Name: ____________________ Fall, 2016

Applied Statistics Comprehensive Examination

You must give complete explanations to receive full credit.

1. (20 points) An official of a plastics company claimed that the company employed 30% white women, 5% minority women, 50% white men and 15% minority men. To test this claim, an affirmative action committee randomly sampled 150 employees and found 40 white females, 15 minority females, 80 white males and 15 minority males.

(a) (10 points) Test the official's claim at a 5% level of significance.

(b) (10 points) Estimate the P-value and give an interpretation.

2. (25 points) A, B, and C are three explanatory variables in a multiple linear regression with $n = 28$ cases. The following table shows the residual sums of squares and degrees of freedom for all models.

Model variables	Residual sum of squares	Degrees of freedom
None	8,100	27
A	6,240	26
B	5,980	26
C	6,760	26
A,B	5,500	25
A,C	5,250	25
B,C	5,750	25
A,B,C	5,160	24

(a) (5 points) Based on the criterion of the estimate of σ^2, which is the best model in the table? Explain why and report the associated estimate of σ^2.

(b) (5 points) Based on the criterion of adjusted R^2, which is the best model in the table? Explain why and report the associated adjusted R^2.

(c) (5 points) Explain why the models selected in parts (a) and (b) should agree.

(d) (10 points) Use forward selection to determine a best model. At each step, perform an extra-sum-of-squares F-test to see whether the additional variable is significant ($\alpha = 0.05$) and state which variable enters the model at that step.

3. (20 points) A calculus teacher ran an all-day Saturday seminar to try to improve the algebra skills of some of her weaker students. To assess the effectiveness of the seminar, she gave the six students an algebra test both before and after the seminar, making sure that the tests were different, but of comparable difficulty. The results are given in the table below. Using $\alpha = 0.05$, test for evidence that the seminar improved the algebra skills of the students. State any necessary assumptions, and suggest how you might assess them.

Student	Before	After
1	50	80
2	40	60
3	60	80
4	50	60
5	60	50
6	50	70

4. (40 points) Students were given different drug treatments before studying for their exams. Some were given a memory drug, some a placebo drug, and some no treatment. The percent improvement in their exam scores (%) over their initial baseline exams are shown below for the three treatment groups:

Statistic	Memory Drug	Placebo	No Treatment
	70	37	3
	77	43	10
	83	50	17
	90	57	23
	97	63	30
Sum:	417	250	83
Mean:	83.4	50.0	16.6
Variance:	112.3	109.0	112.3

(a) (5 points) State the effects model and define all terms in the model. State the assumptions.

(b) (5 points) Create a meaningful plot of the data and comment on the results.

(c) (10 points) Carry out an ANOVA to test the null hypothesis that all three treatments have the same mean.

(d) (10 points) Use the Student-Newman-Keuls procedure for multiple comparisons to determine which treatment maximizes the mean improvement.

(e) (5 points) Estimate the difference between the mean percent improvement for the memory drug versus the other two treatments combined using a 95% confidence interval.

(f) (5 points) Briefly discuss how you would assess whether the model assumptions are met.

5. (20 points) After losing a lawsuit for under-filling cans of tuna, a local tuna company has revamped its quality control process for ensuring that tuna cans contain at least the required 5.0 ounces of meat on average. After each large batch of cans is produced, six cans are selected at random, and the meat in each can is weighed. If the average amount of meat in the six cans is below 4.9 ounces, then the entire batch is rejected and the equipment is carefully reset. Suppose that we think of this procedure as a test of $H_0 : \mu = 5$ against $H_a : \mu < 5$, where μ is the population mean amount of meat per can. You may assume that the amount of meat put in each can is normally distributed.

(a) (10 points) If the standard deviation is 0.1 ounces, what is the α level for the test?

(b) (10 points) If the standard deviation is 0.1 ounces and the population mean amount of meat per can is 4.8 ounces, what is the power of the test?

6. (10 points) Suppose you are designing a study in order to determine whether there is a difference in mean response between the various levels of the main factor (Factor A). You believe that Factor B will contribute considerable variability to the response and you are thinking about incorporating it into the design. Discuss the advantages and/or disadvantages if you incorporate Factor B into the design as a blocking variable and:

(a) (5 points) Factor B contributes significant variability to the response.

(b) (5 points) Factor B does not contributes significant variability to the response.

7. (20 points) A soft drink dispensing machine is said to be out of control if the standard deviation of the contents exceeds 1.15 deciliters. A random sample of 25 drinks from this machine is obtained and yields a standard deviation of 2.03 deciliters. Using $\alpha = 0.05$, determine if these data indicate that the machine is out of control. State any necessary assumptions for your hypothesis test, and suggest how you might assess them.

8. (20 points) A railroad company used two types of wheel mounts that differ in the way they handle track irregularities: type A (spring equalized) and type B (frame equalized). After a one year period, it was found that 20 out of 150 cars using type A wheel mounts needed service and 18 out of 180 cars using type B wheel mounts needed service. Let π_A and π_B be the population proportions of cars that needed service.

(a) (10 points) Find a 95% confidence interval for the difference $\pi_A - \pi_B$.

(b) (10 points) Use this interval to decide if there is a difference between π_A and π_B, and justify your answer.

9. (25 points) In a study of the effects of predatory intertidal crab species on snail populations, researchers investigated the relationship between closing forces and propodus heights of the claws for three species of crabs. The total sample size was 38. The table below shows output from the least squares fit to the separate-lines model for the regression of log force (y) on log height (x_1). The regression model for log force was

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3,$$

where $y = \log(\text{force})$, $x_1 = \log(\text{height})$, $x_2 = \begin{cases} 1, & \text{if } \textit{species 2}; \\ 0, & \text{otherwise.} \end{cases}$, and $x_3 = \begin{cases} 1, & \text{if } \textit{species 3}; \\ 0, & \text{otherwise.} \end{cases}$

Variable	Estimate	SE	t-stat	p-value
Constant	0.5191	1.0000	0.5191	0.6073
x_1	0.4083	.4868	0.8387	0.4079
x_2	-4.2992	1.5283	-2.8131	0.0083
x_3	-2.2992	1.7606	-1.4123	0.1675
$x_1 \times x_2$	2.5653	0.7354	3.4885	0.0014
$x_1 \times x_3$	1.6601	0.7889	2.1043	0.0433

(a) (10 points) State the assumptions that are necessary for this multiple regression. Using the following plots, comment on whether the assumptions appear to be met. If you cannot assess an assumption using these plots, name an alternative assessment method.

(b) (5 points) Conduct a test of the hypothesis that the slope in the regression of log force on log height is the same for species 2 as it is for species 1. Use $\alpha = 0.05$.

(c) (10 points) Find a 95% confidence interval for the amount by which the slope for species 3 exceeds the slope for species 1.

Fit Diagnostics for y