STAT 3504
Assignment #2

DUE: Wed. Jan. 27

1. Given that, by definition, SSTO = $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{..})^2$ and SSTR = $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\overline{Y}_{i.} - \overline{Y}_{..})^2$, prove that alternative (computational) forms of SSTO and SSTR are given by

$$SSTO = \sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{..}^2}{n_T} \quad \text{and} \quad SSTR = \sum_{i=1}^{r} \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{n_T}$$

2. **To be done by hand**

#7 on assign 1 cont.: In a study to determine whether students in B.A., B.Sc, and B.B.A. programs earn different amounts in summer jobs, a random sample of 5 students from each of these programs was asked to report what they earned the previous summer. The results ($1000) are given below. Assume the fixed effects, single factor ANOVA model $Y_{ij} = m_i + e_{ij}$ is appropriate.

	Unit j				
factor level i	1	2	3	4	5
B.A.	3.3	2.5	4.6	5.4	3.9
B.Sc.	3.9	5.1	3.9	6.2	4.8
B.B.A.	4.0	6.2	6.3	5.9	6.4

a) In question #7 on Assignment 1 you found the following residuals and fitted values, and then plotted the residuals against the fitted values to look for possible assumption violations.

$\hat{m}_1 = \overline{Y}_{1.} = 3.94$ $\hat{m}_2 = \overline{Y}_{2.} = 4.78$ $\hat{m}_3 = \overline{Y}_{3.} = 5.76$

where B.A. = 1, B.Sc. = 2, B.B.A. = 3

B.A.: $e_{11} = -0.64$, $e_{12} = -1.44$, $e_{13} = 0.66$, $e_{14} = 1.46$, $e_{15} = -0.04$
B.Sc.: $e_{21} = -0.88$, $e_{22} = 0.32$, $e_{23} = -0.88$, $e_{24} = 1.42$, $e_{25} = 0.02$
B.B.A.: $e_{31} = -1.76$, $e_{32} = 0.44$, $e_{33} = 0.54$, $e_{34} = 0.14$, $e_{35} = 0.64$

b) Now use the Brown-Forsythe test to test for unequal variances. Use $a = 0.05$.

c) Considering ALL 15 observations together, what is the value of the second largest residual? What is its rank? What is its expected value under normality? What is the third smallest residual? What is its rank? What is its expected value under normality?

OVER

Use SAS to help you do the following question. Hand in a copy of BOTH code and output.

Remember to put a FOOTNOTE statement with you name and student number at the beginning of your program. Remember to ALWAYS SAVE your program.

SAVE YOUR OUTPUT for this question as it will be needed in the next assignment.

3. For problem #18.17 in your text the data is in the data disk that comes with your text under the file name **ch18pr17.dat** It is also in **z:\stat3504a\ch18pr17.dat**

   column 1 = response, column 2 = factor level (winding speeds),
   column 3 = observation number

   In an experiment to study the effect of the speed of winding thread (1: slow, 2: normal, 3:fast, 4:maximum) onto 75-yard spools, 16 runs of 10,000 spools each were made at each of the 4 winding speeds. The response variable is the number of thread breaks during the production run. The results (in time order) are as given in the data set.

   a) What is the study unit? What is the factor? What are the factor levels
   b) Assuming single factor fixed levels Anova Model 16.2, give the **complete** model specification (including assumptions).
   c) Also have the fitted value, median, **standard deviation**, and **variance** for each factor level printed out. These will needed for Assignment 3.
   d) Plot the residuals vs the fitted values. What is indicated by this plot? Is the result here consistent with that of part (b)? Why or why not?
   e) Use the Brown-Forsythe test to test for inequality of treatment variances. Use a = .05. What is the p-value of the test? Are your results consistent with your diagnosis in part (d)?
   f) Obtain normal probability plots of the residuals for each treatment (factor level). Do there seem to be any obvious violation of the normality assumptions? Explain.
   g) For each treatment, obtain the correlation coefficient between the ordered residuals and the expected values under normality. Use this to test whether there evidence of non-normality? Use a = .05.