<u>Rank-based ANOVA</u>

Stat 505
3-6-18

(1)

Let's revisit the 1-way ANOVA

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \begin{array}{l} i = 1, \dots, a \\ j = 1, \dots, n \end{array} \qquad N = na$$

Usual assumptions: $\varepsilon_{ij} \sim$ iid $N(0, \sigma^2)$

If the assumptions are not met, and n is small, then an alternative method is recommended.

(2)

Apply the <u>rank</u> <u>transformation</u>.

Rank all N of the data values from 1 (smallest) to N (largest)

- If there are ties, average the ranks for the items involved in the tie.

Compute F:

$$SS_{TOT} = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$$

After ranking, $SS_{TOT} = \sum_{i=1}^{a} \sum_{j=1}^{n} (R_{ij} - \bar{R}_{..})^2$

3

Sum of integers from 1 to N is $\frac{N(N+1)}{2}$

Sum of squares from 1 to N is $\frac{N(N+1)(2N+1)}{6}$

$$\bar{R}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{b} R_{ij} = \frac{1}{N}\frac{N(N+1)}{2} = \frac{N+1}{2}$$

$$SS_{TOT} = \sum_{i=1}^{a}\sum_{j=1}^{n}(R_{ij}-\bar{R}_{\cdot\cdot})^2 = \sum_i\sum_j R_{ij}^2 - N\bar{R}_{\cdot\cdot}^2$$

$$= \frac{N(N+1)(2N+1)}{6} - N\left(\frac{N+1}{2}\right)^2 = \frac{N(N^2-1)}{12}$$

4

This implies that $SS_{TRT} + SS_{ERR} =$ constant,

so MS_{TRT} & MS_{ERR} are no longer independent.

Compute $SS_{TRT} = \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2$

$$= n\left(\sum_{i=1}^{a}\bar{y}_{i\cdot}^2 - a\,\bar{y}_{\cdot\cdot}^2\right)$$

After ranking,

$$SS_{TRT} = n\left(\sum_{i=1}^{a}\bar{R}_{i\cdot}^2 - a\,\bar{R}_{\cdot\cdot}^2\right)$$

(5)

$$= n \sum_{i=1}^{a} \bar{R}_{i\cdot}^{2} - N\left(\frac{N+1}{2}\right)^2$$

$$= n \sum_{i=1}^{a} \left(\frac{R_{i\cdot}}{n}\right)^2 - \frac{N(N+1)^2}{4}$$

$$SS_{TRT} = \frac{1}{n} \sum_{i=1}^{a} R_{i\cdot}^2 - \frac{N(N+1)^2}{4}$$

In the usual F test, we have that

under H_0, $\frac{SS_{TRT}}{\sigma^2} \sim \chi^2_{a-1}$

(6)

But after ranking, what is σ^2?

Before, it was $V(\varepsilon_{ij}) = V(y_{ij})$

Now, $\sigma^2 = V(R_{ij})$

$= V(\text{discrete uniform distribution})$

$= \frac{N^2-1}{12}$

$= \frac{1}{N} SS_{TOT}$

7

Now
$$\frac{SS_{TRT}}{\sigma^2} = \frac{\frac{1}{n}\sum_{i=1}^{a} R_{i.}^2 - \frac{N(N+1)^2}{4}}{\frac{N^2-1}{12}}$$
$$= \frac{12}{N^2-1}\left[\frac{1}{n}\sum_{i=1}^{a} R_{i.}^2 - \frac{N(N+1)^2}{4}\right]$$
$$= \frac{N}{N-1}\underbrace{\left[\frac{12}{N(N+1)}\sum_{i=1}^{a}\frac{R_{i.}^2}{n} - 3(N+1)\right]}_{H}$$

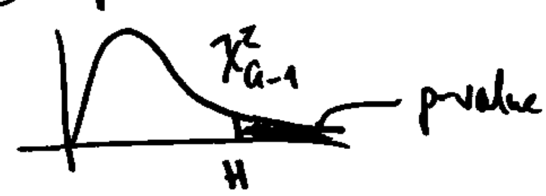
H is the Kruskal-Wallis test statistic

8

H has an approximate χ^2_{a-1} distribution.

(So does $\frac{N}{N-1}H$)

Kruskal-Wallis test:

1. Rank the data
2. Find the sum of the ranks for each of the a levels of the factor
3. Compute H
4. 

(9)

But what is actually being tested?

$$E[g(Y)] \neq g(E[Y])$$

So we are no longer testing $H_0: \mu_1 = \mu_2 = \dots = \mu_a$

After ranking, you are actually testing the null hypothesis that the population <u>medians</u> are all equal.

(10)

#2 from in-class portion of mid-term

$$Y_{ijk} = \mu + \tau_i + \gamma_{ij} + \varepsilon_{ijk} \qquad i = 1,2,3$$

$$\sum_{i=1}^{3} \tau_i = 0 \qquad \sum_{j=1}^{3} \gamma_{ij} = 0 \qquad \begin{array}{l} j = 1,2,3 \\ n = 1,2,3 \end{array}$$

Write the matrix form

$$
\underset{27\times 1}{\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{131} \\ y_{132} \\ y_{133} \\ \vdots \\ y_{332} \\ y_{333} \end{bmatrix}}
=
\underset{27\times 9}{\left[\begin{array}{c|cc|cccccc}
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\
\hline
\vdots & & & & & & & & \\
1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\
1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1
\end{array}\right]}
\underset{9\times 1}{\begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{bmatrix}}
+ \underset{27\times 1}{\vec{\varepsilon}}
$$

Source	df
A	2
G	6
ERR	18
TOT	26

12

#3 from in-class midterm

$a = 3 \quad k = 2 \quad \lambda = 2$

$N = ar = bk$ and $\lambda = \dfrac{r(k-1)}{a-1}$

$2 = \dfrac{r(2-1)}{3-1}$

$\therefore \underline{r = 4}$ (b)

$N = 3 \cdot 4 = b \cdot 2$

$\therefore \underline{b = 6}$ (a)

13

(c)

BLK

TRT	1	2	3	4	5	6
1	✓	✓		✓	✓	
2	✓		✓	✓		✓
3		✓	✓		✓	✓