Tutorial: ANOVA in Stata

In this example, we will use data from the California Health Interview Survey (CHIS). From their website (http://www.chis.ucla.edu): CHIS is the nation's largest state health survey. Conducted every two years on a wide range of health topics, CHIS data gives a detailed picture of the health and health care needs of California's large and diverse population. CHIS is conducted by the UCLA Center for Health Policy Research in collaboration with many public agencies and private organizations.

In 2009, CHIS surveyed more than 47,000 adults, more than 12,000 teens and children and more than 49,000 households. We will use a sample of 500 adults for this lab (`CHISANOVA.dta`). Suppose we are interested in the relationship between number of hours worked (per week) and health, as measured by BMI. Would we expect those who worked longer hours to be healthier than those who worked shorter hours, or vice versa? Number of hours worked per week is divided into 5 categories: 0-10, 10-25, 25-35, 35-45, 45+.

1. How many people are in each category?
2. We now wish to run an ANOVA. Are the assumptions for ANOVA met?
3. What are the null and alternative hypotheses for this test?
4. Perform the hypothesis test at the $\alpha = 0.05$ level.

   Conduct a oneway ANOVA in Stata using the `oneway` command:

```
. oneway bmi work_cat, tabulate

           |           Summary of bmi
   work_cat |        Mean   Std. Dev.       Freq.
------------+------------------------------------
       0-10 |   26.431579   5.9410147          38
      10-25 |   26.429189   5.7075504          74
      25-35 |     24.3495   4.1477871          60
      35-45 |   27.128351    5.647101         188
        45+ |   27.854928   6.1797228         140
------------+------------------------------------
      Total |     26.8419   5.7540637         500

                       Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      550.823688      4   137.705922      4.27     0.0021
 Within groups      15970.6916    495   32.2640234
------------------------------------------------------------------------
    Total           16521.5153    499   33.1092491

Bartlett's test for equal variances:  chi2(4) =  11.7543  Prob>chi2 = 0.019
```

You may also use the following drop-down menus to access the `oneway` command: Statistics / Linear models and related / ANOVA/MANOVA / One-way ANOVA.

What are:

(a) your test statistic,

(b) the degrees of freedom,

(c) the p-value,

(d) your decision, and

(e) your interpretation?

5. We have rejected the null hypothesis, thus we have evidence that at least one pair of means are not equal. Perform all possible pairwise comparisons using the Bonferroni correction.

6. Which pairs of means are significantly different?

7. A colleague of yours, who has the same dataset, calculates the means for each work category. After looking at these means he takes the group with the largest mean (45+) and the group with the smallest mean (25-35) and performs a t-test (without a Bonferroni correction). He tells you that since he only did one test, he does not need to correct for multiple comparisons and that his method is valid. Do you agree? Why or why not?