

STAT nnn
George McCabe
Name:

EXAM I.2
mm dd, yyyy

Show all work on these pages

- (25) 1. (a) In a linear model the design (X) matrix is 4×20 . What is the dimension of the error vector?
- (b) For an analysis of the model $Y = \beta_0 + \beta_1 X + e$, the estimate of β_1 was 6.4 with a standard error of 1.6. What is the value of the t statistic for testing the null hypothesis that $\beta_1 = 0$?
- (c) Refer to the previous problem. What additional information would you need to find a P -value for this hypothesis test and how would you use it?
- (d) The correlation between X and Y is .5. What proportion of the variation in Y would be explained by a linear regression using X as an explanatory variable?
- (e) Suppose the estimated regression equation is $\hat{Y} = 10 + 5X$. If the analysis was rerun with X replaced by $U = 5X + 2$, give the estimated regression equation.

- (25) 2. (a) In a multiple linear regression analysis, all of the P -values for tests on the individual coefficients are large, say greater than .6, but the P -value for the F -test is .001. Explain how this can happen.
- (b) Refer to the previous problem. What additional analysis would you perform to improve the model?
- (c) The 95% confidence interval for a regression coefficient is -2.1 to 8.6. What can you say, if anything, about the P value for the hypothesis test that this coefficient is zero?
- (d) Data were collected on a sample of 25 women and 25 men. Separate simple linear regressions were run for each group. For women the slope was 30 with a standard error of 8; for men the slope was 20 with a standard error of 10. Calculate the value of the t statistic that you would use to compare the two slopes and give its degrees of freedom.
- (e) Two different laboratory procedures were used to find the amount of calcium in a collection of orange juice samples. The correlation between the two measures was .95. Does this imply that the value given by the first procedure was approximately the same as the value given by the second procedure for each of the samples analyzed? Explain why or why not.

- (25) 3. For this question refer to the output from PROC REG given as OUTPUT FOR PROBLEM 3 and the plot that follows. The data analyzed are from problem 1.22 in the text. Read the description of the data on page 39.
- (a) Give the estimated regression equation. (Be sure to round off in an appropriate way.)

 - (b) Is there a linear relation between hardness and time? Give an estimate, a test statistic with degrees of freedom and P -value to support your statement.

 - (c) Why are the predicted values for the first four cases the same?

 - (d) Does the output give confidence intervals for mean hardness or prediction intervals for a future observation? What is the interval for the mean hardness of items run for 24 hours?

 - (e) Suppose the data have been taken from a process where you would like to produce items with an average hardness of 220 (Brinell units). What time value would you choose to achieve this?

- (25) 4. For this question refer to the SAS output given as OUTPUT FOR PROBLEM 4. The data are from a study designed to relate the taste of cheddar cheese to concentrations of various chemicals that naturally are found in cheese. The variables are taste (TASTE), acetic acid (ACETIC), hydrogen sulfide, (H2S), and lactic acid (LACTIC). Acetic acid and hydrogen sulfide are expressed in logs.
- (a) Give the fitted regression model for the analysis using all three explanatory variables.

 - (b) What do the significance tests for this analysis say about the relationship of the three chemicals to taste. Be specific.

 - (c) There is a significance test given with the correlation table for the pair taste and acetic. There is also a test for the acetic coefficient in the regression analysis. Give the results of these two tests and summarize what this means.

 - (d) Give the fitted model for the analysis using H2S and LACTIC. Why are the coefficients for H2S different in the two analyses?

 - (e) Compare the percent of variation in taste explained by the two models. Is the difference significant? Give a test statistic and P -value.

OUTPUT FOR PROBLEM 3

Model: MODEL1

Dependent Variable: HARDNESS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	5297.51250	5297.51250	506.506	0.0001
Error	14	146.42500	10.45893		
C Total	15	5443.93750			

Root MSE	3.23403	R-square	0.9731
Dep Mean	225.56250	Adj R-sq	0.9712
C.V.	1.43376		

Parameter Estimates

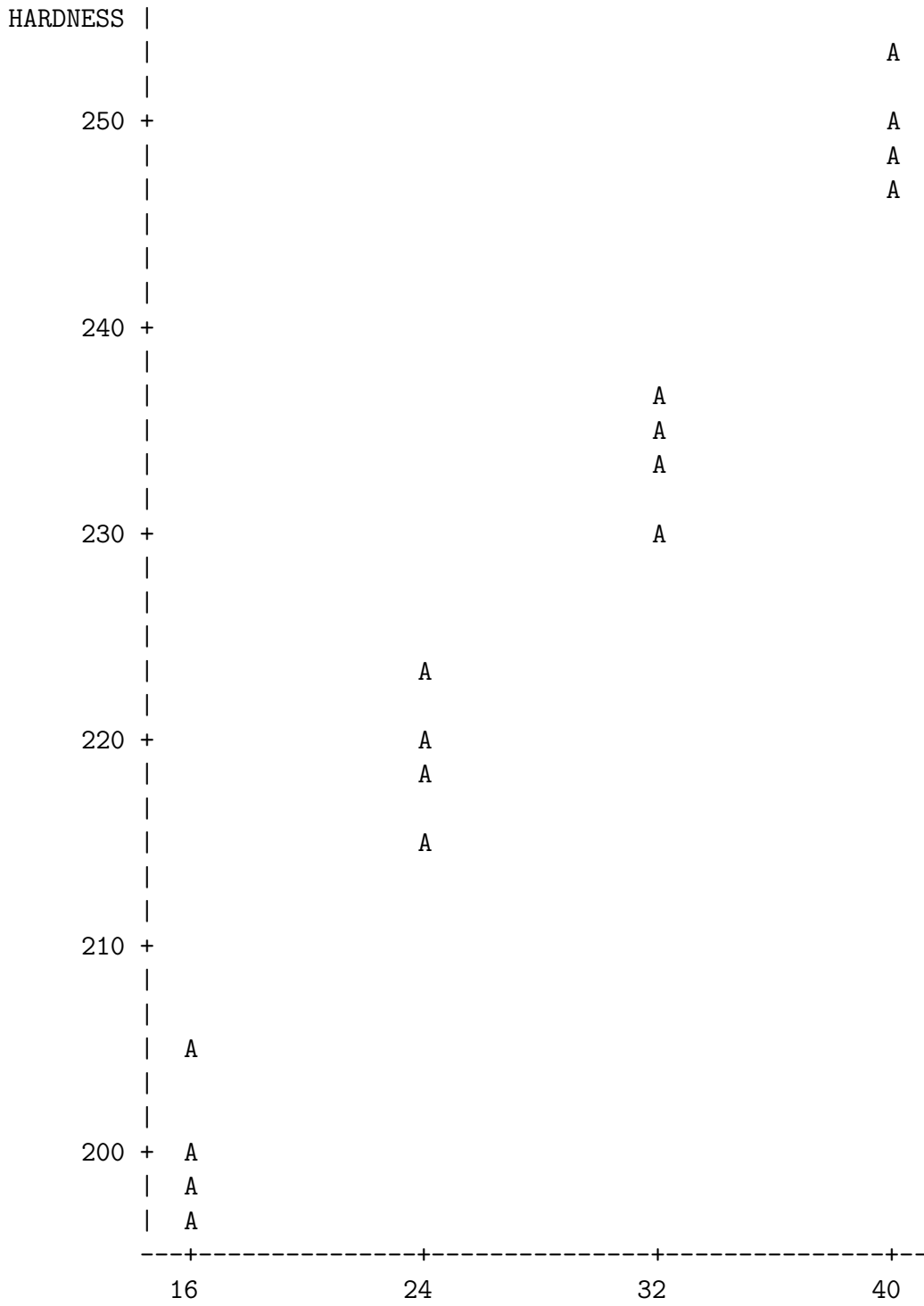
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	168.600000	2.65702405	63.454	0.0001
TIME	1	2.034375	0.09039379	22.506	0.0001

Obs	TIME	Dep Var HARDNESS	Predict Value	Std Err Predict	Lower95% Mean	Upper95% Mean
1	16	199.0	201.2	1.353	198.2	204.1
2	16	205.0	201.2	1.353	198.2	204.1
3	16	196.0	201.2	1.353	198.2	204.1
4	16	200.0	201.2	1.353	198.2	204.1
5	24	218.0	217.4	0.886	215.5	219.3
6	24	220.0	217.4	0.886	215.5	219.3
7	24	215.0	217.4	0.886	215.5	219.3
8	24	223.0	217.4	0.886	215.5	219.3
9	32	237.0	233.7	0.886	231.8	235.6
10	32	234.0	233.7	0.886	231.8	235.6
11	32	235.0	233.7	0.886	231.8	235.6
12	32	230.0	233.7	0.886	231.8	235.6
13	40	250.0	250.0	1.353	247.1	252.9
14	40	248.0	250.0	1.353	247.1	252.9
15	40	253.0	250.0	1.353	247.1	252.9
16	40	246.0	250.0	1.353	247.1	252.9

Obs	TIME	Residual	Std Err Residual	Student Residual	-2-1-0 1 2	Cook's D
1	16	-2.1500	2.937	-0.732	*	0.057
2	16	3.8500	2.937	1.311	**	0.182
3	16	-5.1500	2.937	-1.753	***	0.326
4	16	-1.1500	2.937	-0.391		0.016
5	24	0.5750	3.110	0.185		0.001
6	24	2.5750	3.110	0.828	*	0.028
7	24	-2.4250	3.110	-0.780	*	0.025
8	24	5.5750	3.110	1.792	***	0.130
9	32	3.3000	3.110	1.061	**	0.046
10	32	0.3000	3.110	0.096		0.000
11	32	1.3000	3.110	0.418		0.007
12	32	-3.7000	3.110	-1.190	**	0.057
13	40	0.0250	2.937	0.009		0.000
14	40	-1.9750	2.937	-0.672	*	0.048
15	40	3.0250	2.937	1.030	**	0.112
16	40	-3.9750	2.937	-1.353	**	0.194

Sum of Residuals 0
Sum of Squared Residuals 146.4250
Predicted Resid SS (Press) 194.0132

Plot of HARDNESS*TIME. Legend: A = 1 obs, B = 2 obs, etc.



PROGRAM AND OUTPUT FOR PROBLEM 4

Correlation Analysis

4 'VAR' Variables: TASTE ACETIC H2S LACTIC

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
TASTE	30	24.5333	16.2554	736.0	0.7000	57.2000
ACETIC	30	5.4980	0.5709	164.9	4.4770	6.4580
H2S	30	5.9418	2.1269	178.3	2.9960	10.1990
LACTIC	30	1.4420	0.3035	43.2600	0.8600	2.0100

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 30

	TASTE	ACETIC	H2S	LACTIC
TASTE	1.00000 0.0	0.54954 0.0017	0.75575 0.0001	0.70424 0.0001
ACETIC	0.54954 0.0017	1.00000 0.0	0.61796 0.0003	0.60378 0.0004
H2S	0.75575 0.0001	0.61796 0.0003	1.00000 0.0	0.64481 0.0001
LACTIC	0.70424 0.0001	0.60378 0.0004	0.64481 0.0001	1.00000 0.0

Dependent Variable: TASTE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	4994.47558	1664.82519	16.221	0.0001
Error	26	2668.41109	102.63120		
C Total	29	7662.88667			

Root MSE	10.13071	R-square	0.6518
Dep Mean	24.53333	Adj R-sq	0.6116
C.V.	41.29364		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-28.876770	19.73541835	-1.463	0.1554
ACETIC	1	0.327741	4.45975656	0.073	0.9420
H2S	1	3.911841	1.24843029	3.133	0.0042
LACTIC	1	19.670543	8.62905483	2.280	0.0311

Dependent Variable: TASTE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	4993.92131	2496.96066	25.260	0.0001
Error	27	2668.96535	98.85057		
C Total	29	7662.88667			

Root MSE	9.94236	R-square	0.6517
Dep Mean	24.53333	Adj R-sq	0.6259
C.V.	40.52593		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-27.591815	8.98182528	-3.072	0.0048
H2S	1	3.946267	1.13569217	3.475	0.0017
LACTIC	1	19.887204	7.95900907	2.499	0.0188