Name _____ Fall 2015

Applied Statistics Comprehensive Examination

- Calculators are permitted on this part of the examination.
- When you are asked to construct a confidence interval, always interpret the interval in terms of the problem.
- When you are asked to perform a hypothesis test, always write down the null and alternative hypotheses and write the conclusions in terms of the problem.

Regression – Linear Models

1. (25 pts) To study the relationship between the number of employed people (in million) (x) and the gross domestic product (GDP, in billions of US dollars) (y), a researcher has collected data from 12 countries. Use the following information to answer the questions:

$$\sum_{i=1}^{12} x_i = 581 \quad \sum_{i=1}^{12} (x_i - \bar{x})^2 = 377 \quad \sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y}) = 64$$

$$\sum_{i=1}^{12} y_i = 53 \quad \sum_{i=1}^{12} y_i^2 = 267$$

Source	df	SS	MS	F
Regression				
Error		22.08		
Total				

(a) (5 pts) Write out the fitted linear regression line.

(b) (5 pts) State the assumptions that are necessary for this simple linear regression.

(c) (10 pts) Complete the ANOVA table.

(d) (5 pts) What hypothesis is being tested in the above ANOVA table? State and conduct the test at the significance level of 0.05.

2. (25 pts) Suppose a realtor wants to model the appraised price (in thousands of euros) of an apartment in Vitoria, Spain, as a function of two predictors: living area (in square meters) and the apartment's energy efficiency (in two levels, high and low). Use the SAS output below to help answer the following questions.

(a) (5 pts) Conduct the global hypothesis test for the model at the significance level of 0.05.

(b) (5 pts) Calculate and interpret the value of R^2 for this model.

(c) (5 pts) Based on the fitted model in the output, write down separate prediction equations for each energy efficiency category.

(d) (10 pts) Test for equality of the slopes of the models in part (c) at the significance level of 0.05.

Multiple Regression with interaction

The GLM Procedure

Dependent Variable: price

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	739411.007	246470.336	174.26	<.0001
Error	214	302682.981	1414.406		
Corrected Total	217	1042093.988			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
area	1	682389.1004	682389.1004	482.46	<.0001
energy	1	31069.8909	31069.8909	21.97	<.0001
area*energy	1	25952.0161	25952.0161	18.35	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
area	1	400141.5499	400141.5499	282.90	<.0001
energy	1	11849.6175	11849.6175	8.38	0.0042
area*energy	1	25952.0161	25952.0161	18.35	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	22.31	B	13.135	1.70	0.0909
area	2.97	B	0.143	20.73	<.0001
energy low	73.09	B	25.250	2.89	0.0042
energy high	0.00	B	.	.	.
area*energy low	-1.21	B	0.283	-4.28	<.0001
area*energy high	0.00	B	.	.	.

3. (30 pts) Suppose a completely randomized experiment with four treatments and five observations per treatment is planned. But inadvertently, an observation is lost from the third treatment. Assume an effects model and let G be the grand total of all the observations and T_i be the total of the observations in the i^{th} treatment group.
    a. (10 pts) Write the normal equations in matrix form.
    b. (10 pts) Write the normal equations in matrix form using the "set to zero" restriction.
    c. (10 pts) Write the normal equations in matrix form using the "sum to zero" restriction.

4. -(20 pts) Consider an experiment to compare the effectiveness of two different pesticides, A and B say, both applied in two different forms: spray (A_1 and B_1) and powder (A_2 and B_2). A control treatment (i.e., no pesticide), C say, is also included in the experiment in order to establish any effectiveness of the pesticides at all. Thus, altogether there are 5 treatments (A_1, A_2, B_1, B_2 and C) and each treatment is randomly applied to *r* uniformly infested plots of land. Obtain a complete set of <u>meaningful</u> orthogonal contrasts and describe what each contrast is testing.