ST512

Fall Semester, 2008

Quiz 1

Name:____

Directions: Answer questions as directed. Please show work. Partial credit may be awarded for correct expressions given in incomplete answers. To save time, it is ok to give answers as numeric expressions without carrying out every last operation. If the question asked for the standard error of the estimated mean of a population at $x = 14$ in a simple linear regression, the following response would receive full credit:

$$\widehat{SE}(\hat{\beta}_0 + 14\hat{\beta}_1) = \sqrt{16.4\left(\frac{1}{38} + \frac{(14-10.8)^2}{190}\right)}$$

There is a blank page at the end for extra space. You may also use the back of the page if necessary.

1. In NFL game from a sample of $n = 682$ games, two measurements are made: the published point spread x and margin of victory y for the favored team. A simple linear regression of y on x was fit with SAS. Code and *selected* output are given below. For all significance tests, use $\alpha = .05$.

```
proc reg data=spreads ;
   model outcome=spread;
run;
                              The SAS System                              1
                              The REG Procedure

           Root MSE              13.26051     R-Square      0.0755
           Dependent Mean         6.09673     Adj R-Sq      0.0742
           Coeff Var            217.50217

                         Parameter        Standard
   Variable      DF       Estimate           Error     t Value     Pr > |t|
   Intercept      1        0.02755         0.96668        0.03       0.9773
   spread         1        1.14291         0.15446        7.40       <.0001
```

(a) Obtain a 95% confidence interval for the slope term (β_1) from the regression model. At level $\alpha = .05$, is it plausible that $\beta_1 = 0$? How about $\beta_1 = 1$?

(b) Is $\mu(x = 0) = 0$ a plausible value for the mean outcome when $x = 0$? Conduct an appropriate test.

(c) Let the average point spread and outcome be denoted by $\bar{x}$ and $\bar{y}$, respectively. (Note that the latter is given in the output as $\bar{y} = 6.10$.) Consider games in which one team is favored by $x = \bar{x}$ points.

i. Report an estimate of the mean margin of victory for the favored team in these $(x = \bar{x})$ games, along with a standard error.

ii. Report an estimate of the standard deviation of the margin of victory for the favored teams in these $(x = \bar{x})$ games.

iii. Obtain a 95% prediction interval for the amount by which a team favored by this amount $(x = \bar{x})$ will win in an upcoming game.

(d) By writing (up,down, same) indicate how the answers to questions i) and ii) in part (c) would change if games where $x = \bar{x} + 1$ were considered:

i. mean ? standard error ?

ii.

2. An experiment with $n = 20$ observations is run to minimize peanut kernel damage y, during the shelling process. Consider the two multiple linear regression models below. The predictor variables are spacing between sheller bars (x) and stroke frequency, ($x_h = 0$ for low frequency and $x_h = 1$ for high frequency):

$$\begin{aligned}\mu_1(x, \text{high}) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x_h + \beta_4 x_h x + \beta_5 x_h x^2 \\ \mu_2(x, \text{high}) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x_h\end{aligned}$$

The models are fit using the SAS code on the next page, where predictors are given in the model statements in the same order as above. Assume kernel damage varies normally about its mean, with constant variance.

(a) Report an F-ratio (and degrees of freedom) for a test comparing the two models.

(b) Estimate the mean kernel damage when a spacing of $x = 1$ *in.* is used at high frequency. Give two answers, one for each model.

(c) Let the estimated covariance matrices of $\hat{\beta}$ under models μ_1 and μ_2 be denoted by $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, respectively. Give matrix expressions for the standard errors of each of the two estimated means in part (b).

```
proc reg data=goobers;
   model damaged=x x2 high highx highx2 ;
   model damaged=x x2 high ;
run;
```

The REG Procedure

Model: MODEL1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	62.49729	12.49946	32.18	<.0001
Error	14	5.43868	0.38848		
Corrected Total	19	67.93598			

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	46.21242	9.26303	4.99	0.0002
x		1	-87.17297	22.69791	-3.84	0.0018
x2		1	43.66730	13.17293	3.31	0.0051
high		1	-25.53382	9.58744	-2.66	0.0185
highx		1	53.62141	23.42934	2.29	0.0382
highx2		1	-27.29604	13.59290	-2.01	0.0643

Model: MODEL2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	54.31576	18.10525	21.27	<.0001
Error	16	13.62021	0.85126		
Corrected Total	19	67.93598			

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	24.16788	3.48654	6.93	<.0001
x		1	-38.03888	8.31808	-4.57	0.0003
x2		1	17.81967	4.80912	3.71	0.0019
high		1	-0.75035	0.47862	-1.57	0.1365

3. A dentistry experiment randomizes subjects to two groups ($n = 20$ each). One group receives a control toothpaste, the other an experimental fluoride toothpaste. Average daily plaque scores, y, are measured as the response, along with a covariate of brushing frequency called *compliance*, which is centered about its observed mean (1.984) leading to a mean-zero covariate, $z = compliance - 1.984$. SAS code and output pertinent to this problem are given on the next page. In testing for an effect of the fluoride treatment, consider the model in which mean plaque score depends linearly on *compliance* (and hence on z), and in which this dependence is constant across treatments.

   (a) Report a statistic, p-value, and associated degrees of freedom for a test for a fluoride treatment effect on plaque score after controlling for *compliance*.

   (b) Report the mean plaque score for each treatment, adjusted to the overall mean compliance of 1.984 (or $z = 0$).

Group	Adjusted mean
control	
fluoride	

   (c) Estimate the difference between mean plaque score for the two treatments at a given compliance. Report a standard error.

   (d) Give the proportion of observed variation in plaque score explained by a simple linear regression on z, where the treatment (toothpaste type) is ignored.

(e) Tough problem: given that the unadjusted mean plaque score was higher for the control group than for the fluoride group, recover the unadjusted means for each group using the output.

```
proc reg data=teeth2;
   model plaque=z fluoride /ss1 ss2; /* fluoride is an indicator for fluoride group */
run;
```

****

The SAS System

The REG Procedure

Model: MODEL1

Dependent Variable: plaque

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	34.13180	17.06590	17.30	<.0001
Error	37	36.50552	0.98664		
Corrected Total	39	70.63731			

Root MSE	0.99330	R-Square	0.4832
Dependent Mean	7.40873	Adj R-Sq	0.4553
Coeff Var	13.40709		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	7.77946	0.22243	34.97	<.0001	2195.57355	1206.88547
z	1	-0.90661	0.17450	-5.20	<.0001	28.66618	26.63263
fluoride	1	-0.74145	0.31502	-2.35	0.0240	5.46562	5.46562