THE UNIVERSITY OF BRITISH COLUMBIA

FORESTRY 430 and 533

MIDTERM EXAMINATION: October 16, 2006 Instructor: Val LeMay

Time: 50 minutes

40 Marks FRST 430 50 Marks FRST 533 (extra questions)

This examination consists of 8 pages (2 questions). **There are two extra part-questions for FRST 533 students only.**

(15) 1. Satellites can be used to measure relectances for different wavelengths. For a small land area, you obtain values for the wavelengths that appear green to the human eye, for 30 X 30 m areas (pixels). For 10 of these pixels, you obtain ground measurements that allow you to calculate the volume per ha for these selected pixels. You then use these data to obtain a regression equation to predict ground volume per ha from green wavelength measures. Once you have this, you will be able to apply the equation to all pixels to obtain estimated volume per ha for every pixel. However, you have used all your money to get these data, and cannot afford to pay for a statistical package. Therefore, you use EXCEL to do your analyses.

Using EXCEL you graph volume per ha (Y) versus green wavelength measures (X) (page 2). You are not sure if this is a linear relationship so you also graph the logarithm of volume per ha (Y) versus the logarithm of the green measures (X) (page 3).

a. Based on these two graphs, which variables (original or transformed) appear to be result in a more linear trend?

b. Using **your choice of original or transformed data only**, use the EXCEL data and preliminary calculations to obtain:

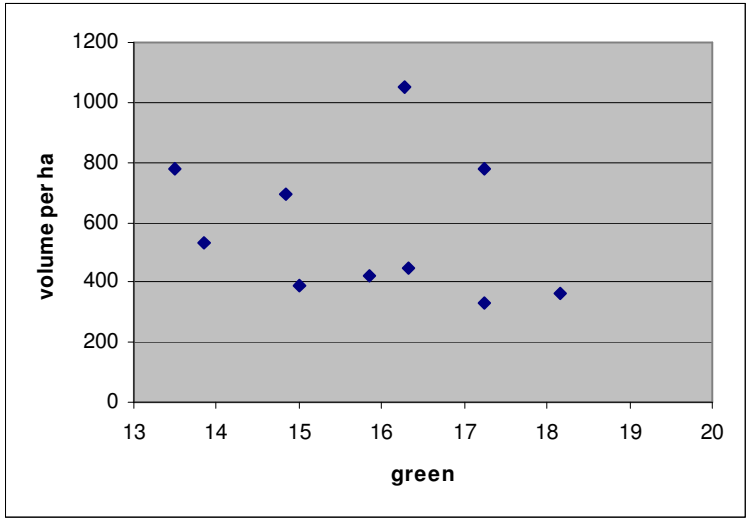
   1) the estimated slope and intercept
   2) the sum of squares Y, sum of squares regression, and sum of squares error
   3) the r^2 value
   4) the standard error of the estimate (SEE)

c. FRST 533 only: Based on the results, is this a good equation? Give specific evidence from your analysis.

volume (y)	green (x)	yi-meany	(yi-meany)sq	xi-meanx	(xi-meanx)sq	(xi-meanx) X (yi-meany)
450	16	-130	16838	0.5	0.2	-64.9
365	18	-214	45736	2.3	5.4	-499.0
422	16	-157	24762	0.0	0.0	-3.7
532	14	-48	2271	-2.0	3.9	94.2
777	17	197	38904	1.4	2.0	279.4
1051	16	471	222161	0.5	0.2	213.2
334	17	-245	60251	1.4	2.0	-347.7
388	15	-192	36734	-0.8	0.7	159.7
779	14	199	39776	-2.3	5.4	-465.4
697	15	118	13863	-1.0	1.0	-117.7

Means	579	16

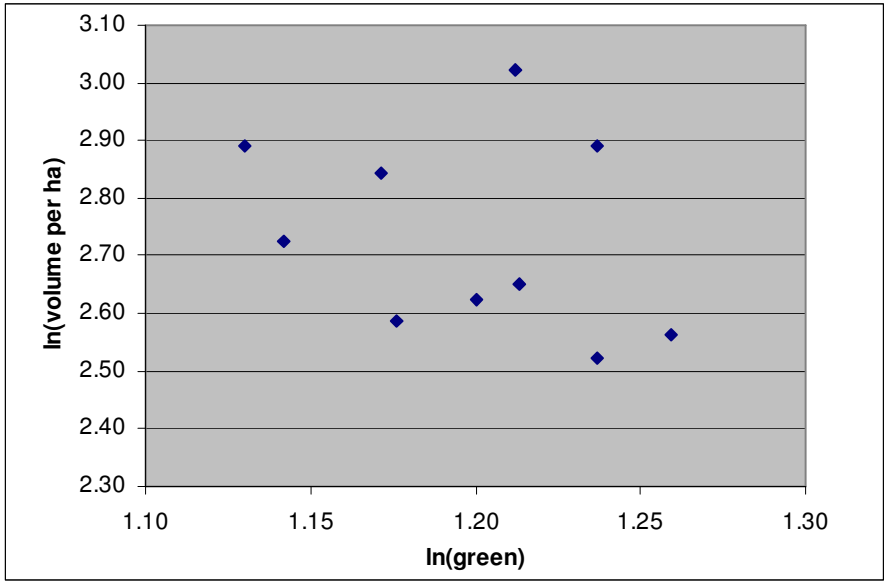
Sums:		0	501296	0.0	21.0	-751.9



volume	green	lnvolume (y)	lngreen (x)	yi-meany	(yi-meany)sq	xi-meanx	(xi-meanx)sq	(xi-meanx) X (yi-meany)
450	16	2.65	1.21	-0.08	0.01	0.02	0.00	0.00
365	18	2.56	1.26	-0.17	0.03	0.06	0.00	-0.01
422	16	2.63	1.20	-0.11	0.01	0.00	0.00	0.00
532	14	2.73	1.14	-0.01	0.00	-0.06	0.00	0.00
777	17	2.89	1.24	0.16	0.02	0.04	0.00	0.01
1051	16	3.02	1.21	0.29	0.08	0.01	0.00	0.00
334	17	2.52	1.24	-0.21	0.04	0.04	0.00	-0.01
388	15	2.59	1.18	-0.14	0.02	-0.02	0.00	0.00
779	14	2.89	1.13	0.16	0.03	-0.07	0.00	-0.01
697	15	2.84	1.17	0.11	0.01	-0.03	0.00	0.00

Means:	2.73	1.20

Sums:		0.00	0.26	0.00	0.02	-0.02



(25) 2. I wanted to know if I could use the amount of shrub cover and rock cover (ground covered by shrubs and by rocks) to predict the biomass of bees (used the logarithm of beesbm), as I am somewhat allergic to bee stings. Regression was used to fit the following equation to predict beesbm:

$$MODEL\,1: \quad \ln \hat{bees} = b_0 + b_1 \times \ln shrubso + b_2 \times shrubso + b_3 \times \ln rocks$$

where:

- b_0 to b_3 are the estimated coefficients;

(see SAS outputs starting on page 5).

NOTE: Take values from the outputs where-ever possible
ALSO indicate what alpha level you used for all tests.

(a) Using just the residual plot and the normality plot (and normality tests), which assumptions of multiple linear regression can you check? Are these met?

(b) What are the R^2 and SEE values?

(c) Test whether the regression is significant. Show the hypothesis, test statistic, p-value or critical value from a table, and the decision.

(d) Test whether each of the variables is significant. **Show a general hypothesis for all variables to be tested.** Then for each variable, give the test statistic, the p-value (or critical values from a table), and the decision.

(e) Is this a good model? Give evidence to support your statement.

(f) FRST 533 only: For this problem, would you recommend using selection methods (e.g., forward, backward, stepwise, R square selection methods)? Why or why not?

The REG Procedure

Model: MODEL1

Dependent Variable: lnbees

Number of Observations Read	35
Number of Observations Used	25
Number of Observations with Missing Values	10

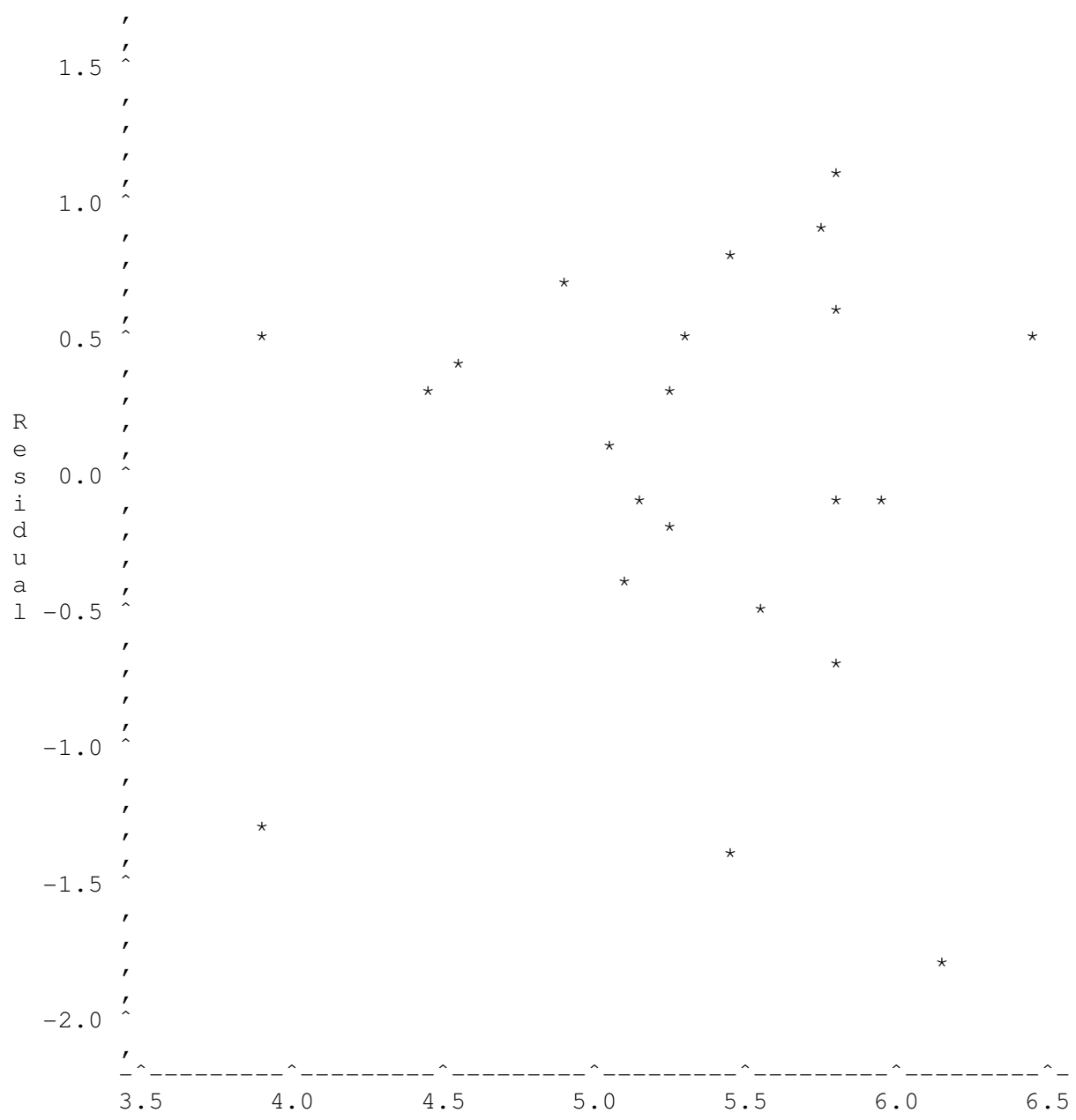
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	10.68279	3.56093	5.70	0.0051
Error	21	13.11965	0.62475		
Corrected Total	24	23.80244			

Root MSE	0.79041	R-Square	0.4488
Dependent Mean	5.35714	Adj R-Sq	0.3701
Coeff Var	14.75428		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	2.51112	1.66268	1.51	0.1459
lnshrubs		1	3.62474	1.32124	2.74	0.0122
shrubso	shrubso	1	-0.37114	0.12365	-3.00	0.0068
lnrocks		1	-0.44227	0.15029	-2.94	0.0078



NOTE: 10 obs had missing values. 3 obs hidden.

The UNIVARIATE Procedure
Variable: resid1 (Residual)

Moments

N	25	Sum Weights	25
Mean	0	Sum Observations	0
Std Deviation	0.73935921	Variance	0.54665204
Skewness	-0.8542429	Kurtosis	0.24245464
Uncorrected SS	13.1196489	Corrected SS	13.1196489
Coeff Variation	.	Std Error Mean	0.14787184

Basic Statistical Measures

Location		Variability	
Mean	0.000000	Std Deviation	0.73936
Median	0.132897	Variance	0.54665
Mode	.	Range	2.90491
		Interquartile Range	0.87379

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	1.5	Pr >= |M|	0.6900
Signed Rank	S	17.5	Pr >= |S|	0.6472

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.936894	Pr < W	0.1255
Kolmogorov-Smirnov	D	0.118759	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.089697	Pr > W-Sq	0.1481
Anderson-Darling	A-Sq	0.570031	Pr > A-Sq	0.1299

The UNIVARIATE Procedure
Variable: resid1 (Residual)

Quantiles (Definition 5)

Quantile	Estimate
100% Max	1.121086
99%	1.121086
95%	0.864051
90%	0.796936
75% Q3	0.500728
50% Median	0.132897
25% Q1	-0.373064
10%	-1.285679
5%	-1.390941
1%	-1.783824
0% Min	-1.783824

Extreme Observations

------Lowest------		------Highest-----	
Value	Obs	Value	Obs
-1.783824	32	0.569543	17
-1.390941	7	0.747106	4
-1.285679	26	0.796936	18
-0.725528	13	0.864051	6
-0.672329	35	1.121086	25

Missing Values

Missing Value	Count	-----Percent Of----- All Obs	Missing Obs
.	10	28.57	100.00

The UNIVARIATE Procedure
Variable: resid1 (Residual)

Stem	Leaf	#	Boxplot
1	1	1	
0	55556789	8	+-----+
0	11334	5	*--+--*
-0	42111	5	+-----+
-0	775	3	
-1	43	2	
-1	8	1	0

----+----+----+----+

Normal Probability Plot