Tema 4: INFERENCIA ESTADÍSTICA - I

Intervalos de confianza para la media, muesras grandes, varianza conocida

Biología sanitaria 2017/18. Universidad de Alcalá

M. Marvá. Actualizado: 2018-03-20

La inferencia estadística:

- Predecir una característica poblacional a partir de una muestra
- La probabilidad permite asignar una fiabilidad a la predicción

Ejemplo: X la concentración de colesterol en sangre en población.
Tomar una muestra y

- Estimación puntual: calcular media muestral ¡¡cambia con la muestra!!
- Estimación por intervalos: buscar un intervalo en el que probablemente esté la media (valor aproximado)

En concreto, se buscan valores a y b tales que, dada una probabilidad p:

$$P(a < \mu_X < b) = p$$

Así

- conocemos el margen de error para esa medida
- conocemos la probabilidad de que la afirmación e sea cierta

Detalles en secciones 6.1 y 6.2 del libro

El teorema central del límite, versión 2.0

Sea X una v.a. con media μ_X y varianza σ_X^2. Se toma una muestra $X_1, X_2, \cdots, X_n$ de tamaño n (son copias independientes idénticas de X) y se calcula la media muestral

$$\bar{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Conforme crece n, la distribución de la media muestral se aproxima a una normal. En concreto:

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right) \qquad \Leftrightarrow \qquad \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \sim N(0,1)$$

Se acepta que para n>30 (muestras grandes) la aproximación es buena

La media muestral es un estimador de la media poblacional

Objetivo: dada una característica X, hacer predicciones tipo

$$P(a < \mu_X < b) = 0.9$$

Si $n > 30$, sabemos que

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$$

Ejercicio: si $W \sim N(\mu_W, \sigma_W)$, dibuja un intervalo para la v.a. W tal que

1. Tenga probabilidad 0.9.
2. Sea lo más pequeño posible.

Ejercicio: si $Z \sim N(0, 1)$, **determina** un intervalo para Z tal que

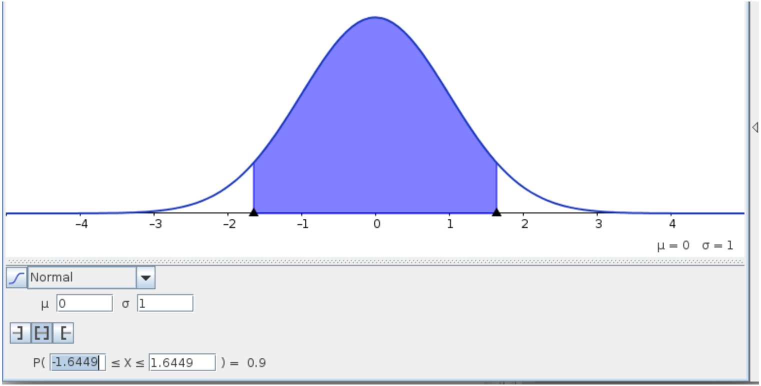
1. Tenga probabilidad 0.9.
2. Sea lo más pequeño posible.

Valores críticos para la distribución normal estándar $Z \sim N(0,1)$

El menor intervalo tal que

$$P(a < Z < b) = 0.9$$

es aquel en el que a es el percentil 5 y b el percentil 95.
OJO: se usa la notación $a = z_{0.95} \quad b = z_{0.05}$, pensando en la cola derecha, y se llaman valores críticos



Observa que, por simetría, $z_{0.95} = -z_{0.05}$

Por un lado, el menor intervalo con probabilidad 0.9 para $Z \sim N(0,1)$ es

$$P(-z_{0.05} < Z < z_{0.05}) = 0.9 \tag{1}$$

Por otro lado, si tomas muestras de tamaño n de X y conoces σ_X:

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right) \Leftrightarrow \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} = Z \sim N(0,1) \tag{2}$$

Sustituimos (2) en (1) y tenemos

$$P\left(-z_{0.05} < \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} < z_{0.05}\right) = 0.9 \tag{3}$$

Reorganizando términos en (3)

$$P\left(\bar{X} - z_{0.05}\frac{\sigma_X}{\sqrt{n}} < \mu_X < \bar{X} + z_{0.05}\frac{\sigma_X}{\sqrt{n}}\right) = 0.9$$

La probabilidad de que la **media poblacional** esté en ese intervalo es 0.9

Intervalo de confianza para la media: varianza conocida, n>30

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

- El **nivel de confianza** $1-\alpha$ es el grado de certidumbre que imponemos
- La **semianchura** del intervalo $z_{\alpha/2}\cdot\frac{\sigma_X}{\sqrt{n}}$ mide la precisión de la estimación.
- El error estándar de la muestra es $\frac{\sigma_X}{\sqrt{n}}$.

Ejemplo: De una población con varianza $\sigma^2 = 9$ conocida se toma una m.a.s. de tamaño 40 con media muestral $\bar{X} = 176$. Calcula el IC_{μ_X} con $nc = 0.95$

```
barX = 176; sigma2 = 9; library(asbio)
(barX + c(-1,1)*qnorm(.975)*sigma2/sqrt(40))
```

```
## [1] 173.2109 178.7891
```

```
ci.mu.z(summarized = TRUE, xbar = barX, sigma = sigma2, conf = 0.95,n = 40)
```

```
##
## 95% z Confidence interval for population mean
## Estimate      2.5%    97.5%
## 176.0000 173.2109 178.7891
```

Intervalo de confianza para la media: varianza conocida, n>30

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

Algunas observaciones importantes

- Aumentar sólo el tamaño de la muestra mejora la precisión pero no la probabilidad de acertar.
- Aumentar sólo el nivel de confianza $(1 - \alpha)$ mejora la probabilidad de acertar pero no la precisión.

Experimentar con el fichero GeoGebra correspondiente

Aún más importante

La construcción del intervalo de confianza es posible porque conocemos la distribución del estimador media muestral

- ¿Y si no conocemos la varianza de la población de estudio?
- ¿Y si la población de estudio no sigue una distribución normal?