Examples for 11/19/15, Part 4

Kruskal-Wallis test for equivalence of means:

Let $f(x)$ be a density of a continuous random variable with mean 0.

Assume Y_{ij}, $i = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, J$, are independent random variables with density $f(x - \mu_j)$.

(The J populations have no parametric assumptions, they are assumed to have densities with a common shape, but perhaps different centers.)

$H_0: \mu_1 = \mu_2 = \ldots = \mu_J$ $\qquad$ H_1: not all of the μ_j are equal.

$\mu_i \neq \mu_j$ for at least one pair i and j.

Let r_{ij} be the respective rank of a data point when all the data is ranked from smallest to largest.

Let $\bar{r}_j$ be the mean of the ranks for each group. Let $\bar{r} = \dfrac{N+1}{2}$ be the grand mean of the ranks.

Test statistic:

$$K = \frac{12}{N(N+1)} \sum_{j=1}^{J} n_j \left(\bar{r}_j - \bar{r}\right)^2 = \frac{12}{N(N+1)} \sum_{j=1}^{J} n_j \left(\bar{r}_j - \frac{N+1}{2}\right)^2 = \frac{12}{N(N+1)} \sum_{j=1}^{J} n_j \bar{r}_j^{\,2} - 3(N+1).$$

Reject H_0 if $K > \chi^2_\alpha(J-1)$.

1. Six samples of each of four types of cereal grain grown in a certain region were analyzed to determine thiamin content, resulting in the following data (μg/g):

Wheat	5.2	4.5	6.0	6.1	6.7	5.7
Barley	6.5	8.0	6.1	7.5	5.9	5.6
Maize	5.8	4.7	6.4	4.9	6.0	5.2
Oats	8.3	6.1	7.8	7.0	5.6	7.2

Does this data suggest that at least two of the grains differ with respect to true average thiamin content? Use $\alpha = 0.05$.

W	**M**	**M**	W	**M**	B	O	W	**M**	B	W	**M**
4.5	**4.7**	**4.9**	5.2	**5.2**	5.6	5.6	5.7	**5.8**	5.9	6.0	**6.0**
1	**2**	**3**	4.5	**4.5**	6.5	6.5	8	**9**	10	11.5	**11.5**

W	B	O	**M**	B	W	O	O	B	O	B	O
6.1	6.1	6.1	**6.4**	6.5	6.7	7.0	7.2	7.5	7.8	8.0	8.3
14	14	14	**16**	17	18	19	20	21	22	23	24

Wheat $\bar{r}_W = 9.5$ **Maize** $\bar{r}_M = 7.6666\overline{6}$

Barley $\bar{r}_B = 15.25$ Oats $\bar{r}_O = 17.5833\overline{3}$ $\bar{r} = 12.5$

$$K = \frac{12}{24 \cdot 25}\left[6\cdot(9.5-12.5)^2 + 6\cdot(15.25-12.5)^2 + 6\cdot\left(7.6666\overline{6}-12.5\right)^2 + 6\cdot\left(17.5833\overline{3}-12.5\right)^2 \right] = 7.8916\overline{6}.$$

A correction for ties can be made, but this correction usually makes little difference in the value of K unless there are a large number of ties.

$\chi^2_\alpha(J-1) = \chi^2_{0.05}(3) = 7.815.$ $K > 7.815.$ **Reject H_0** at $\alpha = 0.05$.

```r
Wheat <- c(5.2,4.5,6.0,6.1,6.7,5.7)
Barley <- c(6.5,8.0,6.1,7.5,5.9,5.6)
Maize <- c(5.8,4.7,6.4,4.9,6.0,5.2)
Oats <- c(8.3,6.1,7.8,7.0,5.6,7.2)

Grain <- c(rep("Wheat",6), rep("Barley",6), rep("Maize",6), rep("Oats",6))
Thiamin <- c(Wheat, Barley, Maize, Oats)

Cereal <- data.frame(Grain, Thiamin)

kruskal.test(Thiamin ~ Grain, data = Cereal)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Thiamin by Grain
## Kruskal-Wallis chi-squared = 7.9158, df = 3, p-value = 0.04779
```