Chapter 8

non-parametric stats

(non-P)

Non-P means no "model".

No: $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$ or $Y_i \overset{iid}{\sim} Bern(\theta)$

parameters

Still have iid, mean $\mathbb{E}\{Y_i\}$, variance $\mathbb{V}ar\{Y_i\}$

Fewer choices/assumptions.

- The CLT can be non-P, but in Ch4 we used it in a P-context.

(A) Bootstrap

↳ inherently simulation-based

↳ main procedure is sampling with replacement

Consider any statistic of interest.

$s(y_1, \ldots, y_n)$ e.g. $\bar{Y}$, $\underline{\text{med}(y_1, \ldots, y_n)}$

s or s^2 ← sd or var

statistic

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \to \begin{matrix} Y_1^* \to s(Y_1^*) \\ Y_2^* \to s(Y_2^*) \\ \vdots \\ Y_B^* \to s(Y_B^*) \end{matrix} \Bigg\} \begin{matrix} \text{collection} \\ \text{an} \\ \text{empirical} \\ \text{sampling} \\ \text{distribution} \end{matrix}$$

sample w/ replacement

B → # bootstrap resamples

$s(y_1, \ldots, y_n)$

The bootstrap resamples are best understood or utilized in a CI calculation.

The simlest way to test a hypothesis with a Bootstrap is to "invert" the CI.

- if the value you're wondering about (H_0) is outside the CI, then reject H_0

Chapter 8
non-p

(B) Permutation Tests

non-p version of Ch5
a two-sample test

$Y_1, \dots, Y_{n_Y} \overset{iid}{\sim} F_Y \qquad X_1, \dots, X_n \overset{iid}{\sim} F_X$

$H_0: F_Y = F_X$
$H_1: F_Y \neq F_X$

Permutation tests target a particular aspect of F_Y & F_X via a choice of statistic on the combined sample

$$s(Y, X) = s(\underbrace{y_1, \dots, y_{n_Y}}_{\bar{Y}}, \underbrace{x_1, \dots, x_{n_X}}_{\bar{X}})$$

e.g. $s(y,x) = \bar{Y} - \bar{X}$

$s(y,x) = sd(y)/sd(x)$

Idea: if H_0 is true, the value of the stat $s(y_1, \ldots, y_{n_y}, x_1, \ldots, x_n)$ won't change much if you permute the combined sample.

- if $n_y + n_x$ is small, you can exhaustively enumerate all permutations.
- in practice, one randomizes