Lecture 12

ANOVA alternatives

We will use the data from the previous lecture for the examples here. The data was normal enough with homogeneous (equal) variances in each sample that using a non-parametric test is not particularly desirable. So while we will apply these tests below to the data, standard ANOVA is the preferable alternative. There are 3 alternatives to standard one-way ANOVA that we will discuss below: Kruskal-Wallis, Moods Median Test, and Welch's ANOVA.

Kruskal-Wallis Test

When there are serious outliers, the Kruskal-Wallis Nonparametric version of one-way ANOVA based on the ranks of the data may be useful. The Kruskal-Wallis statistic is ANOVA-like:

$$KW = \frac{\text{Between Group Sum of Squares of Ranks}}{\text{variance in the ranks}}$$

The Kruskal-Wallis statistic is approximately Chi-Square distributed with $I - 1$ degrees of freedom where I is the number of populations. The null hypothesis is that all the of I populations have the same distribution versus the alternative hypothesis that they do not all have the same distribution.

In Minitab, Kruskal-Wallis is under Stat > Nonparametrics > Kruskal-Wallis...

Mood's Median Test

The procedure behind Mood's Median test is this: Find the overall median. For each population/group, count the number of data points that are above (below) the median. Form a contingency table based on these counts and ask if the distribution is the same for all populations. If not, then it seems unlikely that the common median is not the common median for each population - some populations are different than others. The Mood's Median statistic is also Chi-Square distributed with $I - 1$ degrees of freedom.

Welch's ANOVA

Welch's ANOVA is a generalization of Welch's t-test, of course. It allows for unequal variances in each population and is an important tool when the design is unbalanced and the variances are unequal. Unfortunately, it is not pre-programmed into Minitab. I believe it is pre-programmed into SPSS. It is possible to calculate Welch's F-test by hand following these steps:

1. For each population, calculate the within population sample variance, s_i^2.
2. For each population, calculate a weight $w_i = n_i/s_i^2$ where n_i is the sample size from the i^{th} population.
3. Calculate a weighted mean: $\bar{X} = \sum_{i=1}^n w_i \bar{X}_i / \sum_{i=1}^n w_i$ where $\bar{X}_i$ is the sample mean from the i^{th} population.
4. Let I be the number of population. Calculate Welch's F-statistic as

$$F = \frac{\left(\sum_{i=1}^n w_i(\bar{X}_i - \bar{X})^2\right)/(I-1)}{1 + \frac{2(I-2)}{I^2-1}\sum_{i=1}^n \frac{1}{n_i-1}\left(1 - \frac{w_i}{\sum_{j=1}^n w_j}\right)^2}$$

5. The numerator degrees of freedom is $I-1$. The denominator degrees of freedom is given by the formula:

$$\frac{I^2-1}{3\sum_{i=1}^n \left(\frac{1}{n_i-1}\right)\left(1 - \frac{w_i}{\sum_{j=1}^n w_j}\right)^2}$$

Exercises for Lecture 12

1. –
2. –