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 most appropriate test. 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 following table ranks the original data. That is, the numbers in parentheses are the ranks of the raw number (the percentage of women on the venire) in the overall data set:

Trial Judge	Judge A	Judge B	Judge C	Judge D	Judge E	Judge F
16 (6.5)	40 (43.5)	36 (41.5)	34 (37.5)	24 (18)	33 (36)	22 (15.5)
18 (10)	30 (30)	32 (34)	30 (30)	30 (30)	36 (41.5)	21 (14)
14 (3)	16 (6.5)	32 (34)	32 (34)		28 (23.5)	31 (32)
6 (1)	35 (39.5)	27 (21.5)	29 (26.5)		20 (12.5)	27 (21.5)
18 (10)	50 (46)	29 (26.5)	24 (18)		18 (10)	17 (8)
15 (4.5)		45 (45)	28 (23.5)		22 (15.5)	29 (26.5)
15 (4.5)			20 (12.5)		40 (43.5)	26 (20)
9 (2)			35 (39.5)			29 (26.5)
24 (18)						34 (37.5)

For instance, 6% is the smallest percent women on any venire and so it is ranked first (1). The percentage occurred on one of the trial judge's venires. There are three 32%'s in the data set and each is given the rank (34) because these would naturally fall in ranks 33, 34, and 35 and rank 34 is the average of these. Similarly, the two 16%'s get the average rank of 6.5 because they would have naturally fallen into ranks 6 and 7.

The Kruskal-Wallis statistic is the following ANOVA-like statistic:

$$
\begin{aligned}
KW &= \frac{\text{Between Group Sum of Squares of Ranks}}{\text{variance in the ranks}} \\
&= \frac{12\sum_{i=1}^{I} n_i(\bar{R}_i - \bar{R})}{N(N+1)}
\end{aligned}
$$

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. Thus, rejecting the null hypothesis may not mean the centers of the distributions are different if the shape of the distributions are also different. In particular, if the variances are unequal, then that might be enough to cause you to reject that the distributions are the same even if the medians of the distributions are the same. This is mainly a problem if the design is unbalanced. Outliers can mean unequal variances but they can mean that you just happened by chance to sample an extreme point in one population whereas you did not happen to sample them in another. So the Kruskal-Wallis test is useful for dealing with unequal variances and lack of normality due to the presence of outliers.

In Minitab, Kruskal-Wallis is under Stat > Nonparametrics > Kruskal-Wallis... We will discuss the following output in class.

```
Kruskal-Wallis Test: percent_women versus judge

Kruskal-Wallis Test on percent_women

judge          N  Median  Ave Rank      Z
Judge_A        5   35.00      33.1   1.69
Judge_B        6   32.00      33.8   2.01
Judge_C        8   29.50      27.7   0.97
Judge_D        2   27.00      24.0   0.05
Judge_E        7   28.00      26.1   0.55
Judge_F        9   27.00      22.4  -0.28
Trial_Judge    9   15.00       6.6  -4.21
Overall       46              23.5

H = 21.40  DF = 6  P = 0.002
H = 21.45  DF = 6  P = 0.002  (adjusted for ties)

* NOTE * One or more small samples
```

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 overall data 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. It is less powerful than Kruskal-Wallis or ANOVA, but it is robust to distributional considerations. That is, it is a true test of the medians being the same rather than a test of whether the distributions are the same.

In Minitab, Mood's Median test is under Stat > Nonparametrics > Mood's Median Test... We will discuss the following output in class.

```
Mood Median Test: percent_women versus judge

Mood median test for percent_women
Chi-Square = 14.16    DF = 6    P = 0.028

                                        Individual 95.0% CIs
judge        N<=  N>  Median  Q3-Q1   --+---------+---------+---------+----
Judge_A        1   4    35.0   22.0        (--------------*------------)
Judge_B        1   5    32.0    9.8                  (---*-------)
Judge_C        3   5    29.5    8.5               (----*--)
Judge_D        1   1    27.0    6.0               (--*-)
Judge_E        4   3    28.0   16.0            (------*-------)
Judge_F        5   4    27.0    8.5              (----*-)
Trial_Judge    9   0    15.0    6.5   (----*-)
                                      --+---------+---------+---------+----
                                       12        24        36        48

Overall median = 28.0
* NOTE * Levels with < 6 observations have confidence < 95.0%
```

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:

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

The following macro for conducting Welch's ANOVA is available from the WEB site for this course. To use it, your labels must be in the first column of the Minitab worksheet and your data must be in the second column. The other columns should be empty.

```
GMACRO
Welch

# Lables MUST be in C1
# Data MUST be in C2

Name k1 "I"
Let I = 1
Let k3 = N(c1) - 1
Do k2 = 1 : k3
If c1(k2) ~= c1(k2 + 1)
Let I = I +1
ENDIf
EndDo

Name c3 'Mean' c4 "ByVar" c5 "Variance" c6 "N"
Statistics C2;
By C1;
Mean 'Mean';
GValues 'ByVar';
Variance 'Variance';
N 'N'.

Name c7 'weights'
Let 'weights' = 'N'/'Variance'

Name c8 'WeightedMean'
Let k3 = I
Do k2 = 1:k3
Let c8(k2) = sum('weights'*'Mean')/(sum('weights'))
EndDo

Name c9 'F'
Name c10 'numerator df'
```

```
Name c11 'denominator df'

Let 'F' = sum('weights'*('Mean' - 'WeightedMean')**2)/(I - 1)
Let 'F' = 'F'/( 1 + (2*(I-2)/(I**2 - 1))*sum( (1/('N' -1))*(1 - 'weights'/sum('weights'))**2

Let c10 = I -1
Let c11 = (I**2 -1)/(3* sum( (1/('N' -1))*(1 - 'weights'/sum('weights'))**2))

Name c12 'p-value'
Let k2 = I -1
Let k3 = Floor(c11(1), 0)
CDF 'F' 'p-value';
F k2 k3.

Let c12 = 1 - c12

ENDMACRO
```

References and Readings

[1] Hans Zeisel. Dr. Spock and the case of the vanishing women jurors. *The University of Chicago Law Review*, 37:1–18, 1969.

Exercises for Lecture 12

1. –
2. –