Stat 346 Homework 2

Nathan Jarus

Feb. 13, 2014

1 Problem 1, KNN #2.10

2 Problem 2

2.1 Part a

The confidence interval for β_1 is defined as follows:

$$\beta_1 \pm t_{1-\alpha\backslash 2,n-1}(s(b_1)) \tag{1}$$

Substituting the provided values, we get:

$$3 \pm 2.09302 \tag{2}$$

2.2 Part b

$$H_0 : \beta_1 = 4 \tag{3}$$

$$H_a : \beta_1 \neq 4 \tag{4}$$

Based on this, we can set up a test statistic:

$$t^* = \frac{b_1 - 4}{s(b_1)} = -1 \tag{5}$$

How does one calculate the p value?
Ultimately, I think we fail to reject H_0.

3 Problem 3

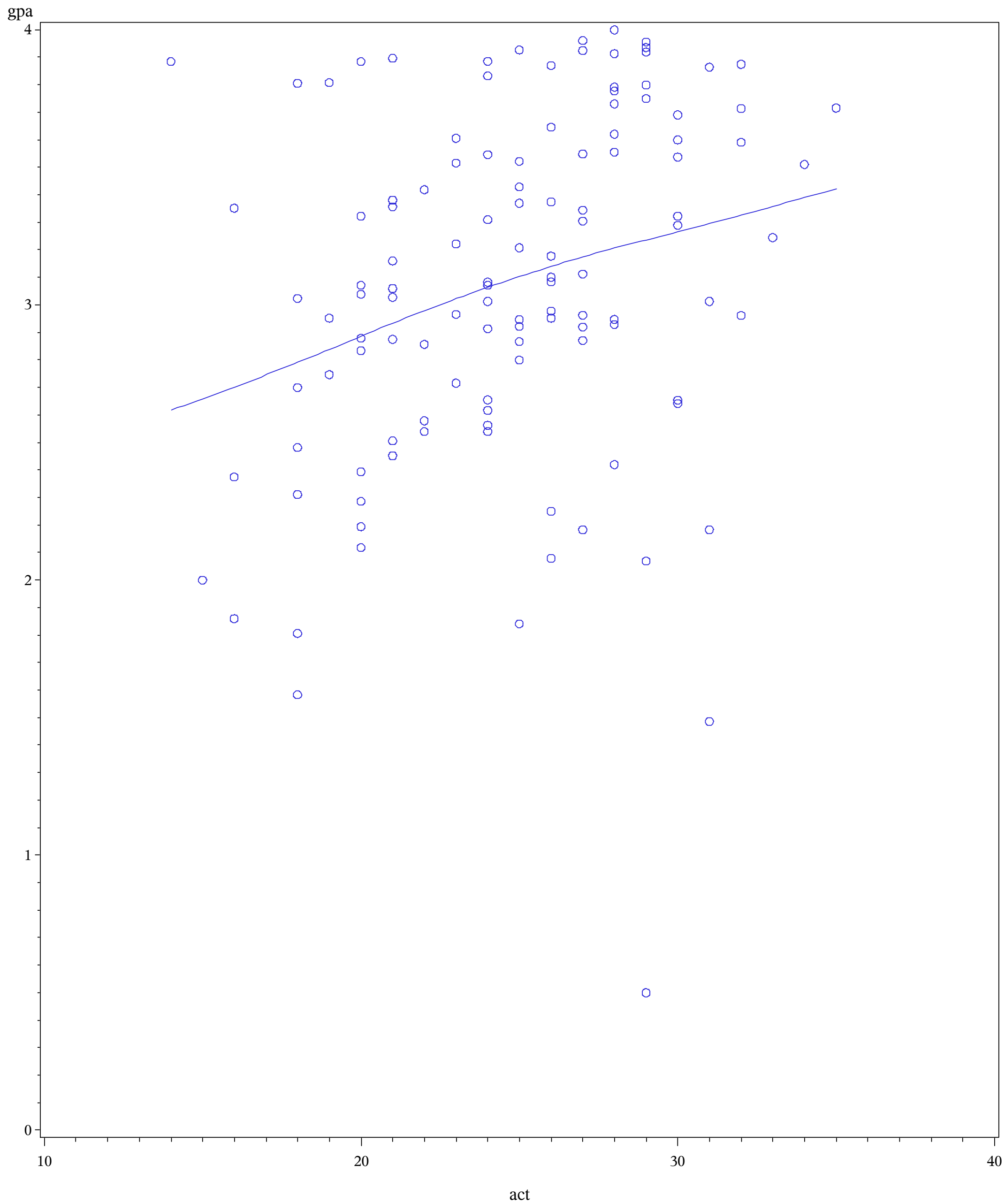
3.1 Part a

The scatterplot is shown in Figure 3.1 With sufficient smoothing, the relationship is vaguely linear.

16:58 Wednesday, February 12, 2014 **1**

GPA and ACT

Scatter plot of ACT Score vs. GPA with Smoothing Line N=80



3.2 Part b

GPA and ACT

Scatter plot of ACT Score vs. GPA with Smoothing Line N=80

The REG Procedure

Model: MODEL1

Dependent Variable: gpa

Number of Observations Read	120
Number of Observations Used	120

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	3.58785	3.58785	9.24	0.0029
Error	118	45.81761	0.38828		
Corrected Total	119	49.40545			

Root MSE	0.62313	**R-Square**	0.0726
Dependent Mean	3.07405	**Adj R-Sq**	0.0648
Coeff Var	20.27049		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**95% Confidence Limits**	
Intercept	1	2.11405	0.32089	6.59	<.0001	1.47859	2.74951
act	1	0.03883	0.01277	3.04	0.0029	0.01353	0.06412

The estimated regression equation is

$$\hat{Y} = 2.11405 + 0.03883 * \hat{X} \tag{6}$$

For the slope, the 95% confidence limits are 0.01353 and 0.06412. Since they are both positive, we can conclude with confidence > 95% that there is a positive linear relationship between GPA and ACT scores.

3.3 Part c

A significance test for the slope, b_1 uses the hypothesis

$$H_0 : b_1 = 0 \tag{7}$$

$$H_a : b_1 \neq 0 \tag{8}$$

Our test statistic is 3.04 with 118 DOF and the resulting p-value is 0.0029. Since p < 0.05, we reject H_0, meaning that there is a significant linear relationship between GPA and ACT scores.

3.4 Part d

The lowest X value in the data set is 14, so an ACT score of 0 (the intercept) would not be within the scope of the model.

3.5 Part e

Our mean response for an ACT score of 21 is a GPA of 2.92948. The 95% confidence interval is $(2.7826, 3.0763)$.

3.6 Part f

The prediction is the same as for Part E, 2.92948. However, the confidence interval is $(1.6868, 4.1721)$.

3.7 Part g

Percent of variance accounted for by ACT score:

$$\frac{3.58785}{49.40545} * 100 = 7.26205\% \tag{9}$$

Correlation: 0.26948

3.8 Part h

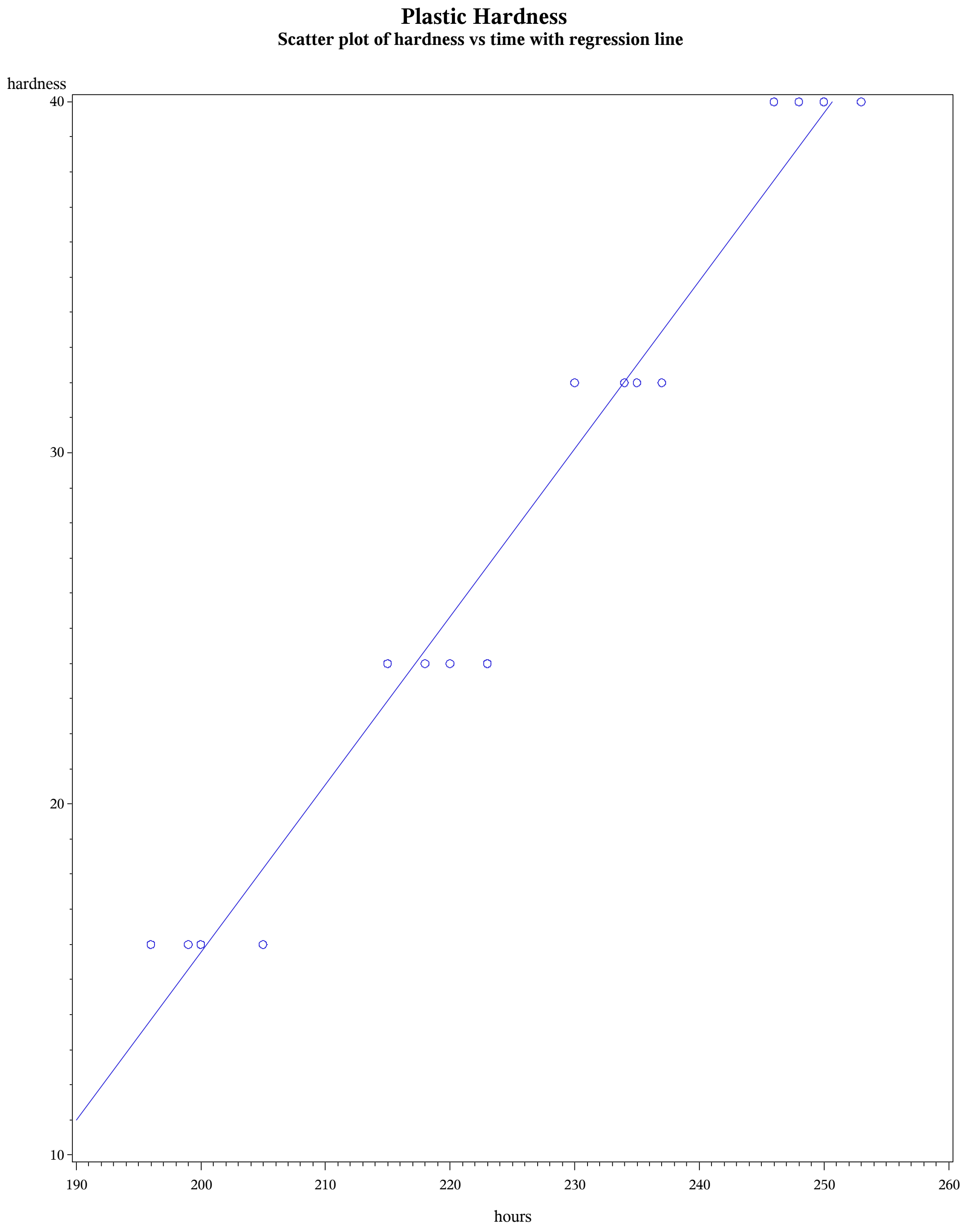
I do not think that ACT scores make a good predictor for GPA. The confidence intervals on predictions are very wide, and ACT score accounts for only 7% of the variance in GPA.

4 Problem 4

4.1 Part a

As shown in Figure 4.1, the data definitely follows a linear pattern, but seems to be sorted into buckets. This may demonstrate a nonlinear relationship between hardness and time, or it may be a limitation of the measurement method for hardness.

12:44 Thursday, February 13, 2014 **1**



4.2 Part b

Plastic Hardness

Scatter plot of hardness vs time with regression line

The REG Procedure

Model: MODEL1

Dependent Variable: hardness

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	1245.57198	1245.57198	506.51	<.0001
Error	14	34.42802	2.45914		
Corrected Total	15	1280.00000			

Root MSE	1.56817	**R-Square**	0.9731
Dependent Mean	28.00000	**Adj R-Sq**	0.9712
Coeff Var	5.60059		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**95% Confidence Limits**	
Intercept	1	−79.89337	4.81005	−16.61	<.0001	−90.20990	−69.57683
hours	1	0.47833	0.02125	22.51	<.0001	0.43275	0.52392

Estimated regression equation:

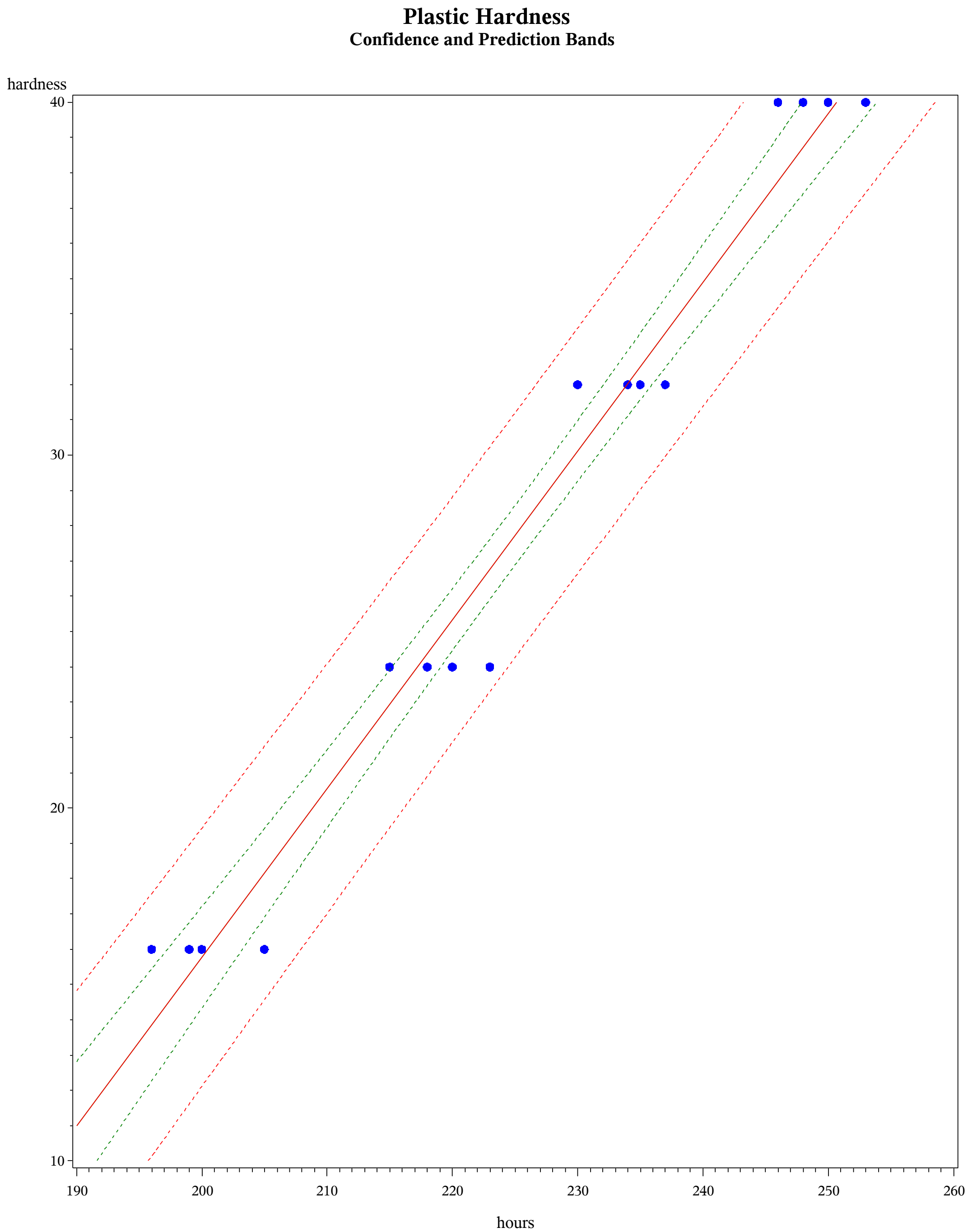
$$\hat{Y} = -79.89337 + 0.47833 * \hat{X} \tag{10}$$

The slope 95% confidence interval is $(0.43275, 0.52392)$.

4.3 Part c

As shown in Figure 4.3, the prediction band is wider to account for the extra variation due to the fact that new observations will probably not fall on the regression line.

12:44 Thursday, February 13, 2014 **1**



Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	45.20	45.20	43.88	<0.0001
Error	70	72.1	1.03		
Total	71	117.3			

Tab. 9: Problem 5 ANOVA Table

4.4 Part d

97.31031% of the variance in hardness is due to time. The correlation coefficient between hardness and time is 0.98646.

4.5 Part e

Time seems to be a very good predictor for plastic hardness.

5 Problem 5

$$H_0 : b_1 = 0 \tag{11}$$

$$H_a : b_1 \neq 0 \tag{12}$$

From the values in table 9, we can reject the null hypothesis and conclude that we have a reasonably good fit.