Problem 2.26

(b) Test by means of an F test whether or not there is a linear association between the hardness of the plastic and the elapsed time. Use a = .01. State the alternatives, decision rule, and conclusion.

Testing H_0: $\beta_1 = 0$ vs H_a: $\beta_1 \neq 0$. Reject H_0 if P-value < 0.01. Since P-value < 0.0001, we reject H_0 and conclude that there is a significant linear relationship between hardness and time.

(d) Calculate r^2 and r.

The values of r^2 and r can be obtained from the SAS output: $r^2 = 0.9731$ and so $r = 0.9865$. (Choose the positive square root because $b_1 > 0$).

Root MSE	3.23403	R-Square	0.9731
Dependent Mean	225.56250	Adj R-Sq	0.9712
Coeff Var	1.43376		

2. Problem 2.31

a)

Source	*SS*	*df*	*MS*
Regression	93,462,942	1	93,462,942
Error	455,273,165	82	5,552,112
Total	548,736,107	83	

b). $H_0: \beta_1 = 0$, $H_a: \beta_1 = 0$. $F^* = 93,462,942 / 5,552,112 = 16.8338$, $F(.99; 1, 82) = 6.9544$. If $F^* \leq 6.9544$ conclude H_0, otherwise H_a. Conclude H_a. $(t^*)^2 = (-4.102895)^2 = 16.8338 = F^*$. $[t(.995; 82)]^2 = (2.63712)^2 = 6.9544 = F(.99; 1, 82)$

so the two tests are, indeed, equivalent

c) $SSR = 93,462,942$ which is 17.03% or 0.1703

d) -0.4127

7. *Describe the distribution of the explanatory variable. Show the plots and output that were helpful in learning about this variable.*

Using PROC UNIVARIATE we see there are 20 observations ranging between 3.9 and 6.3 with a mean of 5 and median of 4.85; their standard deviation is 0.69. There are no extreme observations (i.e., ones far away from the others) as shown in the histogram plot below. The

distribution appears to be reasonably symmetric but not completely so; it has a slight skew to the right.

The UNIVARIATE Procedure
Variable: testscore

Moments

N	20	Sum Weights	20
Mean	5	Sum Observations	100
Std Deviation	0.69282032	Variance	0.48
Skewness	0.49695171	Kurtosis	-0.595083
Uncorrected SS	509.12	Corrected SS	9.12
Coeff Variation	13.8564065	Std Error Mean	0.15491933

Basic Statistical Measures

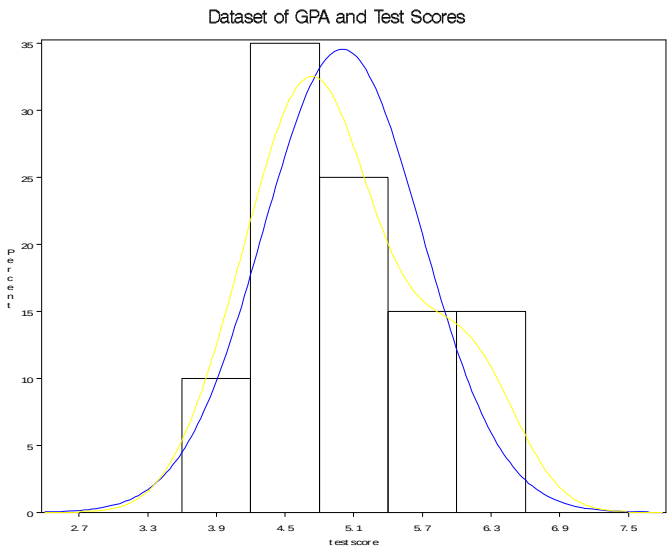
Location		Variability	
Mean	5.000000	Std Deviation	0.69282
Median	4.850000	Variance	0.48000
Mode	4.700000	Range	2.40000
		Interquartile Range	0.90000

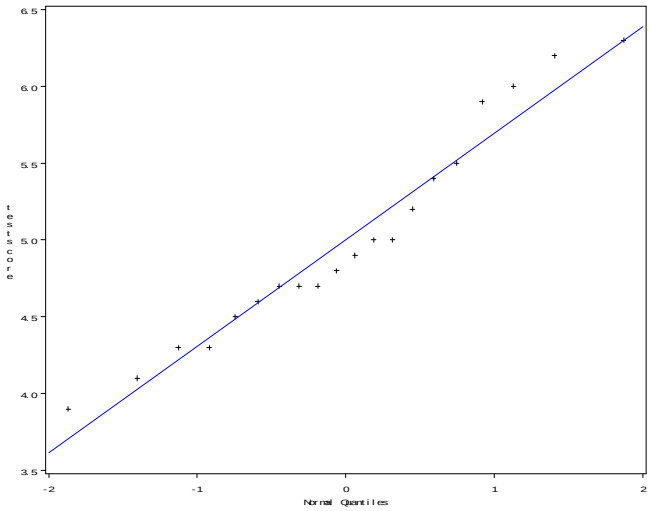
Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
3.9	4	5.5	1
4.1	19	5.9	18
4.3	16	6.0	7
4.3	10	6.2	6
4.5	5	6.3	14

Stem	Leaf	#	Boxplot
6	023	3	|
5	59	2	|
5	0024	4	+--+--+
4	5677789	7	*-----*
4	133	3	|
3	9	1	|

----+----+----+----+





8. *Run the linear regression to predict GPA from the entrance test score and obtain the residuals (do not include a list of the residuals in your solution).*

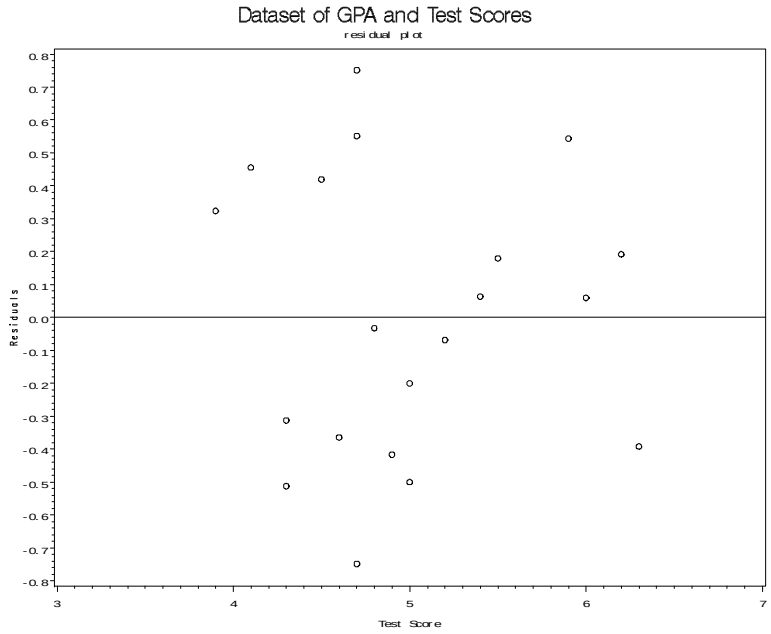
   a. *Verify that the sum of the residuals is zero by running* `PROC UNIVARIATE` *with the output from the regression.*

The UNIVARIATE Procedure
Variable: resid (Residual)

Moments

N	20	Sum Weights	20
Mean	0	Sum Observations	0
Std Deviation	0.4234117	Variance	0.17927747
Skewness	0.05677081	Kurtosis	-1.0342174
Uncorrected SS	3.40627193	Corrected SS	3.40627193
Coeff Variation	.	Std Error Mean	0.09467773

   b. *Plot the residuals versus the explanatory variable and briefly describe the plot noting any unusual patterns or points.*



There does not appear to be any obvious pattern or outlier in this residual plot. It looks like a random scatter of points and the variance is reasonably constant.

c. Plot the residuals versus the order in which the data appear in the data file and briefly describe the plot noting any unusual patterns or points.

There is some suggestion that the residuals may become more negative over time but it does not appear to be a big problem.