X se distribuye según una ley normal N(μ,σ). ¿cuál debe de ser el tamaño de la muestra para que al estimar μ mediante la media muestral, $\overline{x_n}$, al nivel de confianza del 95% se cometa un error inferior a 0,3?

Solución:

En este problema lo que fijamos es el valor de error máximo, E=0.3, y buscamos el valor de n. Al ser el nivel de confianza de 95%, α=0,05 y por tanto $z_{\alpha/2}$=1.96. Despejando n del error podemos calcular fácilmente el valor de n:

E= $z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$, siendo E≤0.3 → $n \geq \left(\frac{z_{\alpha/2}\cdot\sigma}{E}\right)^2 = \left(\frac{1.96\cdot 2.5}{0,3}\right)^2 = 274.5 \rightarrow n \geq 275$

2.3. Resumen

En los problemas de intervalos de confianza para la media, m, de una población grande con desviación típica, σ, conocida (o aproximada a partir de la desviación muestral, s) intervienen los siguientes elementos:

- El tamaño de la muestra: n
- La media muestral $\bar{x}=\bar{x}_n$
- La distribución muestral de la media, $\bar{x}$: N(μ, $\sigma/\sqrt{n}$)
- El nivel de confianza (en tanto por cien): 95%, 99%, etc
- El nivel de significación, α, que es el tanto por uno de la diferencia 100-nivel confianza (0,05, 0,01, etc)
- El valor crítico de la variable tipificada, $z_{\alpha/2}$, que es el valor que cumple $p(z\leq z_{\alpha/2})=1-\alpha/2$ (se encuentra en la tabla de la distribución normal).
- El radio del intervalo o error máximo es E= $z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$
- El centro de la muestra, $\bar{x}=\bar{x}_n$
- El intervalo de confianza IC=[$\bar{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \bar{x}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$]

Ejercicios resueltos:

Tipos de ejercicios que pueden plantearse son básicamente ***tres tipos de problemas:***

- Dada una distribución conocido σ, n, α y la media $\bar{x}=\bar{x}_n$ determinar el intervalo de confianza
- Dada una distribución N(μ,σ) y conocidos σ, α, y el intervalo de confianza o E calcular el tamaño de la muestra.
- Dados n, σ e I determinar la media muestral $\bar{x}=\bar{x}_n$ y el nivel de confianza α.

P1. Una variable aleatoria X se distribuye con media μ desconocida y varianza σ=2,5. Se extrae una muestra aleatoria de tamaño n=100 y si tiene que su media $\bar{x}_{100}$=4,3. Construir un intervalo de confianza de μ al 95%.

Solución: Según el enunciado α=0,05, y por tanto $z_{\alpha/2}$=1.96, y el error máximo es entonces E= $z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}=1.96\cdot\frac{2.5}{\sqrt{100}}=0.49$. Así el IC=($\bar{x}_{100}$-E, $\bar{x}_{100}$+E)=(3.81, 4.79).

P2. Los estudiantes de Bachillerato de España duermen un número de horas diarias que se distribuye de forma normal con media μ desconocida y desviación σ=3. A partir de una muestra aleatoria de tamaño 30 se ha obtenido una media muestral de 7 horas. Hallar un intervalo de confianza, al 96%, para la media de horas de sueño, μ.

Solución:

Sea X="horas de sueño" que sigue la distribución normal N(μ,σ=3). Con fin de estimar el valor de μ tomamos la muestra con n=30 y media $\bar{x}_{30}$=7 horas, que es un estimado puntual de μ.

La media de las muestras de tamaño n=30, $\bar{x}_{30}$ siguen un distribución normal: $N(\mu,\frac{3}{\sqrt{30}})=N(\mu,1.73)$. El error máximo cometido con μ y con intervalo de confianza del 96% (α=0,04) será E= $z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$. Para calcular $z_{\alpha/2}$ miramos la tabla de la distribución normal el valor que cumple $P(z\leq z_{\alpha/2})=1-\alpha/2=0.98 \rightarrow z_{\alpha/2}=2.06$, y por tanto el error máximo es $E=2.06\cdot\frac{3}{\sqrt{30}}=1.1$.

Con los datos calculados antes se cumple entonces que IC=($\bar{x}_{30}$-E, $\bar{x}_{30}$+E)=(5.9, 8.1).

P3. El peso de los niños de 10 semanas de vida se distribuye según una normal con varianza de 87g. ¿cuántos niños serán suficientes para estimar con una confianza del 95% el peso medio de esa población con un error que no supere a 15g?

Solución: La variable aleatoria peso X sigue distribución N(μ,σ=87) con μ desconocida y con error inferior E≤15g.

Se cumple que E= $z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\leq 15 \rightarrow n\geq\frac{z_{\alpha/2}\cdot\sigma}{E}=129{,}641$

Por tanto el número de datos de la muestra ha de ser al menos de 130.

P4. a) Determinar el intervalo de confianza con el 95% para la media de una variable normal que tiene una desviación típica σ=3, teniendo en cuenta que se ha obtenido una muestra de tamaño 100 para el que $\bar{x}=\bar{x}_{100}=5$. b)¿Cuánto debería haber sido el tamaño de la muestra si se quiere obtener un intervalo de confianza para la media, también al 95%, con amplitud de 0,4.

Solución: La variable aleatoria X sigue una distribución normal con μ desconocida y σ=3. Por tanto $\bar{X}=\bar{X}_{100}$ sigue una ley normal N(μ, $\frac{3}{\sqrt{100}}$)

a) Para nivel confianza del 95%, por lo que α=0,05 y por tanto $z_{\alpha/2}$=1.96. El intervalo de confianza será $(\bar{x}-E,\ \bar{x}+E)$ con E= $z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$=0,588, con lo que IC=(4.412, 5.588)

b) Ahora 2E=2· $z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\leq 0{,}4 \rightarrow n\geq\frac{2\cdot z_{\alpha/2}\cdot\sigma}{0.4}=864.36$. Luego n tiene que ser mayor que 865.

3. Intervalo de confianza para una proporción.

Recordemos el teorema del límite para las proporciones: Si la distribución de una población grande tiene una prporción p de que ocurra un suceso A (q=1-p de que no ocurra), entonces la variable aleatoria P, de las porporciones muestrales extraidas de esa población se aproxima si el tamaño es grande ($n\geq 30$) a una distribución normal: $P\rightarrow N(\mu_p=p,\ \sigma_p=\sqrt{\frac{p\cdot q}{n}})$.

Con el fin de estimar el valor de p se toma una muestra aleatoria simple de tamaño n, que proporciona una media de proporción p_n, que es el estimador puntual de p. Pero si $n\geq 30$ entonces $\overline{P_n}$ sigue la distribución normal $N(\mu_p=p,\ \sigma_p=\sqrt{\frac{p\cdot q}{n}})$, con lo que la variable tipificada $Z=\frac{p_n-p}{\sqrt{\frac{p\cdot q}{n}}}$. Si conocemos el nivel de confianza determinamos α, y con este $z_{\alpha/2}$. De esta forma el error $E=z_{\alpha/2}\cdot\sqrt{\frac{p\cdot q}{n}}$ y por tanto $IC=(p_n-E,\ p_n+E)$.

Al igual que en el partado anterior podemos calcular el tamaño de la muestra mímo conocido el error máximo cometido. Sólo hay que despejar n de la fórmula del error:

$$n\geq\left(\frac{z_{\alpha/2}}{E}\right)^2\cdot p\cdot q$$

Ejercicios resuelto:

P.1 En una encuesta realizada entre 50 personas de una gran población, se han encontrado que el procentaje de individuos con gafas es del 25%. Determinar un intervalo de confianza al 99% para la proporción poblacional, p, de los individuos con gafas.

Solución: Para n=50 se cumple que $p_{50}=25/100=0.25$. Se cumple que la variable aleatoria P, de la proporición muestral sigue la normal $N(\mu_p=p\ ,\ \sigma=\sqrt{\frac{p\cdot q}{n}})$.

Como el nivel de confianza es del 99%, α=0,01 y $z_{\alpha/2}$=2.575. Con lo que el error máximo es E= $z_{\alpha/2}\cdot\sqrt{\frac{p\cdot q}{n}}$ =2.575$\cdot\sqrt{\frac{0.25\cdot 0.75}{50}}$ =0,16, y así el intervalo de confianza para la proporicón de gente con gafas es IC=(p_{50}-E, p_{50}+E)=(0.9, 0.41).

P.2. En una muestra de 120 personas extraida de cierta población muy numerosa, 20 de ellas eran portadores de un virus. Estima el intervalo de confianza para el porcentaje de personas que son portadores del virus en dicha población con un nivel de confianza del 90% y al 99%.

Solución:

Tenemos una muestra de una población, y una variable binomial (A=tener virus, $\bar{A}$=no tenerla). Si llamamos $\overline{P_n}$ a la media de la proporición, esta sigue una distribución normal con media $\bar{p}$=20/120=1/6 y desviación $\mu=\sqrt{\frac{p\cdot q}{n}}=0{,}034 \rightarrow$ N(1/6, 0.034).

a) Si la confianza es del 90%, entonces α=0.1 y $z_{\alpha/2}$=1.645. De esta forma el error máximo es E= $z_{\alpha/2}\cdot\sqrt{\frac{p\cdot q}{n}}$ =0.056 → IC=[0.111, 0.223]

b) Si la confianza es del 99%, entonces α=0.01 y $z_{\alpha/2}$=2.575. De esta forma el error máximo es E= $z_{\alpha/2}\cdot\sqrt{\frac{p\cdot q}{n}}$ =0.088 → IC=[0.079, 0.254]

Ejercicios propuestos:

P1. Sabemos que la edad de una población se comporta como una N(μ,10). Para estimar extraemos una muestra de tamaño 100, cuya media resulta ser de 37. Estimar μ mediante un intervalo de confianza del 90%.

P2. El peso de los alumnos de Bachillerato de cierta ciudad tiene una media desconocida y una desviación típica σ=5.4 kg. Tomamos una muestra aleatoria de 100 alumnos de Bachillerato de esa ciudad. Si la media poblacional es de 60 kg. Calcular al nivel de confianza del 99% el intervalo de confianza para el peso medio de todos los alumnos de Bachillerato de la ciudad.

P3. Se hizo una encuesta a 325 personas mayores de 16 años y se encontró que 120 iban al cine regularmente. Hallar con un nivel de confianza del 94% un intervalo para estudiar la proporción y el porcentaje de los ciudadanos que van al cine regularmente.

P4. Tomando al azar una muestra de 500 personas de una determinada comunidad se encontró que 300 leían la prensa regularmente. Halla, con confianza del 90% un intervalo para estudiar la proporción de lectores entre las personas de esa comunidad.

P5. El 60% de los empleados de una fábrica están a favor de trabajar los sábados. Se toma una muestra aleatoria. ¿Cuál deberá ser el tamaño de la muestra para que con un nivel de confianza del 95% el error máximo admisible en la estimación de 0.08?

P6. En una comunidad autónoma se sabe que la desviación típica del número de días que dura un contrato temporal es igual a 57 días. Indica el número mínimo de contratos en los que se han mirado su duración para que el intervalo con un nivel de confianza del 95% que da la duración media de un contrato de ese tipo tenga una amplitud no mayor de 10 días.