Chap 12

non-P Correlation

& SLR

Ⓐ non-P Correlation

Data: $(x_1, y_1), \ldots, (x_n, y_n)$

Spearman's ρ_s

first rank two samples (x & y) separately

$r_1^{(y)}, \ldots, r_n^{(y)}$ & $r_1^{(x)}, \ldots, r_n^{(x)}$

& put back into pairs $(r_i^{(x)}, r_i^{(y)})$ $i = 1, \ldots, n$

Then calculate ordinary Pearson-ρ correlation on the rank pairs.

$\rightarrow \hat{\rho}_s$

WTK: $H_0: \rho_s = \phi$ (no association)

v. $H_1: \rho_s \neq \phi$ ⟸ only ϕ

Simulation: randomly permuting ranks for two sets of $\{1, \ldots, n\}$, forming random rank pairs

→ & calculate $\hat{\rho}_s$

Math: SSP, but more complicated b/c of the cross terms in Pearson-ρ

↳ use CDF pSpearman in R. library (SuppDists)

Kendall's - τ

↳ Rather than looking at n pairs, looks at the "agreement" of all pairs of pairs (x_i, y_i) & (x_j, y_j) of which there are $\binom{n}{2} = \frac{n(n-1)}{2}$ total

concordant if $\frac{y_i - y_j}{x_i - x_j} > 0$ discordant if $\frac{y_i - y_j}{x_i - x_j} < 0$

half concordant / discordant if $y_i = y_j$

dicard the pair if $x_i = x_j$

Let n_c & n_d denote the number of concordant & discordant pairs

[see book for algorithm]

$$\hat{\tau} = \frac{n_c - n_d}{n_c + n_d}$$

$n_c + n_d \longrightarrow \binom{n}{2} = \frac{n \cdot (n-1)}{2}$ if no ties in xs

symmetry problem
x & y switched

but cor(x, y, method = "kendall") solves it.

WTK $H_0: \tau = \phi$
v. $H_1: \tau \neq \phi$

simulation is easy, just calculate $\hat{\tau}$ w/ totally random X & Y

→ Anything, even random ranks.

math: R has builtin CDF pkendall

Chap 12

(B) non-P SLR

Correlation (A) is linear, and non-P SLR leans on that connection

Testing: Spearman's ρ
CI: Kendall's τ

Data are (x_i, y_i) pairs &

$$\mathbb{E}\{Y_i\} = \beta_0 + \beta_1 x_i \quad i = 1, \ldots,$$

Suppose WTK $H_0: \beta_1 = \beta_1^{(0)}$
v. $H_1: \beta_1 \neq \beta_1^{(0)}$

Calculate zero-intercept (or intercept-free) residuals

$$u_i = y_i - \beta_1^{(0)} x_i \qquad i = 1, \ldots, n$$

like e_i but β_0 not involved

Then, use Spearman's ρ on (x_i, u_i).

<u>CI</u> Each concordant/discordant pair involve a calculation of rise/run

$$S_{ij} = \frac{y_i - y_j}{x_i - x_j} \quad \frac{\Delta y}{\Delta x} \equiv \text{slope}$$

$$1 \le i < j \le n \qquad \underbrace{n_s = \binom{n}{2}}_{\text{unless discard some pairs}} = \frac{n(n-1)}{2}$$

Once you have the S_{ij}, you must figure out which ones lie "near" the desired quantiles under Kendall's τ

→ qkendall or via simulation

(q)

using this with s_{ij} require

rescaling the quantile to indices $\{1, \ldots, n_s\}$

for s.

<u>Prediction</u> for non-P SLR

Sensible to use median for everything

$$\hat{\beta}_1 = s^{(n/2)} \quad \& \quad \hat{\beta}_0 = y^{(n/2)} - \hat{\beta}_1 x^{(n/2)}$$

However, if you want uncertainty (like PI or CI) you must simulate

→ there is no "closed form"

→ see book.