Chapter 11
Non-P Scale

(A) Squared Ranks

$$Y_1, \ldots, Y_{n_y} \sim F_y \qquad X_1, \ldots, X_{n_x} \sim F_x$$

$$H_0: \mathbb{V}ar\{Y\} = \mathbb{V}ar\{X\}$$
$$\text{v. } H_1: \mathbb{V}ar\{Y\} \neq \mathbb{V}ar\{X\}$$

Convert each y_i & x_j into its absolute deviation from the sample average.

$$V_i = |y_i - \bar{y}| \qquad i = 1, \ldots, n_y$$
$$u_j = |x_j - \bar{x}| \qquad j = 1, \ldots, n_x$$

Assign ranks to the combined sample

$$V_1, \ldots, V_{n_y}, u_1, \ldots, u_{n_x}$$
$$\rightarrow r_1^{(v)}, \ldots, r_{n_y}^{(v)}, r_1^{(u)}, \ldots, r_{n_x}^{(u)}$$

Several common test stats

$$\bar{r}^2 \equiv \bar{r}^{2(v)} = \sum_{i=1}^{n_y} r_i^{2(v)}$$

Sampling distribution easy by simulation, but harder (doable) by <u>math</u>.

See Appendix B.3

Involves solving a SSP

A CDF is provided as psqranks in ranks.R on book webpage.

Chap 11
Continued

(B) Kruskal-Wallis (KW)
(non-p ANOVA)

data Y_{ij} $\quad i = 1, \dots, n_j \quad j = 1, \dots, m$

$$H_0 : \mathbb{E}\{Y_{ij}\} = \mu \quad \text{for all } ij$$
$$\text{v. } H_1 : \mathbb{E}\{Y_{ij}\} \neq \mathbb{E}\{Y_{ik}\} \quad \text{for some } j \neq k$$
(for all i)

Let r_{ij} denote the rank of the combined sample of all $n = \sum_{j=1}^{m} n_j$ observations

Denote $\bar{r}_j = \sum_{i=1}^{n_j} r_{ij} \quad j = 1, \dots, m$

Combine size-adjusted squares of these rank aggregates

Test stat: $\bar{r}^2 = \sum_{j=1}^{m} \bar{r}_j^{\,2} / n_j$

KW test by math

SSP-based technically doable, but large $\bar{r}^2$ leads to huge computation.

A CLT-like argument provides

$$\frac{\bar{R}^2 - n(n+1)^2/4}{\frac{1}{n-1}\left[\sum_{j=1}^{m}\sum_{i=1}^{n_j} r_{ij}^2 - \frac{n(n+1)^2}{4}\right]} \sim \chi^2_{m-1} \text{ as } n \to \infty$$