Statistics 514: Problem Set No. 3

Thursday, September 27/Session 12

1. In 1968 Dr. Benjamin Spock was tried in U. S. District Court of Boston on charges of conspiring to violate the Selective Service Act. The defense challenged the method by which jurors were selected for this case, claiming that women were underrepresented. In fact, they argued, this district judge had a history of jury pools in which women were underrepresented. This argument was made by comparing this judge's pools (% women) with those of six other judges.

Spock's	Other District Judges					
Judge	A	B	C	D	E	F
6.4	16.8	27.0	21.0	24.3	17.7	16.5
8.7	30.8	28.9	23.4	29.7	19.7	20.7
13.3	33.6	32.0	27.5		21.5	23.5
13.6	40.5	32.7	27.5		27.9	26.4
15.0	48.9	35.5	30.5		34.8	26.7
17.7		45.6	31.9		40.2	29.5
18.6			32.5			29.8
23.1			33.8			31.9
15.2			33.8			36.2

   (a) Test the hypothesis that the mean percentage is the same among the other six judges using ANOVA and the Kruskal-Wallis test ($\alpha = 0.05$). *Note*: A data set containing only judges 2-7 can be created using the following commands:

```
data newname;
  set oldname;
  if judge = 1 then delete;
```

   (b) Test the hypothesis that the mean percentage for Spock's judge is lower than the others, using both a t-test and the Kruskal-Wallis test ($\alpha = 0.05$). In this case, there are only two groups. *Note*: This data set can be created using the following commands:

```
data newname1;
  set oldname;
  if judge > 1 then judge = 2;
```

2. Four different designs for a digital computer circuit are being studied to compare the amount of defects. The following data has been given (`defects.dat` on the class website):

```
design defect
1   7
1   2
1   4
1   7
1   2
2   11
2   6
2   9
```

```
2   7
2   5
3   16
3   13
3   11
3   13
3   13
4   5
4   5
4   4
4   6
4   6
```

(a) Is the number of defects present the same for all four designs ($\alpha = 0.05$)?

(b) Analyze the residuals from this experiment. In particular, what do you think about the normality assumption? Can you use any formal tests to support your conclusion?

(c) Use the Kruskal-Wallis test for the experiment and compare the results in (a).

3. To study the effects of pesticides on birds, an experimenter randomly (and equally) allocates $N = 65$ chicks to five diets (a control and four with different pesticide included). After a month, each chick's calcium content (mg) in 1 cm length of bone was measured resulting in the following:

		Pesticide			
	Control	1	2	3	4
Mean	11.54	11.00	11.42	11.44	11.28
St. Dev.	0.27	0.47	0.31	0.42	0.31

Construct the ANOVA table (i.e., compute the between and within SS) and test if there appears to be any differences in means (use $\alpha = 0.01$).

4. An experiment is conducted to study the effect of hormones injected into test rats. There are two distinct hormones – A, B – each with two distinct levels. For purposes here, we will consider this to be four different treatments labeled $\{A, a, B, b\}$. Each treatment is applied to six rats with the response being the amount of glycogen (in mg) in the liver.

Treatment	Responses					
A	106	101	120	86	132	97
a	51	98	85	50	111	72
B	103	84	100	83	110	91
b	50	66	61	72	85	60

Three contrasts are of interest. They are

Comparison	A	a	B	b
Hormone A vs Hormone B	1	1	-1	-1
Low level vs High level	1	-1	1	-1
Equivalence of level effect	1	-1	-1	1

(a) Compute the one-way ANOVA (using `proc glm`) to see if there are any treatment differences.

(b) Are these contrasts orthogonal? Why or why not?

(c) Compute the single degree of freedom sum of squares to test each null hypothesis. Interpret the results. (Note: Be careful when using a character string variable to denote the treatment levels. The order of the treatments that SAS uses in the `contrast` statement is different from A, a, B, b.)

5. Assume that a psychologist has performed a study to compare four different treatments for alleviating agoraphobia. Three subjects have been randomly assigned to each of four types of therapy: rational-emotive (R-E), psychoanalytic (P), client-centered (C-C), and behavioral (B). The following posttest scores were obtained on a fear scale, where higher scores indicate more severe phobia:

R-E	P	C-C	B
2	10	4	8
4	12	6	10
6	14	8	12

(a) Carry out the model comparison necessary to test whether there is a statistically significant difference between the means of the four groups. State the models, estimate their parameters, calculate the predicted scores and errors for each individual subject, and determine the value of F and its significance.

(b) Calculate the t-value for comparing for each pair of means. You should have six such t values. Note that with equal n,

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{(s_1^2 + s_2^2)/n}}$$

Hint: there is a peculiar relationship among the four s_j^2 values for these data. This should simplify your task considerably.

(c) Square each of the t-values you calculated in the previous question. Do you see any relationship between these six t^2 values and the F value you calculated in the first question?

6. A graduate student has conducted a four-group study in which he tested the following three planned comparisons:

	1	2	3	4
C_1	1	-1	0	0
C_2	0.5	0.5	-1	0
C_3	1/3	1/3	1/3	-1

The sums of squares for the three comparisons are 75, 175, and 125, respectively. The value of MSE equals 25, and there were 11 participants in each group. The student's advisor wonders whether the overall F-test of $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ would be statistically significant for these data. Can you help her?

(a) Is it possible to perform the test of the overall null hypothesis from the available information? If so, is the test significant? If it is not possible, explain why not.

(b) Find the observed F value for each of the planned comparisons tested by the student. Which, if any, are statistically significant with an experimentwise error rate of 0.05?

(c) What relationship, if any, is there between the single observed F value of Problem 6a and the three observed F-values of Problem 6b?

7. A manufacturer suspects that the batches of raw material furnished by a supplier differ significantly in calcium content. There are a large number of batches currently in the warehouse. Five of these are randomly selected for study. A chemist makes five determinations on each batch and obtains the following data:

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38

(a) Is there a significant variation in calcium content from batch to batch? Use $\alpha = 0.05$.

(b) Estimate the components of variance.

(c) Find a 95 percent confidence interval for $\sigma_\tau^2/(\sigma_\tau^2 + \sigma^2)$.

(d) Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

8. Five tire types (brand/model combinations like Goodyear/Arriva) in the size 175/80R-13 are chosen at random from those available in a metropolitan area, and six tires of each type are taken at random from warehouses. The tires are placed (in random order) on a machine that will test tread durability and report a response in thousands of miles. The data follow:

Brand	Miles					
1	55	56	59	55	60	57
2	39	42	43	41	41	42
3	39	41	43	40	43	43
4	44	44	42	39	40	43
5	46	42	45	42	42	44

Compute a 99% confidence interval for the ratio of type-to-type variability to tire-within-type variability (σ_α^2/σ^2). Do you believe that this interval actually has 99% coverage? Explain.

9. A researcher has conducted an experiment with six independent groups of 12 participants each. Although the overall F-test was nonsignificant, he decided to use Scheffé's method of multiple comparisons. He claims that his calculations revealed that the average of the first three groups was significantly different from that of the last three. How would you respond to this claim?

10. For Problem 3.1 in Montgomery, $MS_E = 12825.7$. What would the minimum difference have to be between any two means for you to conclude they are significantly different if

(a) You performed the LSD comparison procedure?

(b) You performed Tukey's multiple comparison procedure (Table VII)?

(c) You performed Scheffé's procedure?

(d) Explain the relationship between power and the minimum difference, and state which of the four is the most powerful and least powerful comparison procedure.

11. A psychologist has tested 10 independent hypotheses. She has decided to control the false discovery rate for this set of hypotheses at 0.01 . The ten p-values she has obtained are as follows: 0.04, 0.15, 0.02, 0.31, 0.06, 0.62, 0.01, 0.03, 0.46, 0.08 . Which, if any, hypotheses can she reject controlling the FDR at 0.01? Which, if any, hypotheses can she reject controlling the experimentwise error rate at 0.01? Which, if any, hypotheses can she reject controlling the strong experimentwise error rate at 0.01?

12. In an experiment with five groups and 25 degrees of freedom for error, for what number of contrasts is the Bonferroni procedure more powerful than the Scheffé procedure?