Stat 401 B/xm - Exam 2, Fall 2001

Please put your name on the back of the last page

This exam has five questions, each with multiple parts, and a bonus. The first four questions are worth from 10 to 20 points each. The last question is worth 40 points. Some questions extend over two pages.

Please write your answers in the spaces provided. If you need more room, continue on the back of the page. Partial credit is given in all problems, so please show your work.

This exam is closed book. If you need a critical value, please ask. The last page is a formula sheet. If you want a formula and you don't see it, please ask.

1. 10 pts. a) Complete the blanks in the following ANOVA table.

Source	d.f.	Sums of squares	Mean square	F statistic
Model		55.83		
Error	32			
Corr. Total	35	262.84		

b) How many observations were used in this study?

c) Could this ANOVA table be describing the fit of a simple linear regression? Why or why not?

2. 20 pts. A researcher is evaluating ways to reduce erosion from freeway roadsides by adding material to the soil along the roadside. Plots on a newly build roadside were randomly assigned to one of 5 treatments (listed below). Grass was seeded into all plots in the spring. In late summer, the amount of soil erosion (grams of soil eroded) was measured on all plots. The treatments and the sample averages were:

Treatment	average erosion	contrast coefficients
1) control, nothing added	57.2	
2) topsoil added	30.5	
3) composted sewage sludge from Cedar Rapids	25.3	
4) composted sewage sludge from Debuque	27.5	
5) composed sewage sludge from Des Moines	29.1	

These data were analyzed using 1-way Analysis of Variance.

a) The p-value from the 4 d.f. F-test that compares the equal means model to the different means model is 0.011. Carefully state an appropriate conclusion from this test.

b) One of the goals of the study is to estimate the effect of compost, compared to just adding topsoil. In the table above, please write the coefficients for the linear contrast to estimate the difference between the topsoil treatment and the average of the three sewage sludge treatments.

c) Please estimate this contrast.

d) A second goal was to see if there were any differences between the three sewage sludge treatments. This question can be answered by comparing two models. Using words or symbols (μ's with subscripts), please describe the two models:

full model:

reduced model:

3. 20 pts. A study of southern corn leaf blight infection looked at the relationship between light levels and the amount of leaf infected. 75 plants were randomly assigned to one of 5 light levels, from 0.4 to 2.0 units of light (15 plants per light level). Each plant was inoculated with the blight, and after an appropriate length of time, the percent of leaf area infected was measured on each plant. The fitted regression equation was %leaf area infected = 11.3 - 5.25*light.

   Diagnostics, including residual plots and a lack-of-fit test, indicate no problems with linearity or other assumptions. Other potentially useful information includes:
   pooled standard deviation = $\sqrt{\text{MSE}} = \hat{\sigma} = 1.2\%$.
   Mean value of light = $\overline{X} = 1.2$ units.
   Variance of light values = $\sigma_x^2 = 0.21$.
   Standard error of the predicted mean % leaf area infected at 2.4 units of light = $s_{\hat{Y}_o} = 0.81$.

   a) Please provide two sentences that interpret (or explain) the estimated slope and the intercept.

   b) The researcher wants to make conclusions about individual plants growing in 2.4 units of light. Please predict the % leaf area infected for a plant grown in 2.4 units of light.

c) Is it reasonable to use the regression line to make this prediction? Why or why not?

d) The standard error of a predicted % infection for an individual plant growing at 2.4 units of light is closest to (circle your choice):

0.46%, 0.81%, 1.2%, 1.45%, 2.01 %.

4. 10 pts. The statistical estimation and testing of a regression model (like the relationship between infection and light levels in problem 3) makes assumptions. Please list them and indicate whether each is relatively important or relatively less important. For at least two of the assumptions, please indicate a method to evaluate that assumption.

Assumption	Importance	Method for diagnosis

5. 40 pts. The following analyses are part of a study of corn response to potassium (K) fertilization experiment on the Caribbean island of Antigua. This was a randomized experiment, with potassium fertilizer treatments randomly assigned to plots at an agricultural research station. There were 10 plots at each of 5 levels of K fertilizer: 0 (control), 2.5, 5, 10, and 30 units. Each plot received the customary levels of N and P. The response is the corn yield per plot. The general goal of the experiment was to describe the yield response to addition of K fertilizer. Specific questions include:
1) Does the amount of K fertilizer affect the yield?
2) Is the response to fertilization linear?

Be sure to outline what you did (if anything) to decide on a particular approach. Note that there are many possible analyses. The enclosed SAS output includes some of them, but I couldn't include everything. If you want something else, please ask me.

The analyses in the included SAS or JMP output are:

Analysis	Response	SAS page(s)	JMP pages
Plot of obs.		1	1
1-way ANOVA	yield	2-4	
1-way ANOVA	log yield	8-10	
linear reg.	yield	5-7	
linear reg.	log yield	11-13	

Formulae

I have tried to anticipate formulae you might need, but I have not copied all formulae. If you think you need something else, don't assume you're wrong. Just ask for the formula.

$$
\begin{aligned}
\overline{X} &= \sum_{i=1}^{N} X_i/N \\
\text{Var } X &= \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2 \\
s_x &= \sqrt{\text{Var } X} \\
\text{s.e. } \overline{X} &= \frac{s_x}{\sqrt{N}} \\
\text{pooled s.d. } s_p &= \sqrt{\frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1+N_2-2}} \\
\text{pooled s.e. of } \overline{X_1} - \overline{X_1} &= s_p\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \\
\text{T ratio: } T &= \frac{\text{estimate} - \text{parameter}}{\text{s.e. of estimate}} \\
\text{estimate of a linear combination: } g &= \sum l_i \overline{X_i} \\
\text{s.e. of a linear combination: } s_g &= s_p\sqrt{\sum l_i^2/n_i} \\
\text{s.e. of slope: } s_{\beta_1} &= s_p\sqrt{\frac{1}{(n-1)s_x^2}} \\
\text{s.e. of intercept: } s_{\beta_0} &= s_p\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{(n-1)s_x^2}} \\
\text{s.e. of predicted mean at } X_o\text{: } s_{\hat{Y}} &= s_p\sqrt{\frac{1}{n} + \frac{(X_o - \overline{X})^2}{(n-1)s_x^2}} \\
\text{s.e. of predicted observation at } X_o\text{: } &= s_p\sqrt{1 + \frac{1}{n} + \frac{(X_o - \overline{X})^2}{(n-1)s_x^2}} \\
&= \sqrt{s_p^2 + s_{\hat{Y}_o}^2} = \sqrt{\hat{\sigma}^2 + s_{\hat{Y}_o}^2} \\
\text{Residual sum of squares: } \text{SS}_R &= \sum (Y_i - \hat{Y}_i)^2
\end{aligned}
$$