Vídeos docentes sobre

Probabilidad y Teoría de la Decisión

Intervalos de confianza

F. J. Díez Vegas
Dpto. Inteligencia Artificial. UNED

fjdiez@dia.uned.es
www.ia.uned.es/~fjdiez

Estimación por intervalos de confianza

- Vamos a dar un intervalo para el parámetro
  - Escogemos α próximo a 0; p.ej., α = 0'05
  - $1-\alpha$ es el **coeficiente de confianza**; p.ej., 0'95
  - El **nivel de confianza** es $[(1-\alpha)\times 100]\%$; p.ej., 95%
- Buscamos un intervalo cuya probabilidad de contener el verdadero valor del parámetro sea $1-\alpha$
- Cuanto mayor es $1-\alpha$, mayor ha de ser el intervalo
  - Si α = 0, es decir, $1-\alpha$ = 1, entonces el intervalo debe contener todos los valores posibles del parámetro
    - Para la tasa de supervivencia, $IC_{100\%}$ = [0,1]
    - Para la concentración de X en sangre, $IC_{100\%}$ = [0, concentración-máxima]
- Hay que buscar un compromiso entre <u>certeza</u> y <u>precisión</u>

Intervalos de confianza para el segundo ejemplo

- Estudio sobre la relación entre la vitamina X y el síndrome Y
- Concentración de X para personas sanas:
  128 μg/cm³ (desviación estándar 20 μg/cm³)
- **Objetivo:** ¿cuál es μ (concentración de X en enfermos)?
- **Datos:** análisis de sangre en 25 pacientes con síndrome Y
  Promedio de concentración de X: 117 μg/cm³
  - ➢ Es de esperar que el intervalo de confianza para μ esté en torno a 117
- Vamos a fijar el nivel de confianza en el 95%

Construcción del intervalo de confianza

- Distribución poblacional: normal, $N(\mu, \sigma=20)$, con μ desconocido
- Tamaño de la muestra: $n = 25$
- Distribución muestral para el estadístico $\overline{X}$: $\overline{X} \sim N(\mu', \sigma')$

$\mu' = \mu$ (desconocido) $\qquad \sigma' = \dfrac{\sigma}{\sqrt{n}} = \dfrac{20}{\sqrt{25}} = 4$

- Nueva variable aleatoria (estadístico):

$$Z = \frac{\overline{X} - \mu'}{\sigma'} = \frac{\overline{X} - \mu}{4}$$

- Z se denomina **pivote**.
- Su distribución no depende de los parámetros del modelo.
- Distribución normal: $Z \sim N(0, 1)$ (independientemente de cuál sea μ)
- Cada investigador obtiene un $\overline{x}$: el promedio de X en su muestra
  (No puede conocer z porque no conoce μ)

Hacemos unos cálculos...

- Por las propiedades de la distribución normal sabemos que

$$P(-1'96 \le z \le 1'96) = 0'95$$

- Por la definición de Z

a) $z \le 1'96 \;\Rightarrow\; \dfrac{\bar{x}-\mu}{\sigma'} \le 1'96 \;\Rightarrow\; \bar{x}-\mu \le 1'96\,\sigma' \;\Rightarrow\; \bar{x}-1'96\,\sigma' \le \mu$

b) $z \ge -1'96 \;\Rightarrow\; \dfrac{\bar{x}-\mu}{\sigma'} \ge -1'96 \;\Rightarrow\; \bar{x}-\mu \ge -1'96\,\sigma' \;\Rightarrow\; \bar{x}+1'96\,\sigma' \ge \mu$

de modo que

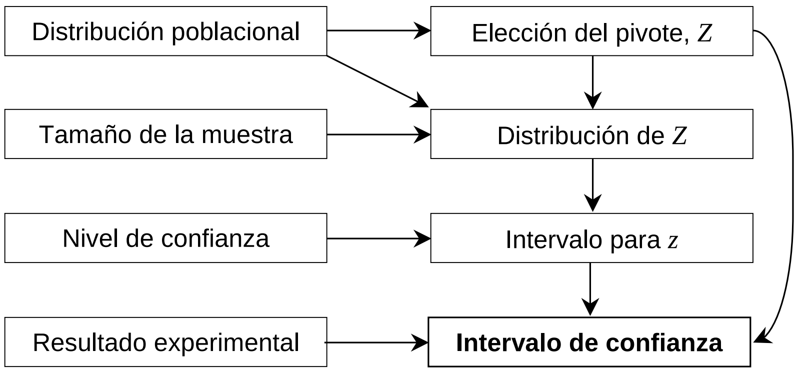
$$-1'96 \le z \le 1'96 \;\Rightarrow\; \bar{x}-1'96\,\sigma' \le \mu \le \bar{x}+1'96\,\sigma'$$

- Por tanto

$$P(\bar{x}-1'96\,\sigma' \le \mu \le \bar{x}+1'96\,\sigma') = 0'95$$

es decir, la probabilidad de que el intervalo $[\bar{x}-1'96\,\sigma',\ \bar{x}+1'96\,\sigma']$ contenga el verdadero valor de μ es el 95%

Cálculo del intervalo de confianza



Interpretación de los intervalos de confianza

- Supongamos que el intervalo de confianza al 95% es [μ_1, μ_2]
- Afirmación: "Hay un 95% de probabilidad de que el verdadero valor de μ esté entre μ_1 y μ_2." ¿Cómo debe entenderse?
- En una probabilidad siempre hay algo aleatorio
- <u>Lo aleatorio</u>, en este caso, <u>es el intervalo</u> que cada investigador obtiene
- <u>El parámetro μ no varía</u>
- Por eso hay quien prefiere decir: "La probabilidad de que el intervalo [μ_1, μ_2] <u>contenga</u> el verdadero valor de μ es el 95%"
    - ➢ Puede entenderse como equivalente a la afirmación anterior, pero no se presta tan fácilmente a una interpretación errónea

Ejemplo

- 20 investigadores buscan el verdadero valor de μ
- Cada uno obtiene un intervalo
- El 95% de los investigadores acierta; el 5% obtiene un intervalo incorrecto
- Problema:
  ninguno de ellos sabe si ha acertado o no