Lecture 11

One-way ANOVA examples

In this lecture, we will discuss the following application of one-way ANOVA: In 1967, Dr. Benjamin Spock and others were put on trial in a U.S. District court in Massachusetts for conspiracy to violate the Military Service Act of 1967. Dr. Spock is that famous author of child care books and was popular with women. When a group of potential jurors was selected from the local residents (the group of potential jurors is called a *venire*), only 9% were women. Since approximately half the residents were women, this seemed surprising. However, due to child care issues among other things, women may not make up half of a typical venire. A comparison was made to the venires of six other district court judges in order to determine if the trial judge (or, rather, his clerk) in Spock's case seemed to have biased the number of women on the venire. It was generally assumed that since women were more opposed to the war than men were and were also familiar with Dr. Spock's books, that a female juror would be less likely to convict him than a male juror would be. The data, as give in the Law Review article by Hans Zeisel, is given in the table that follows:

Trial Judge	Judge A	Judge B	Judge C	Judge D	Judge E	Judge F
16	40	36	34	24	33	22
18	30	32	30	30	36	21
14	16	32	32		28	31
6	35	27	29		20	27
18	50	29	24		18	17
15		45	28		22	29
15			20		40	26
9			35			29
24						34

Since the design is unbalanced and the sample sizes are small, we need to check the normality and equal variance assumptions carefully. Checking normality in each sample separately is not the best procedure. All seven data sets could be normal and yet conducting 7 individual tests for normality may lead us to suspect one is not erroneously because we are conducting multiple tests. Think of it this way: if each of the 7 samples is normal then, with each data set, we have something like a $p = 5\%$ chance of rejecting the null hypothesis that the data is normal erroneously. We conduct 7 independent tests, the probability of making no mistakes (not rejecting any of them) is $(1-p)^7 = (0.95)^7 = 70\%$. In other words, we have a 30% chance of rejecting normality for at least one sample even if all the samples were in fact normal.

A better way is to test for normality on the residuals under the full model. That is, look at the difference of each data value from the sample average for its population. Those differences should be normally distributed with mean zero and the same variance for every population. Then test normality

on those residuals once. Of course, for the pooled residuals to be normal, we need that the variances be equal across the groups. We can test for that first.

The equal variances assumption is tested via the multiple group analog of Levene's test. That is, the absolute difference from the sample median is taken within each population. A one-way ANOVA is conducted on the absolute deviations from the median to see if they could have equal means. If the mean absolute deviations are equal (p-value ¿ 0.10), then the original data effectively meet the equal variances requirement for a standard one-way ANOVA.

These tests are under Stat > ANOVA. We can store residuals when running the ANOVA and test for normality afterwards under Stat > Basic Statistics. We can test for equal variances first using the menu Stat > ANOVA > Test for Equal Variances... We will do this in class.

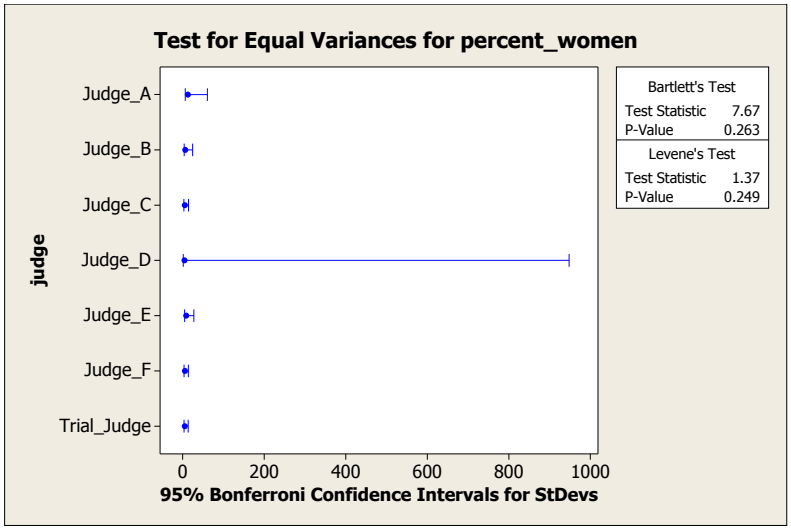

Figure 11.1: Test of equal variances using all of the judges

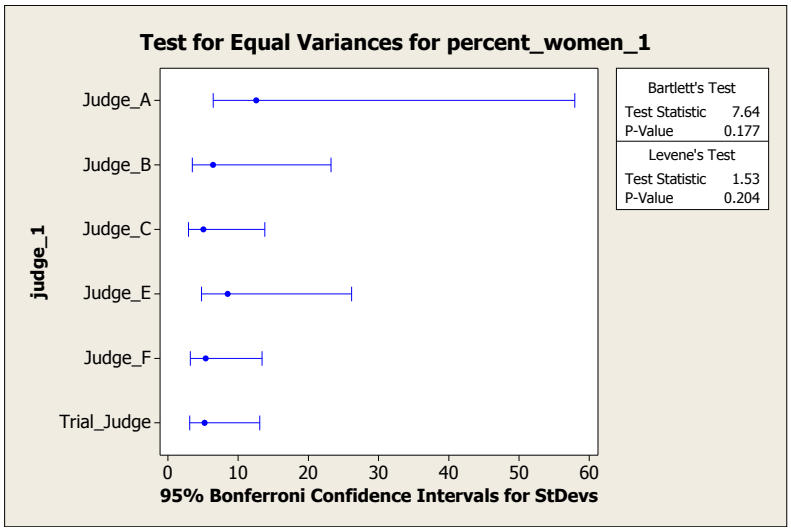

Figure 11.2: Test of equal variances excluding Judge D who has only two observations

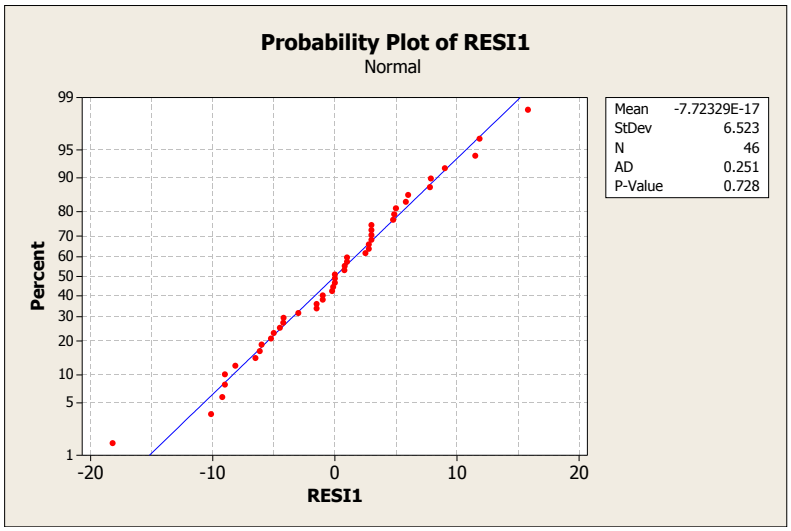

Figure 11.3: Normality plot and test on the residuals of the full model: differences in the data from the particular judge's mean.

Also in class, we will go over the output from the ANOVA and figure out what the various terms indicate.

```
One-way ANOVA: percent_women versus judge

Source  DF      SS     MS     F      P
judge    6  1849.7  308.3  6.28  0.000
Error   39  1914.7   49.1
Total   45  3764.4

S = 7.007   R-Sq = 49.14%   R-Sq(adj) = 41.31%


                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level         N    Mean   StDev  -------+---------+---------+---------+--
Judge_A       5  34.200  12.578                        (-------*-------)
Judge_B       6  33.500   6.411                        (------*------)
Judge_C       8  29.000   5.043                  (-----*------)
Judge_D       2  27.000   4.243          (------------*-----------)
Judge_E       7  28.143   8.494                 (------*------)
Judge_F       9  26.222   5.357                (-----*-----)
Trial_Judge   9  15.000   5.220  (-----*-----)
                                  -------+---------+---------+---------+--
                                      16.0      24.0      32.0      40.0

Pooled StDev = 7.007
```

A write-up of what we see might look something like:

A test for whether the percentage of women on the venires of these seven judges were equal or not gave strong statistical evidence that they were not all equal (based on an ANOVA with $F_{6,39} = 6.28$, p-value $< .001$.)

We will also ask in class if we can pool the data from the six other district court judges. an ANOVA of those judges alone will suggest that we can. Then we can compare the pooled data to the data from the trial judge in Spock's case. We will see that a one-way ANOVA for a two samples is equivalent to a pooled-variance two-sample t-test. (In fact, the ANOVA F-statistic is just the square of the t-statistic).

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 11

1. –

2. –