THE UNIVERSITY OF BRITISH COLUMBIA

FORESTRY 430 and 533

MIDTERM EXAMINATION: October 14, 2005 Instructor: Val LeMay
Time: 50 minutes
40 Marks FRST 430 50 Marks FRST 533 (extra questions)

This examination consists of 11 pages (2 questions). **There are two extra part-questions for FRST 533 students only.**

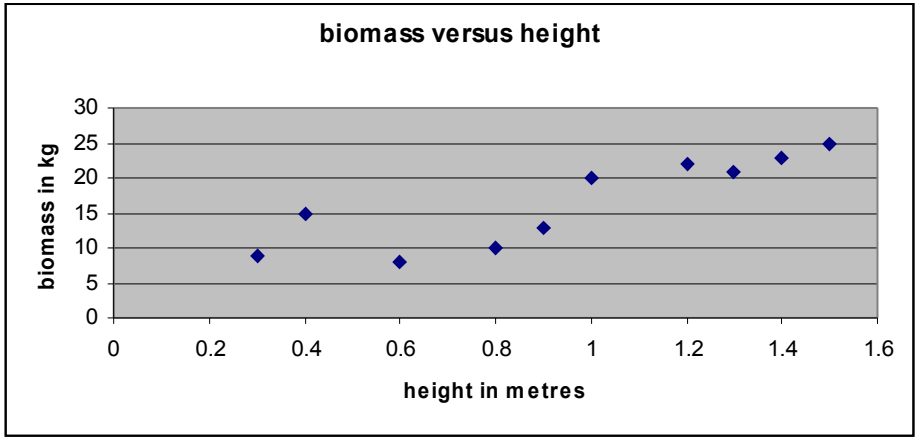
(15) 1. As part of your short-term research job in South Africa, you collect data on biomass (kg) of Protea shrubs, and measure the shrub heights. Since biomass requires destructive sampling of the plants, you would like to develop an equation to estimate biomass from shrub heights. You have a very limited budget and collect data for only 10 plants. Also, you do not have money to buy a statistical package, and there is limited power in the area where you are staying. You manage to get the data into EXCEL, do some preliminary calculations and a graph, and then your laptop has no more power. Using the EXCEL values, calculate:

a. the estimated slope and intercept
b. the sum of squares Y, sum of squares regression, and sum of squares error
c. the r^2 value
d. the standard error of the estimate (SEE)
e. Do you think you might need transformations of the variables, or is this a line? Is the first assumption of simple linear regression met?
f. **FRST 533 only: Is this a good equation in your opinion? How might you improve this equation? Give three possible ways.**

plant	Biomass (y)	Height (x)	biomass-mean	(biomass-mean) sq.	height-mean	(height-mean) sq.	(biomass-mean) X (height-mean)
1	20	1	3.4	11.56	0.06	0.0036	0.204
2	25	1.5	8.4	70.56	0.56	0.3136	4.704
3	13	0.9	-3.6	12.96	-0.04	0.0016	0.144
4	10	0.8	-6.6	43.56	-0.14	0.0196	0.924
5	22	1.2	5.4	29.16	0.26	0.0676	1.404
6	21	1.3	4.4	19.36	0.36	0.1296	1.584
7	15	0.4	-1.6	2.56	-0.54	0.2916	0.864
8	8	0.6	-8.6	73.96	-0.34	0.1156	2.924
9	23	1.4	6.4	40.96	0.46	0.2116	2.944
10	9	0.3	-7.6	57.76	-0.64	0.4096	4.864

mean	16.60	0.94

Sums:			0.00	362.40	0.00	1.56	20.56



(25) 2. Regression was used to fit the following equation to predict crown ratio for birch trees:

$$ALL: \quad \hat{C}R = b_0 + b_1 \times ccf + b_2 \times dbhsq$$

where:

- b_0 to b_2 are the estimated coefficients;
- CR is the crown ratio (length of live crown relative to total tree height);
- Predictor variables are dbhsq (the measured tree diameter at 1.3 m above ground, squared), and ccf (crown competition factor) to represent competition **(see SAS outputs starting on page 4).**

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

(a) Do the residuals meet the assumptions of regression using the residual plot and the normality plot?

(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:**

(a) How many equations would you need to do in order to get all possible regressions if there were three variables to use as predictors?

(ii) How would you add a class variable, region, with three regions, to the equation?

The REG Procedure
Model: ALL
Dependent Variable: CR CR

Number of Observations Read	93
Number of Observations Used	93

Analysis of Variance

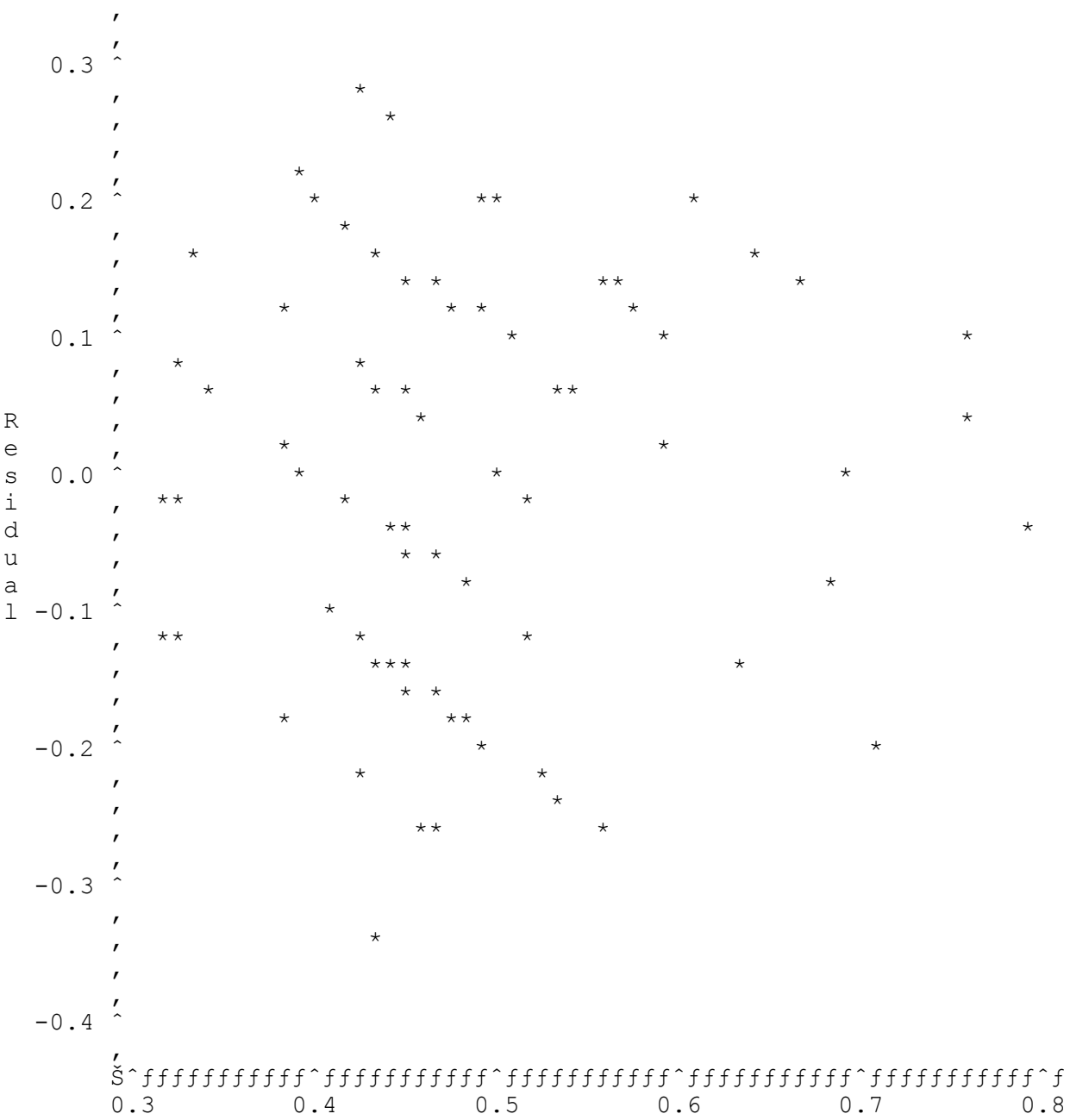
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.04026	0.52013	25.22	<.0001
Error	90	1.85636	0.02063		
Corrected Total	92	2.89661			

Root MSE	0.14362	R-Square	0.3591
Dependent Mean	0.47097	Adj R-Sq	0.3449
Coeff Var	30.49428		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.78709	0.08281	9.50	<.0001
CCF	CCF	1	-0.00217	0.00044287	-4.89	<.0001
dbhsq		1	0.00039252	0.00009161	4.28	<.0001

Plot of resid1*pred1. Symbol used is '*'.



NOTE: 22 obs hidden.

The UNIVARIATE Procedure
Variable: resid1 (Residual)

Moments

N	93	Sum Weights	93
Mean	0	Sum Observations	0
Std Deviation	0.14204856	Variance	0.02017779
Skewness	-0.2184176	Kurtosis	-0.8541527
Uncorrected SS	1.85635697	Corrected SS	1.85635697
Coeff Variation	.	Std Error Mean	0.01472975

Basic Statistical Measures

Location		Variability	
Mean	0.000000	Std Deviation	0.14205
Median	0.005888	Variance	0.02018
Mode	.	Range	0.60953
		Interquartile Range	0.24190

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	1.5	Pr >= |M|	0.8358
Signed Rank	S	36.5	Pr >= |S|	0.8897

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.974452	Pr < W	0.0648
Kolmogorov-Smirnov	D	0.085903	Pr > D	0.0894
Cramer-von Mises	W-Sq	0.121594	Pr > W-Sq	0.0585
Anderson-Darling	A-Sq	0.748436	Pr > A-Sq	0.0494

Quantiles (Definition 5)

Quantile	Estimate
100% Max	0.27880851
99%	0.27880851
95%	0.20440556
90%	0.18174849
75% Q3	0.12044012
50% Median	0.00588825
25% Q1	-0.12145754
10%	-0.19199049
5%	-0.23280473
1%	-0.33072324
0% Min	-0.33072324

Extreme Observations

------Lowest------		------Highest-----	
Value	Obs	Value	Obs
-0.330723	83	0.204406	54
-0.266861	13	0.205416	31
-0.262322	2	0.210140	36
-0.259450	17	0.258583	25
-0.232805	8	0.278809	7

Stem	Leaf	#	Boxplot
2	68	2	
2	0000011	7	
1	55556678	8	
1	1112233334	10	+-----+
0	56667777788899	14	| |
0	0111244	7	*--+--*
-0	444432222220	12	| |
-0	999655	6	| |
-1	44433322211	11	+-----+
-1	9988755	7	
-2	33321	5	
-2	766	3	
-3	3	1	

Multiply Stem.Leaf by 10**-1

The UNIVARIATE Procedure
Variable: resid1 (Residual)

Normal Probability Plot