

Estadística no paramétrica. Trabajo 1.

Máster en Técnicas Estadísticas. Curso 2009-2010

Coeficiente de correlación (permutaciones)

Sea $X = (Y, Z)$ un v.a. bidimensional y consideremos $(Y_1, Z_1), \dots, (Y_n, Z_n)$ una m.a.s. de X . Denotemos por F la distribución de X (distribución conjunta) y por F_Y, F_Z las marginales correspondientes, ambas distribuciones Normales. Queremos contrastar:

$$H_0 : Y, Z \text{ independientes, vs. } H_a : Y, Z \text{ dependientes}$$

Bajo la hipótesis nula de independencia, el coeficiente de correlación $\rho = 0$. Si consideramos como estadístico de contraste el coeficiente de correlación muestral, $\hat{\rho}$, rechazaremos H_0 cuando el coeficiente de correlación muestral sea grande. Por tanto, el p -valor del estadístico de contraste será $\mathbb{P}(\hat{\rho} \geq \hat{\rho}_{obs} | H_0)$.

Bajo H_0 , podemos considerar la muestra observada como:

$$(y_{(1)}, \dots, y_{(n)}, z_{(1)}, \dots, z_{(n)})$$

por lo que podemos obtener una remuestra como:

$$\mathcal{X}^* = \{(y_{(1)}, Z_1^*), \dots, (y_{(n)}, Z_n^*)\}$$

donde (Z_1^*, \dots, Z_n^*) es una permutación de $(z_{(1)}, \dots, z_{(n)})$. Para una m.a.s. de tamaño n tenemos $n!$ posibles permutaciones. En la práctica, se consideran B permutaciones.

1. **Caso 1.** Considera Y, Z independientes con distribución $F_Y = N(0, 4)$ y $F_Z = N(2, 1)$. Para $n = 20$ y $B = 500$, obtén una estimación de la distribución de $\hat{\rho}$. Genera una muestra de (Y, Z) y aproxima el p -valor asociado al coeficiente de correlación muestral.
2. **Caso 2.** Considera Y, Z variables Normales con media cero, varianzas $\sigma_Y^2 = 4, \sigma_Z^2 = 1$ independientes. Para $n = 20$ y $B = 500$, obtén una estimación de la distribución de $\hat{\rho}$ y del p -valor asociado al coeficiente de correlación de una muestra de $(Y, Z) \sim (F_Y, F_Z)$ con correlación $\rho = 0,25$.

3. En el primer caso, obtén un intervalo de confianza por permutaciones para ρ , considerando $\alpha = 0'05$ (intervalo de nivel $(1 - \alpha)$) y evalúa su cobertura real del siguiente modo:
- a) Calcula el IC por permutaciones para ρ , con $B = 500$ y $\alpha = 0'05$.
 - b) Repite el procedimiento anterior $M = 500$ veces y calcula el porcentaje de veces que ρ_0 (valor verdadero de la correlación) cae en el IC. Este porcentaje debe estar próximo al 95 %.