Computational Public Health Statistics Midterm, S05

- ✎ **Name** in usual location (back of last page of exam) *and* blue book, please.
- ✎ Hand in your **Lab Workbook** and **Procedure Notebook** at the beginning of the exam.
- ✎ Complete **Part A** on WebCT (closed book).
- ✎ **Submit your WebCT answers.**
- ✎ Pick up your Procedure Notebook and start **Part B**.
- ✎ **Time limit**: 2 hours.
- ✎ **Hand in your blue book and exam**.

Part B - *Answer in your blue book. Please make sure you include the number and letter of each problem.*

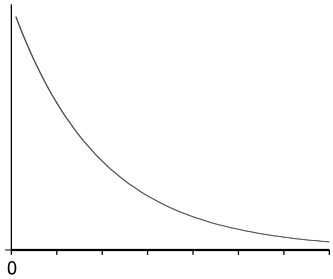
1. *Two groups.* Data from a study using a Liekert scale (5 = strongly agree, 4 = agree, etc.) are:

Group 1:	1	3	4	4	5	5	5	5	5	5
Group 2:	3	3	3	4	4	4	5	5	5	

a. Plot data as side-by-side stemplots on a *common stem* (as in Lab 1, Part A).
b. Describe the *shape* of the Group 1 distribution.
c. Describe the *shape* of the Group 2 distribution.
d. Identify *outliers*, if any. (You can look at the data for this purpopse. This saves time and is no less accurate than more formal methods.)
e. Determine the *median* of Group 1.
f. Determine the *median* of Group 2.

2. *F ratio test*. An F ratio test of two groups with $n_1 = n_2 = 10$ derives an $F_{stat} = 3.99$.
a. Write the null hypothesis tested by the F ratio test.
b. Replicate the curve below in your blue book, including scaling of the axis. Mark the location of the F_{stat} on the curve with proper shading of the p value region.
c. Identify the precise value of the $\alpha = .05$ landmark, and place it at its proper location on your curve.
d. Is the p value less than .05?
e. Do you conclude homoscedasticity or heteroscedasticity?

This is what the F distribution should look like, more or less:



3. *Mean and standard deviation.* Intelligence quotients in 5 monkeys are:

88 100 102 104 106

a. Calculate the *mean* of this sample. (Show work.)
b. Calculate the *sum of squares*. (Show work.)
c. Calculate the *standard deviation*. (Show work.)
d. *Report* the standard deviation with the *proper number of decimal* places.

4. *More monkey data*. A second group of 6 monkeys has a mean intelligence quotient of 113.0 with a standard deviation of 6.8. This group is compared to the prior group of monkeys. SPSS output from the independent *t* test procedure is shown below. *Refer to this output for the next set of questions.*
a. *Show how to calculate* the standard error of the mean difference assuming equal variance. (Plug numbers into formula; standard error shown below.)
b. In plain language, what does the *standard error of the mean difference* quantify?
c. What null hypothesis is tested by the *t* statistic?
d. Report the *t* statistic, its df, and *p* value for the equal variance *t* test?
e. Do the *means* differ significantly (assuming $\alpha = .05$).
f. Do you have evidence that population *variances* differ? Justify your response, supporting your conclusion with statistical results. (One or two sentence response. Please write coherently and legibly.)

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
MONKEYIQ	1	5	100.00	7.071	3.162
	2	6	113.00	6.782	2.769

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
MONKEYIQ	Equal variances assumed	.002	.962	-3.106	9	.013	-13.00	4.186
	Equal variances not assumed			-3.093	8.492	.014	-13.00	4.203

5. *Bud's fascination with monkeys continues.* A third group of monkeys enters the mix. This third group has only 4 monkeys. Their mean IQ is 105.0 (standard deviation = 4.8). You decide to do an ANOVA comparing all three groups. Descriptive statistics for all three groups is:

Group	*n*	Mean	SD
1	5	100.0	7.1
2	6	113.0	6.8
3	4	105.0	4.8
All	15	106.5	8.3

a. List the null hypothesis for the ANOVA test.
b. List the alternative hypothesis for the ANOVA.

Here's a *partial ANOVA table (provided to save time),* showing sums of squares:

	Sum of Squares	df	MS
Between	473.733		
Within	498.000		

Calculate the following:

c. df_B =
d. s^2_B =
e. df_w =
f. s^2_w =
g. F_{stat} =
h. p =
(You may use your *F* table or StaTable to determine the p value.)
i. Assuming $\alpha = .05$, do you retain or reject the null hypothesis?
j. Is the difference significant?
k. When doing post hoc comparisons, why (in general) would you use Bonferroni's (or some other multiple comparisons) correction?
l. Output from the post hoc comparisons using Bonferrroni method is shown below. Interpret the post hoc comparisons output.

Multiple Comparisons

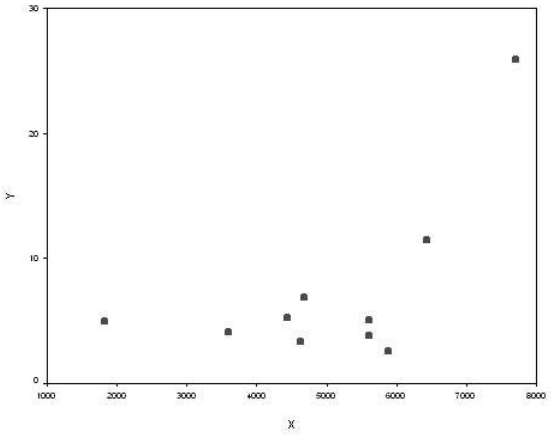
Dependent Variable: MONKEYIQ
Bonferroni

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.
1	2	-13.00	3.901	.018
	3	-5.00	4.321	.809
2	1	13.00	3.901	.018
	3	8.00	4.158	.235
3	1	5.00	4.321	.809
	2	-8.00	4.158	.235

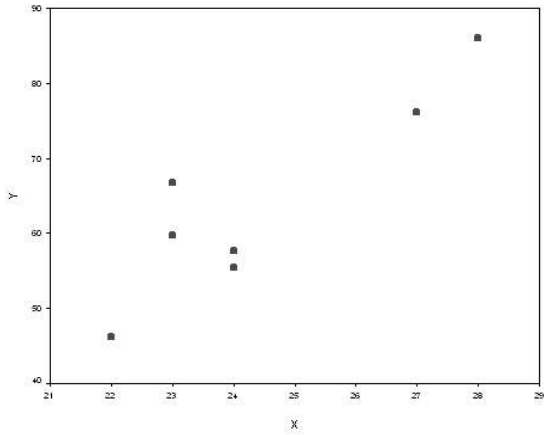
6. CORRELATION AND REGRESSION PROBLEMS

Below are two scatter plots.

Scatterplot 1



Scatterplot 2



a. Describe the data in scatter plot 1. Consider outliers, linearity, and direction of correlation (if any). {Use judgment; do not "squint" too hard.}
b. Are correlational techniques warranted for scatter plot 1? (Assume no data manipulation.)
c. Describe the data in scatter plot 2. {Same advice.}
d. Are correlational techniques warranted for scatter plot 2? (Assume no data manipulation.)

Here is the **SPSS regression output** from a data set. *Assume* linearity and other distributional assumptions.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-69.117	29.035		-2.380	.063
	X	5.448	1.184	.899	4.600	.006

a. Dependent Variable: Y

e. List *distributional* assumptions required for inference.
f. List *validity* assumptions needed for inference?
g. *List* and *interpret* the slope estimate.
h. There are 7 data points for these data. Calculate a 95% confidence interval for the slope.
i. Interpret your 95% confidence interval.