Stat 346 Homework 5

Nathan Jarus

Mar. 20, 2014

1 Problem 1

1.1 Part a

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \\ 1 & 8 \\ 1 & 10 \end{bmatrix}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 6 & 8 & 10 \end{bmatrix}$$

1.2 Part b

$$(X'X)^{-1} = \begin{bmatrix} 1.1 & -0.15 \\ -0.15 & 0.025 \end{bmatrix}$$

1.3 Part c

$$(X'X)^{-1}X'Y = b = \begin{bmatrix} -1.5 \\ 1.05 \end{bmatrix}$$

$b_0 = -1.5$
$b_1 = 1.05$

1.4 Part d

$$s^2\{b\} = \begin{bmatrix} 0.253 & -0.0345 \\ -0.0345 & 0.00575 \end{bmatrix}$$

$s^2\{b_0\} = 0.253$
$s^2\{b_1\} = 0.00575$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	627.81699	209.27233	10.22087	<0.0001
Error	20	61.425	3.07125		
Corrected Total	23	689.26			

1.5 Part e

$b_1 \pm t_{0.975}(3) * s^2\{b_1\}$
1.05 ± 0.01829
Assuming the SLR assumptions hold, there is a >95% probability that there is a significant linear relationship.

2 Problem 2, KNN # 6.1 a

$$X = \begin{bmatrix} 1 & X_{11} & X_{11}X_{12} \\ 1 & X_{21} & X_{21}X_{22} \\ 1 & X_{31} & X_{31}X_{32} \\ 1 & X_{41} & X_{41}X_{42} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

3 Problem 3

3.1 Part a

3.2 Part b

$R^2 = 0.9108$
$R_a^2 = 0.89752$

R_2 would increase if we introduced another prediction variable, but R_a^2 would only change if adding that variable gave a better model fit.

3.3 Part c

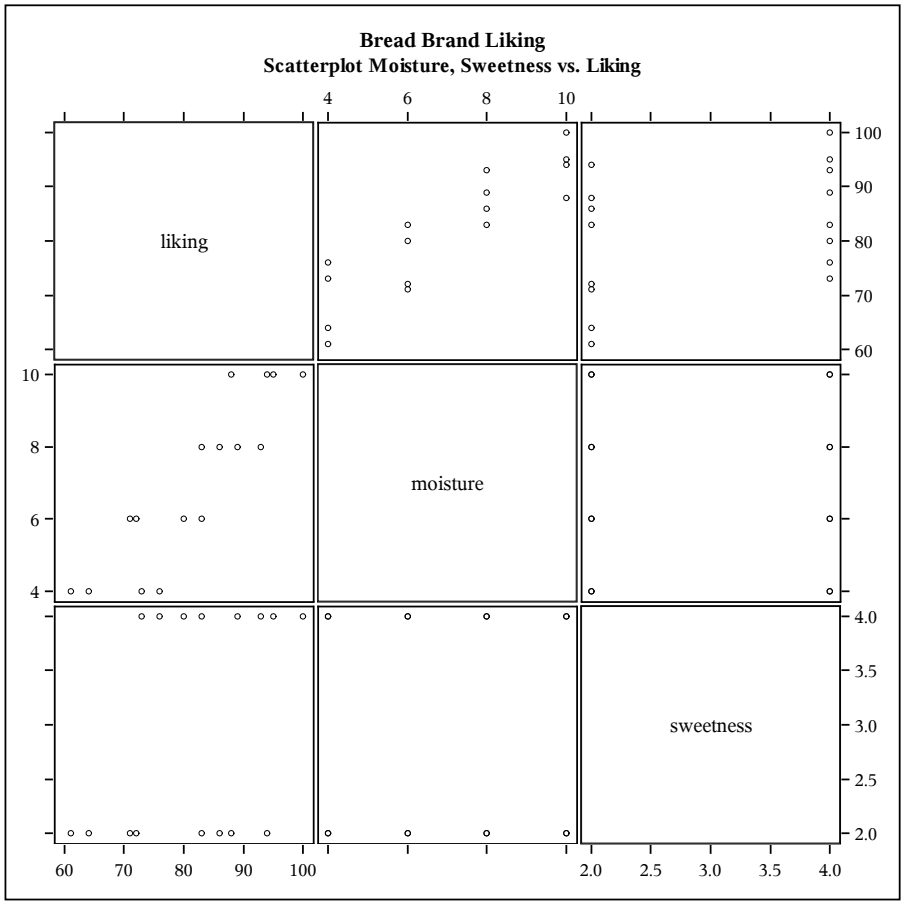
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
H_a : At least one of $\beta_1, \beta_2, \beta_3$ are not 0.

Since our F value, 10.22087, is greater than $F(0.995; 3, 20) = 6.48$, we reject H_0 and assert that there is a linear relationship between at least one predictor variable and the response variable. This test does not tell us which response variable(s) have a linear relationship, however.

4 Problem 4

4.1 Part a

11:54 Thursday, March 20, 2014 1



Bread Brand Liking

Pairwise correlations

The CORR Procedure

3 Variables:	liking moisture sweetness

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**
liking	16	81.75000	11.45135	1308	61.00000	100.00000
moisture	16	7.00000	2.30940	112.00000	4.00000	10.00000
sweetness	16	3.00000	1.03280	48.00000	2.00000	4.00000

Pearson Correlation Coefficients, N = 16				**Prob > |r| under H0: Rho=0**		
	liking	**moisture**	**sweetness**	**Pliking**	**Pmoisture**	**Psweetness**
liking	1.00000	0.89239	0.39458		<.0001	0.1304
moisture	0.89239	1.00000	0.00000	<.0001		1.0000
sweetness	0.39458	0.00000	1.00000	0.1304	1.0000	

These diagnostics give us a basic idea of whether or not we can expect a linear relationship between the variables.

4.2 Part b

Bread Brand Liking

Moisture Regression

The REG Procedure

Model: MODEL1

Dependent Variable: liking

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	1566.45000	1566.45000	54.75	<.0001
Error	14	400.55000	28.61071		
Corrected Total	15	1967.00000			

Root MSE	5.34890	**R-Square**	0.7964
Dependent Mean	81.75000	**Adj R-Sq**	0.7818
Coeff Var	6.54300		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**99% Confidence Limits**	
Intercept	1	50.77500	4.39457	11.55	<.0001	37.69306	63.85694
moisture	1	4.42500	0.59803	7.40	<.0001	2.64477	6.20523

$\hat{Y} = 50.775 + 4.425X_1$ The value of b_1 means that for every unit of moisture the bread has, people will like it 4.425 more.

4.3 Part c

Bread Brand Liking

Sweetness Regression

The REG Procedure

Model: MODEL1

Dependent Variable: liking

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	306.25000	306.25000	2.58	0.1304
Error	14	1660.75000	118.62500		

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Corrected Total	15	1967.00000			

Root MSE	10.89151	**R-Square**	0.1557
Dependent Mean	81.75000	**Adj R-Sq**	0.0954
Coeff Var	13.32295		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**99% Confidence Limits**	
Intercept	1	68.62500	8.61050	7.97	<.0001	42.99291	94.25709
sweetness	1	4.37500	2.72288	1.61	0.1304	−3.73058	12.48058

$\hat{Y} = 68.625 + 4.375X_2$

4.4 Part d

Bread Brand Liking

Multiple Regression

The REG Procedure

Model: MODEL1

Dependent Variable: liking

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	1872.70000	936.35000	129.08	<.0001
Error	13	94.30000	7.25385		
Corrected Total	15	1967.00000			

Root MSE	2.69330	**R-Square**	0.9521
Dependent Mean	81.75000	**Adj R-Sq**	0.9447
Coeff Var	3.29455		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**99% Confidence Limits**	
Intercept	1	37.65000	2.99610	12.57	<.0001	28.62491	46.67509
moisture	1	4.42500	0.30112	14.70	<.0001	3.51794	5.33206
sweetness	1	4.37500	0.67332	6.50	<.0001	2.34676	6.40324

$\hat{Y} = 37.65 + 4.425X_1 + 4.375X_2$

The value of b_1 means that for every unit of moisture has, provided sweetness stays constant, people will like that bread 4.425 more.

4.5 Part e

The intercept changed, but the coefficients for the variables stayed the same. This probably means that the variables are explaining different parts of why people like bread, indicating that the multiple regression should be more accurate than a single regression on either variable.

4.6 Part f

$H_0 : \beta_1 = \beta_2 = 0$
H_a : At least one of β_1, β_2 are not 0.

$F = MSR/MSE$ Since our F value, 129.08, is greater than $F(0.999; 2, 13)$, we reject H_0 and assert that there is a linear relationship between at least one predictor variable and the response variable.

4.7 Part g

Moisture:$(3.51794, 5.332)$
Sweetness:$(2.346766.403)$

We are 99% confident that increasing moisture one unit will cause people to like the bread between 3.51 and 5.32 more. Also, since 0 is not in the confidence interval, we are >99% confident that there is a significant linear relationship between liking bread and moisture content.

4.8 Part h

For Moisture: $H_0 : b_1 = 4.425$
$H_a : b_1 \neq 4.425$
$t = \frac{b_1 - 4.425}{s\{b_1\}}$
For Sweetness: $H_0 : b_2 = 4.375$
$H_a : b_2 \neq 4.375$

$t = \frac{b_2 - 4.425}{s\{b_2\}}$
For both: 13 degrees of freedom
We fail to reject H_0, indicating that both variables contribute to explaining the total variance.

4.9 Part i

$R^2 = 0.9521$
This is the percentage of variation explained by all variables in the model.

4.10 Part j

The multiple regression's adjusted R^2 is the highest at 0.944, meaning that it explains more variation in liking bread than either variable on its own. This is as we predicted in Part e.