STAT 512 Midterm Exam July 8 2005

Name:

Show all your work (you can use this page).

All questions are for 5 points (100 total).

Good luck !

1. You fit the model $Y= \beta_0+\beta_1X_1+\beta_2X_2+\xi$. You use n=25 observations and your estimators are $b_0 = 4$, $b_1 = 1$, $s(b_1)=0.5$, $b_2 = 3$, $s(b_2)=2$ and $s=1$.
a) Test if $\beta_1 = 0$ (Compute the test statistic, give the number of degrees of freedom, corresponding critical value and conclusion).

b) Predict the value of Y for $X_1=2$ and $X_2=5$.

c) $s(\hat{\mu})$ (standard error of the estimator of mean of Y) at $X_1=2$ and $X_2=5$ is equal to 0.3. Estimate the variance of the error of the prediction you made in (b).

d) Construct a corresponding prediction interval.

2. You want to test if a certain medicine has an influence on patient's blood pressure. You are fitting the model : Y(blood pressure)= $\beta_0+\beta_{1*}$age+β_{2*}ind+ β_{3*}ageind + ξ,
where ind is a binary predictor: ind=0 if a patient did not take the medicine (control group), ind=1 if a patient took the medicine (test group). Variable ageind is the product of age and ind.

Below you have the table of Type I and Type II sum of squares:

	Type I	Type II
age	150	130
ind	50	40
ageind	30	30

Total sum of squares (SST) is equal to 530 and n=30.

a) Compute SSM for the full model.

b) Find R^2 for the full model.

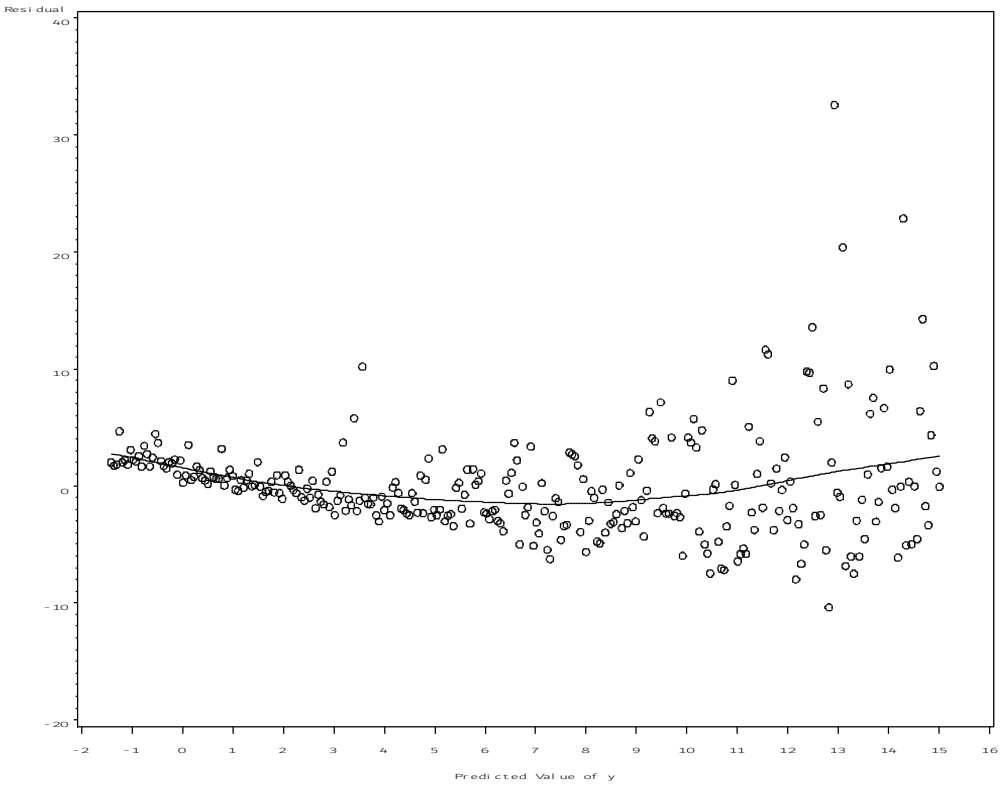
c) Estimate the variance of the error term in the full model.

d) What proportion of the variation of a blood pressure is explained by a simple regression on age ?

e) Compute the sample correlation coefficient between blood pressure and age.

f) Test the hypothesis that the dependence of blood pressure on age is not affected by medicine (Regression lines in test and control groups are the same).

3. Below you are given a plot of residuals from a simple linear regression of Y on X. The residuals are plotted versus predicted values of Y.



a) Do you observe any significant patterns ? Which assumptions of simple linear regression seem to be violated ?

b) Which remedial measures should you take ?

4. Below you are given the output from SAS proc reg.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	
Model	2	50968	25484	6277.00	<.0001
Error	88	357.26788	4.05986		
Corrected Total	90	51325			

Root MSE	2.01491	R-Square	0.9930
Dependent Mean	49.53355	Adj R-Sq	0.9929
Coeff Var	4.06777		

Parameter Estimates

Var	DF	Estimate	Error	t Value	Pr > |t|	Type I SS	Type II SS
Int	1	-0.02877	0.49058	-0.06	0.9534	223275	0.01397
x1	1	7.17878	3.77470	1.90	0.0605	50967	14.68411
x2	1	0.91524	1.88674	0.49	0.6288	0.95533	0.95533

Model Selection Criteria

Number in Model	C(p)	R-Square	AIC	SBC	Variables
1	1.2353	0.9930	128.6970	133.71873	x1
2	3.0000	0.9930	130.4540	137.98658	x1 x2
1	4.6169	0.9928	132.1194	137.14110	x2

a) How many cases you have in your file ?

b) Give the estimator of the standard deviation of the error term.

c) Predict the value of Y for x1=3 and x2=-2.

d) Report the result of the test that coefficients by both explanatory variables are equal to zero. (Give the test statistic, the number of degrees of freedom, P-value and conclusion.)

e) Report the results of t-tests for individual coefficients. (Formulate the hypothesis, give the test statistic, the number of degrees of freedom, P-value and conclusion.)

f) Compare points d) and e) and explain. Which variable would you choose for a simple regression model ?

g) Test for significance of x1 in the simple regression model.

h) Report the best model in terms of C(p), AIC and SBC.