

Statistics 516 - Spring 2003 - Final Exam Solutions

Part I: Answer the two following questions. Five points each.

- 1) In performing a regression, ANOVA, or ANCOVA, what four assumptions must be satisfied? **The errors must be normally distributed, have mean zero, and equal variances at each treatment level, and must be independent.**
- 2) Define what is meant by the p-value (or empirical significance level) of a test. **The probability of observing a test statistic as extreme as the one observed, or more extreme, if the null-hypothesis is true.**

Part II: Answer fifteen of the following sixteen questions. Six points each.

1) Complete the model equation for this regression, identifying the parameters used.

$$y_i = \beta_0 + \beta_{dbh} + \beta_{height} + \beta_{age} + \beta_{grav} + \epsilon_i$$

y_i = weight of the i^{th} tree

β_0 = intercept

β_{dbh} = slope for dbh

β_{height} = slope for height

β_{age} = slope for age

β_{grav} = slope for grav

ϵ_i = error

2) Which of the trees had the most extreme set of x-values? What statistic did you use to tell this?

Tree 16 has a a hat diagonal of 0.3186

3) What set of these variables forms the best regression model for predicting the weight of trees? What did you use to tell this? **dbh, age, and grav has the fewest variables of all the models with C(p) less than p+1 and has the highest adjusted r^2**

4) What problems with the assumptions are apparent from the residual vs. predicted plot? Suggest a transformation that might fix this problem. **Mean of errors isn't zero and the variances increase (horn shape). Take log of y.**

5) Briefly explain why this model is balanced, factorial, with replications, and fixed effects.

It is balanced because each combination occurs exactly 5 times

Because they all occur it is factorial and because they occur more than once it has replication

Fixed effects because we care about these specific conditions.

6) The DF, SS, MS, and F for oven have been deleted. What should they be?

oven + roast + oven*roast=model, so take the differences for the SS and df... use the MSE to get the F

1 14.06596 14.06596 1.1466

7) What hypothesis is being tested by the contrast c1? **That the average gas needed for fresh is equal to the average needed for the three frozen (or formerly frozen conditions).** $\mu_{fresh} = (\mu_{24} + \mu_{12} + \mu_{froz}) / 3$

8) Is the assumption of equal variances met? **Yes. Modified Levene test p-value > 0.05 (it's 0.4042)**

9) Assume the assumption of equal slopes is met. Is there a significant difference in the 9 week weight in pigs for the different weaning stages? **Yes, p-value = 0.0004 for stage.**

10) Assume the assumption of equal slopes is met. What is the estimated standard deviation of the errors in this model for predicting the 9 week weight of pigs? **root-mse=3.392324**

11) Assume the assumption of equal slopes is met. What percent of the variation in 9 week weight in pigs that is explained by this model using wwt and stage? **R-squared=0.706675=70.6675%**

12) Is the assumption of equal slopes met for this model? **Yes. P-value for interaction is > 0.05 (it's 0.5190)**

13) Does a logistic regression seem to fit this data at $\alpha=0.05$? **Yes. Hosmer and Lemeshow p-value > 0.05 (it's 0.2027)**

14) Assuming that the logistic regression does fit this data, is the age of the party member related to their chance of survival at an $\alpha=0.05$ level? **Yes. Likelihood ratio test p-value < 0.05 (it's 0.0186)**

15) What is the predicted chance of a 15 year old surviving the journey?

$\exp(1.8183+15*(-0.0665)) / (1 + \exp(1.8183+15*(-0.0665))) = 0.6944061$

16) Why can't you trust the answer you gave in 15? **We would be extrapolating. (Note, that there is actually a 15 in the data set though... so this was a typo on the exam!!! It would work if I had asked about, say, 10.)**

