Statistics 516 - Spring 2003 - Practice Exam 1 Solutions

Part I:
1) The probability of observing a test statistic as extreme as the one observed (or more extreme) if the null hypothesis is true.

2) The errors must be normally distributed (at each value of the independent variable).
The errors must have mean zero (at each value of the independent variable) -or- The linear form is appropriate.
The errors must have constant variance (across the levels of the independent variable).
The errors must be independent.

3)

Source	SS	DF	MS	F	Prob>F
Regression	21752.0-19209.6=2542.4	6-5=1	2542.4	0.66	0.4529
Error	5*3841.9250=19209.6	5	3841.9250		
Total	21752.0	7-1=6			

Part II:
1) 0.0375 (Draw the picture!)

2) Any of $x^2, x^3, e^x, \sqrt{y}, \ln(y), \frac{-1}{y}$

3) We would be extrapolating; there is no data below 60, so we don't know if the assumptions would still be met at 55".

4) Either there was definitely a mistake in recording the data, or it has an extreme x-value.

Part II: Answer every part of the next three problems. Read each problem carefully, and show your work for full credit. Twenty points each.

5) 100(-0.2912)=-29.12 ... a decrease of 29.12

6) Yes, the p-value in the ANOVA table is < .0001.

7) (145.37, 246.39)

8) From the residual vs. predicted plot we see that the means of the errors are about zero at each predicted value (no wavey pattern), and the variances are about constant (no funnel), except for two outliers. From the q-q plot we see that errors are approximately normally distributed (the points follow the line). Finally, we are told the problem was randomized so the errors are independent.

9) What percentage of the variation in the number of native species is explained by the five predictor variables?
r^2=0.9428 so 94.28%

10) Carefully state what null and alternate hypotheses go with the p-value of 0.0008 in the *Type I Tests* box.
H_0: $\beta_{DistSC}=0$ given Area and DistNear are included in the model We would reject H_0 at any reasonable α.
H_A: $\beta_{DistSC}\neq 0$ given Area and DistNear are included in the model

11) List all of the possible models that are approximately unbiased and say how you determined them.
(Elev and NonNative), (Area, Elev, and NonNative), and (Area, DisctSC, Elev, NonNative) are all C_p at p+1 or less.
Four other models are close as well and I would probably include them: (DistSC, Elev, NonNative), (DistNear, Elev, NonNative), (Area, DistNear, Elev, NonNative), (DistNear, DistSC, Elev, NonNative).

12) Largest Residual or Studentized Residual... San_Salvador.