STA 242/ENV 255
M. Clyde
February 5, 1998.

Name ______________________

Score ______________________

First In–Class Exam

Notes:

1. This is a closed book exam.
2. You may use one sheet of notes.
3. Computer output is at the end of the test.

Problem 1. A biologist wished to study the effect of temperature on the growth of human amniotic cells in a tissue culture. She conducted an experiment using 4 different temperatures and taking 5 samples grown at each temperature for a total of 20 observations. She then fit a straight line regression of Number of Cells (Y) on temperature (X).

a) Complete the partial ANOVA table shown in the following output:

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	Prob > F
Model	*1*	*12.83072*	*12.83072*	*632.1*	0.0001
Error	*18*	0.36618	*.0203*		
C Total	19	13.19690			

b) Calculate the value of R^2 and interpret its value.
$R^2 = 1 - .36618/13.19690 = .9722$. *This means that 97% of the variation in the number of cells can be explained by (the linear) relationship with temperature.*

c) Is there a statistically significant linear relationship between temperature and the number of cells? Describe the formal hypotheses and justify how you reached your conclusion.
Yes. $H_0 : \beta_1 = 0$ *and* $H_a : \beta_1 \neq 0$. *We can use an F-test to test this hypothesis. The F-statistic is 660.2 and has an F-distribution with 1 and 18 df under the null hypothesis. The resulting p-value is .0001, whihc implies very strong evidence against the null hypothesis – ie we would reject for any* $\alpha > .0001$.

d) Based on these results can we conclude that the straight line model is appropriate? Explain.
No. While the F-test and R^2 *values indicate that there is a strong linear relationship, we should look at residual plots to confirm that a linear relationship, and not a curvilinear or nonlinear relationship is more appropriate.*

Problem 2.

The susceptibility of catfish to a certain chemical was determined by immersing individual fish in 2 liters of an emulsion containing the pollutant and measuring the survival time in minutes. A straight line regression was fit using log base 10 of survival time (Y) as the dependent variable, and log base 10 of the concentration as X (measured in parts per million). Refer to the output at the end in order to answer the questions.

a) Test for the significance of the straight line regression and interpret your result.
$H_o : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$. Again we can use either the F-test or a t-test. The t statistic is -.25.84 with a p-value of 18×10^{-15}, which is highly significant.
This implies that there is a very strong relationship between log(survival time) and log(concentration).

b) What is the estimated equation to predict $Y = \log_{10}$ survival time? Give interpretations of the slope and intercept.
Predicted log(survival time) = 7.4911 -.9981 log(concentration).
The intercept reflects that for a log(concentration) = 0 (or a 1 ppm concentration of the pollutant) that the expected log(survival time) is 7.4911 log minutes or $10^7.4911$ minutes – roughly 58 years!
The slope reflects that for every increase in log concentation of 1 unit, we expect the log survival time to decrease by .9981 log minutes. In the original units, this translates into: for every 10 fold increase in concentration, we expect a $10^{-.9981}$ or 0.1% decrease in survival time.
c) Determine a 95% prediction interval for the **survival time** for a new fish exposed to a concentration of 100,000 ppm ($X = 5$). (sample mean is $\bar{X} = 4.5$, and $TSS_X = 2.1$)
the predicted log survival time is 2.5 using the model in (b), which corresponds to a predicted survival time of $10^2.5$ or 316 minutes or 5.27 hours. The prediction interval is predicted log survival time $\pm$ t $SE_{\hat{Y}}$.
$SE_{\hat{Y}} = \sqrt{.0031(1 + \frac{1}{18} + \frac{(5-4.5)^2}{2.1})} = .06$
The t-value is 2.120 based on 16 df.
The Prediction interval is $(2.5 \pm 2.120 * .06)$ = (2.372, 2.628) or (235.5 minutes to 424.6 minutes) or (3.9 hours to 7.08 hours).

d) Does the straight line model seem appropriate. Explain briefly.
Yes. the rsidual versus fitted plot shows no signs of outliers, nonconstant variance or non-linearities. While the noral quantile plot may show some very mild signs of a distribution with tails lighter than a normal, this is really not a problem. "All Clear"

Problem 3. Rathburn (1988) reported on experiments in surveying manatees along Florida's Crystal and Indian Rivers by observation from helicopter and airplane. Helicopters were thought to provide more accurate counts due to their slower speed, but are more expensive. The following results are from the regression of airplane count (Y) on the helicopter count (X) for 10 different days.

Parameter Estimates

Variable	DF	Estimate	Std Error	T Stat	Prob >|T|
INTERCEPT	1	-7.4632	12.4957	-0.5973	0.5669
HELICOPT	1	1.1483	0.2311	4.9692	0.0011

a) What is the interpretation of the slope and intercept in this problem?
Intercept: If the Helicopter count is 0, we have an estimate of -7.46 manatees counted by the airplane.
Slope: For every manatee counted by the helicopter, we expect to count 1.148 by airplane.
b) Construct a 95% confidence interval for the slope. Based on this interval, would you reject that $\beta_1 = 0$? $\beta_1 = 1$? Which seems more relevant to this study? Explain.
t from the tables is2.306 based on 8 df. 95% *CI = (.6154, 1.6809). Based on this interval, we would reject Ho: that the true slope is 0, as the interval excludes 0, but we would not reject Ho: that the true slope is 1, as the interval does contain one. This means we do not have enough evidence to distinguish the slope from 1. The hypothesis that the slope is 1 is really more relevant here, as this means that for every increase in the helicopter count we expect exactly one additional manatee to be counted by the airplane – or that there is a one-to-one increase in the manatee/helicopter counts.*
c) Carry out a t-test of $Ho: \beta_0 = 0$ and interpret your conclusions.
The test statistic is t = -.5973, which has a Student t distribution with 8 df. The p-value is 0.5669, which means that we would not reject Ho for most reasonable choices of α. This together with the results from (b) imply that there is no evidence a gainst a one-to-one correspondence between helicopters and airplanes for counting manatees.
d Can the regression line with slope 1 and y-intercept 0 have a smaller residual sum of squares than the line in (a)? Explain. Which regression line would you choose? Explain briefly.
No way. The least squares estimates always have the smallest residual sum of squares. However, we might be willing to have a model that doesn't fit as well inexchange for the nicer interpretation that the two methods agree. — For the extra energetic student: how would you test that $Ho: \beta_0 = 0 \beta_1 = 1$?