Name_____ Fall 2014

Applied Statistics Comprehensive Examination

Regression and Linear Models

1. (10 points) Upon examining the assumptions for a multiple linear regression model, a statistical analyst suggests that a transformation of the dependent variable and an examination of the DFBetas would be prudent. For each of these methods, identify the problem that supposedly exists in fitting the data to the model and state how this problem could be potentially fixed.

   a. (5 points) Transformation of the dependent variable.
   b. (5 points) Examination of DFBetas.

2. (15 points) Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

   where $i = 1 \ldots n$. If β_0 is known to be 5, determine the least squares estimator for the slope parameter.

3. (25 points) A recent study from the Netherlands (deWind, A., et. al., 2014) examined factors associated with transition from work to (non-disability) early retirement, defined as retirement before age 65. They selected 2,317 employees between the ages of 59 and 63 years to participate in the survey. The following results from a multiple logistic regression on whether the person retired early or not (note that additional variables that they identified were excluded from the summary of results to simplify the problem):

Variable	Parameter Estimate	Standard Error	Odds Ratio
Female (vs. Male)	-0.24	0.17	0.79
Developmental Skills [1]	-0.61	0.13	0.54
Partner's Attitude Towards Retirement:			
Positive	1.35	0.18	3.85
No Partner	0.67	0.23	1.96
Negative/Neutral (Reference)	0		

   [1] A numeric measure of how much the person actively searches for activities in their job that allow them to expand and adapt their knowledge and skills. Higher scores mean higher level of active searching.

a. (15 points) Interpret the coefficients for the following variables:
    i. Female
    ii. Partner having a POSITIVE attitude towards retirement
    iii. Developmental skills

b. (10 points) Conduct the appropriate hypothesis test on the parameter estimate for developmental skills.

4. (30 Points) A company wanted to replace the machines used to make a certain component in one of its factories. Three different brands of machines were available, so the management designed an experiment to evaluate the productivity of the machines when operated by the company's own personnel. Six employees were randomly selected to participate in the experiment, each of whom was to operate each machine two different times. The response was an overall score, which took into account the number and quality of components produced. The partial ANOVA table is given below:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square
Machine		1.800		
Person		1,200		
Machine*Person		400		
Error		36		

a. (10 Points) Write the mathematical model for this design. Be sure to describe each term in the model and state all model assumptions.
b. (10 Points) Fill in the rest of the partial ANOVA table by filling in the cells for the degrees of freedom and the mean squares. In addition, determine the expected mean squares for each term in the ANOVA table.
c. (10 Points) Test the significance of the machine effect at the 0.05 level of significance.

5. (20 Points) Suppose we are interested in the effect of two different kinds of potting soil on the number of days to germination of three varieties of carrot seed. The following table displays the data.

Soil	Variety 1	Variety 2	Variety 3
1	6	13	14
	10	15	24
	11		
2	2	32	18
	5		9
	9		12
	8		

The mean square error from an effects model which includes the main effects for soil and variety along with the interaction effect is 52.

a. (10 Points) Construct two orthogonal interaction contrasts. Estimate one of them using a 95% confidence interval.
b. (10 Points) Using one observation per cell, construct the design matrix using sum-to-zero restrictions