STAT 515 - Fall 2003 - Solutions to the Practice 2 Homework

Page 522: 11.13c, also calculate the SSE and MSE.

SS_{xx}=43.4286, SS_{xy}=39.8571, $\bar{y} = 3.4286$, $\bar{x} = 3.7143$

Using the equations on page 516 we get:

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{39.8571}{43.4286} = 0.9178 \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 3.4286 - 0.9178(3.7143) = 0.0196$$

For the SSE we can either do it the long way like in table 11.4 on page 518, or we can use the shortcut formula on page 531 (which means we need SS_{yy}).

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 41.7143 - (0.9178)39.8571 = 5.1335$$

$y - \bar{y}$	$(y - \bar{y})^2$
4-3.4286	0.3265
3-3.4286	0.1837
0-3.4286	11.7553
1-3.4286	5.8981
8-3.4286	20.8977
5-3.4286	2.4693
3-3.4286	0.1837
	sum=41.7143

MSE = SSE/n-2 = 5.1335/5 = 1.0267

Page 526: 11.19 using SAS. Also, check that the assumptions for performing regression are met, conduct a test of the null hypothesis that beta_1=0, and verify that the degrees of freedom in the ANOVA table are correct.

```
DATA oj;
INPUT sweet pect @@;
CARDS;
5.2   220   5.5   227   6.0   259   5.9   210   5.8   224
5.6   268   5.6   239   5.9   212   5.4   410   5.6   256
5.8   306   5.5   259   5.3   284   5.3   383   5.7   271
5.5   264   5.7   271   5.5   264   5.7   227   5.3   263
5.9   232   5.8   220   5.8   246   5.9   241
;
PROC INSIGHT;
OPEN oj;
FIT sweet = pect;
RUN;
```

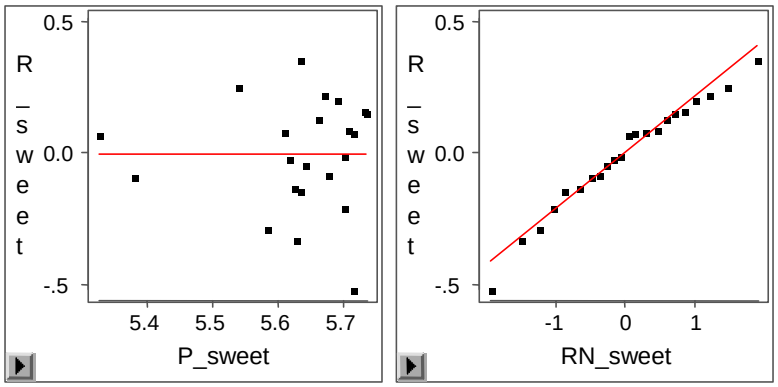
a)

Model Equation				
sweet	=	6.1685	-	0.0021 pect

b) The intercept tells us that (if we could trust the regression line out that far!) an orange with zero pectin would have a sweetness of 6.1685. For every ppm the pectin goes up, the sweetness goes down by 0.0021.

c) 6.1685-300(0.0021)=5.5385

Assumptions)



i) We are not told that the orange samples were random, so we can't be sure of the assumption that the errors are independent.
ii) From the residual vs. predicted plot it does look like a line is appropriate... at each value of predicted sweetness, it looks like the mean of the residuals is around zero.
iii) From the residual vs. predicted plot it is not clear if the variance of the residuals is constant, it looks a little wider (up and down) on the right side than the left... but that could just be because there are only two points on the left hand side.
iv) The errors appear to be normally distributed from the residual vs. predicted plot.

Test and degrees of freedom)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Model	1	0.2273	0.2273	5.18	0.0330
Error	22	1.0193	0.0439		
C. Total	23	1.2465			

The p-value of 0.0330 is less than α=0.05 so we reject the null hypothesis. Pectin can be used to predict sweetness. There are p=2 groups and n=24 observations and the df are indeed p-1=1, n-p=22, and n-1=23.

Page 548: 11.54 (both what r and r^2 tell us)

a) There is a relationship, but it is not a very strong relationship since the math confidence only explains 1.96% (r^2) of the variation/error in the learning of computer skills.

b) There is a stronger relationship between a girls math confidence and computer interest, but it is still not that strong a relationship since the math confidence only explains 10.89% (r^2) of the variation/error in the learning of computer skills.