

NAME: _____

1. Figures 1 and 2 display SAS output from the regression of scores from the first test (test1) on scores from the first homework (hw1) for the 41 students in a statistics course.

- (a) (10 points) What is the response? The predictor? The regressor?

ANS: *Response: TEST1; predictor, regressor: HW1*

- (b) (10 points) Write out the equation of the fitted model. Interpret the slope.

ANS:

$$\hat{TEST1} = 39.7213 + 0.4942HW1$$

The change in predicted TEST1 per unit change in HW1 is 0.4942

- (c) (10 points) Is it wise to interpret the intercept? Why or why not? If it is, Interpret the intercept.

ANS: *It is not wise, since a HW1 value of 0 is well outside the range of the data.*

- (d) (10 points) Construct a 95% confidence interval for the slope. Does the result indicate that there is a significant relationship between TEST1 and HW1?

ANS: *Since the t table in the text does not have entries for 39 degrees of freedom, we will use $t_{40,0.975} = 2.0211$.*

$$0.4942 \pm (0.1338)(2.0211) = (0.22377, 0.76462)$$

The interval does not contain 0, so it indicates a significant relationship between TEST1 and HW1.

- (e) (10 points) By what proportion is the uncertainty in predicting the response reduced by using the regression model?

ANS: 0.2591

- (f) (10 points) What is the Pearson correlation between TEST1 and HW1?

ANS: 0.5090

- (g) (10 points) Evaluate the quality of the fit.

The plot of residuals versus fitted values offers no evidence of a lack of fit. The plots of the studentized residuals do not give reason to doubt the assumption of normality.

- (h) (10 points) Estimate the standard deviation of the random errors.

ANS: $\hat{\sigma} = \sqrt{MSE} = 14.9238$.

- (i) (10 points) Obtain a point estimate and a level 0.95 confidence interval for the mean response for a student having a 60 on the homework, if the mean and variance of the homework scores are 68.0304 and 310.9280, respectively.

ANS: *A point estimate is $39.7213 + (0.4942)(60) = 69.373$. The estimated standard error is*

$$\hat{\sigma}(\hat{Y}) = \sqrt{222.72 \left[\frac{1}{41} + \frac{(60 - 68.0304)^2}{(40)(310.9280)} \right]} = 2.5665.$$

Since the t table in the text does not have entries for 39 degrees of freedom, we will use $t_{40,0.975} = 2.0211$. The level 0.95 confidence interval for the mean response is then $69.373 \pm (2.5665)(2.0211) = (64.186, 74.561)$.

2. (10 points) From the article "Study Says Rigor of High-School Course Work Is the Best Predictor of College Graduation," by Ben Gose, Chronicle of Higher Education, XLV, 39, June 4, 1999, p. A46, which details a study conducted by the U.S. Dept. of Education purporting to show that the rigor of a student's high school curriculum is a better predictor of whether the student will graduate from college than either test scores or high school grades. The following is a quote from the article:

Clifford Adelman, the senior research analyst who led the study, spent two years examining the transcripts and test scores of students who graduated from high school in 1982 to determine what contributes most to the completion of a bachelor's degree.

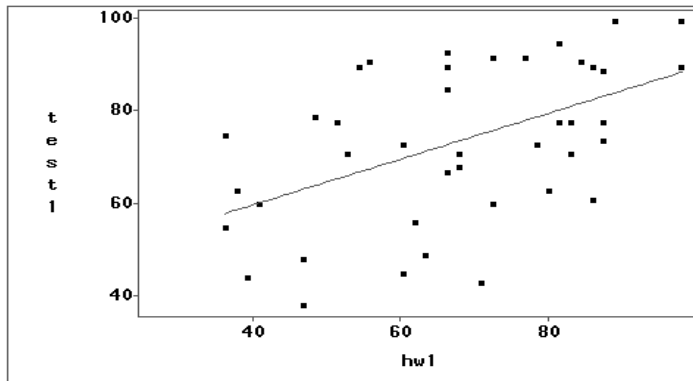
Mr. Adelman established college graduation by the age of 30 as a benchmark. He then worked backward to see which measure of preparedness for college-the student's test scores, class rank, or course work- had the strongest correlation with the earning of a bachelor's degree.

The study was reviewed by Alberto F. Cabrera, an associate professor and research associate at the Center for the Study of Higher Education at Pennsylvania State University. Mr. Cabrera says the study should persuade high-school guidance counselors to encourage as many students as possible to take demanding courses. "The findings make a clear point that the type of courses a student takes has a long-term effect on the student's success," Mr. Cabrera says.

Do you agree with Mr. Cabrera? If so, what in the study's description substantiates his conclusion? If not, tell why his conclusion is unjustified.

ANS: *As described, the study does not support Mr. Cabrera's conclusion. Mr. Cabrera is confusing association with causation. Let us assume that Mr. Adelman found a strong correlation that is indicative of association between high school curriculum rigor and college graduation (setting aside the question of whether obtaining correlations in an exploratory manner can provide convincing evidence of association). This does not show that the former causes the latter. It may be that those who take rigorous courses are good students who are likely to graduate from college. If so, then requiring rigorous high school courses may not ensure higher college rates for mediocre students.*

Model Equation		
test1	=	39.7213 + 0.4942 hw1



Parametric Regression Fit								
Curve	Degree(Polynomial)	Model		Error		R-Square	F Stat	P
		DF	Mean Square	DF	Mean Square			
	1	1	3037.1519	39	222.7197	0.2591	13.64	

Summary of Fit			
Mean of Response	73.3415	R-Square	0.2591
Root MSE	14.9238	Adj R-Sq	0.2401

Parameter Estimates							
Variable	DF	Estimate	Std Error	t Stat	Pr > t	Tolerance	Var Inflation
Intercept	1	39.7213	9.3979	4.23	0.0001	.	0
hw1	1	0.4942	0.1338	3.69	0.0007	1.0000	1.0000

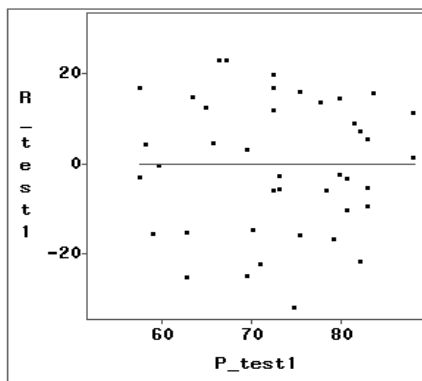


Figure 1: Regression output 1.

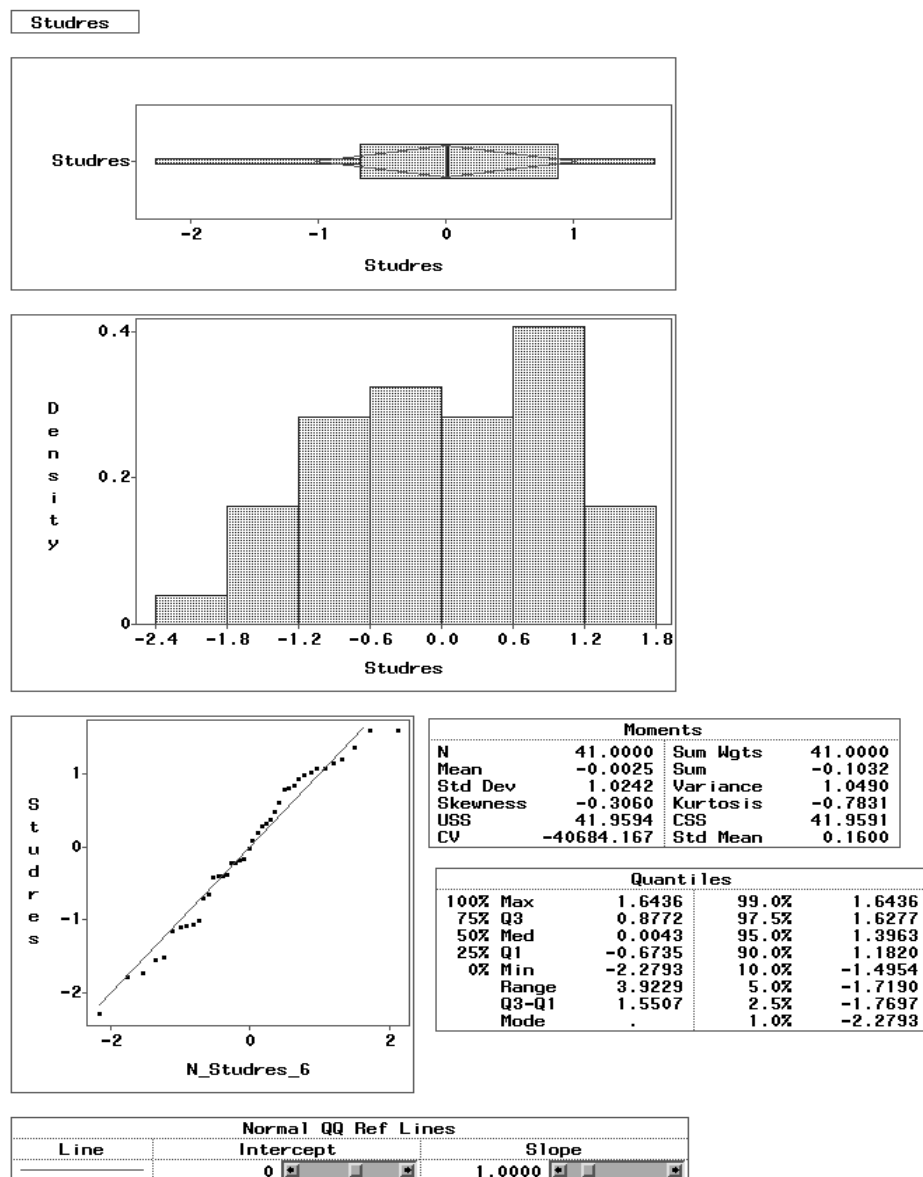


Figure 2: *Regression output 2.*