

1. (25 pts) Is a college's average SAT score a good predictor of what it charges for tuition? A random sample of 10 colleges produced the following data¹.

College	SAT-Math (x)	Tuition (in \$1000s) (y)
Bennington College	560	33
Calvin College	605	19
Queens College	535	11
Flagler College	560	9
Grove City College	636	10
Indiana Univ. of Pa.	534	12
Lafayette College	665	30
Manhattanville College	550	25
Marlboro College	590	27
Mercer University	587	23

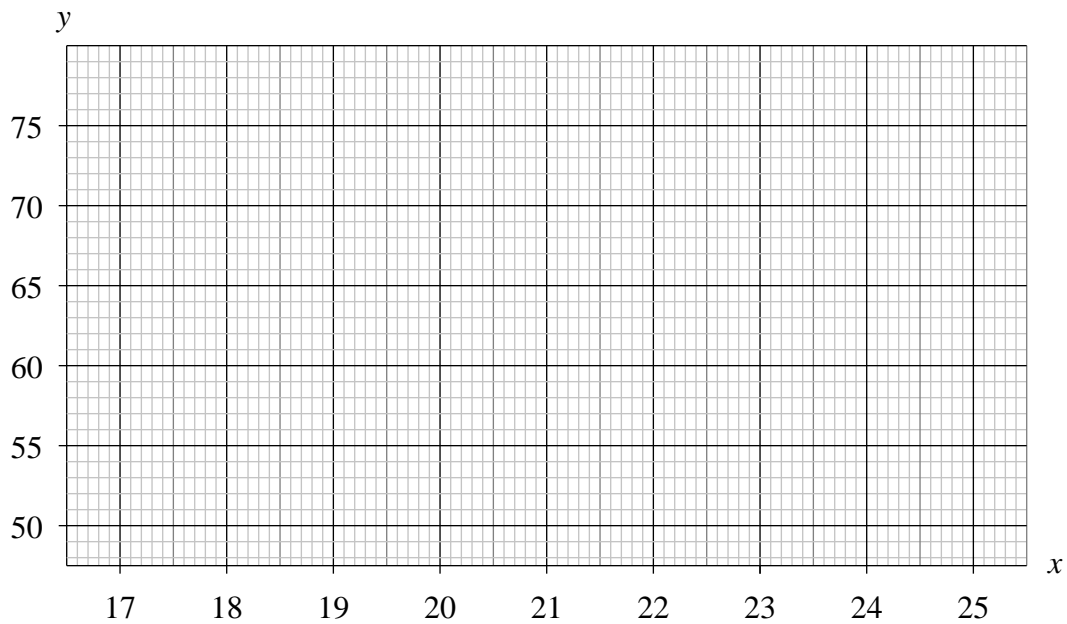
- (a) (5 pts) Use these data to find the least-squares regression line. Write the equation.
- (b) (3 pts) Find the correlation coefficient.
- (c) (3 pts) What does this correlation coefficient in this problem tell us about the association between these two variables?
- (d) (4 pts) *According to the model*, an increase of 10 in the average SAT-Math score for a college would correspond to what change in the expected tuition?
- (e) (4 pts) The average SAT-Math score at Hampden-Sydney College is 573. Use this value and the model to predict the tuition at HSC. (According to CNNMoney.com, HSC's tuition is \$21,878.)
- (f) (3 pts) In this model, which is the explanatory variable and which is the response variable?
- (g) (3 pts) Someone might object to this model by pointing out that it does not take into account important factors such as whether the college is private or the college's location (private colleges and colleges in the northeast are typically more expensive). What is the statistical term for these factors?
2. (32 pts) The Project on Student Debt recently published a report on student debt after graduation from college in 2007. The report included figures on the average amount of debt by state (average for Virginia is \$18,084) and percentage of students who graduate with debt (percentage for Virginia is 59%). A random sample of 10 states provides the following data.²

¹Princeton Review and CNNMoney.com

²www.projectonstudentdebt.org

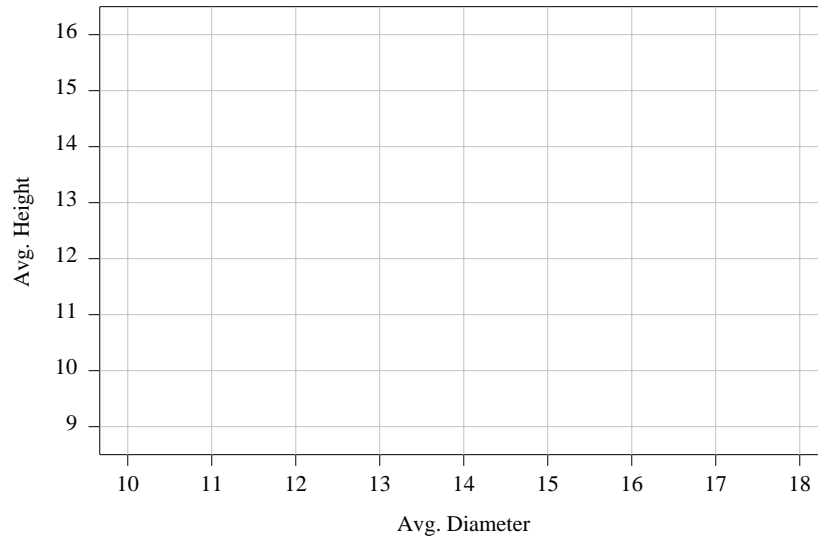
State	Average Debt (\$1000s) (x)	Percentage with Debt (y)
Alaska	25.0	53
Alabama	20.9	61
Delaware	17.4	50
Indiana	21.3	60
Kansas	18.5	61
Massachusetts	21.1	63
Ohio	22.0	67
Pennsylvania	23.6	71
Rhode Island	23.2	67
South Carolina	20.2	59

- (a) (4 pts) On the graph paper below, draw a scatter plot of the data.
- (b) (3 pts) Based on your scatter plot, describe the relationship between average debt and percentage of graduates with debt.
- (c) (5 pts) Find the equation of the regression line.
- (d) (4 pts) Find the correlation coefficient. What does it tell us about the relationship?
- (e) (3 pts) Find the coefficient of determination. What does it tell us about the relationship?
- (f) (4 pts) Given that $SST = 629.6$, find SSR and SSE .
- (g) (4 pts) The average debt in Alabama is \$20,921. Use the regression line to predict the percentage of graduates with debt in Alabama. (The actually percentage is 61%.)
- (h) (5 pts) At the 5% level of significance, test for the significance of the model. That is, test the hypotheses $H_0 : \beta = 0$ and $\rho = 0$ vs. $H_1 : \beta \neq 0$ and $\rho \neq 0$.



3. (29 pts) One would expect the diameter of a tree to correlate with the height of the tree. In fact, one would expect a positive correlation and probably a strong one. The following table (of authentic data) shows the average diameter and average height of 8 different species of northern trees.

Species	Avg. Diameter (cm)	Avg. Height (m)
Balsam fir	11.03	9.27
Balsam Poplar	17.40	15.45
Black spruce	13.25	11.51
Jack pine	16.13	13.92
Red pine	17.57	13.82
Trembling aspen	16.73	15.95
White birch	11.20	11.19
White spruce	14.19	10.46



- (a) (5 pts) Draw the scatterplot of these data in the graph paper above.
- (b) (5 pts) Find the equation of the regression line.
- (c) (3 pts) Graph *accurately* the regression line on the graph paper above.
- (d) (4 pts) Find the correlation coefficient.
- (e) (3 pts) Does the value of the correlation coefficient agree with our expectation that the correlation would be positive and strong? Explain.
- (f) (4 pts) Suppose we knew that the average diameter of an unspecified northern species of tree was 12 cm. Based on the model, what would be our best estimate of the average height of that species?
- (g) (5 pts) Recall that our measure of how well a line fits the data was the sum of squared residuals. Find the value of SSE for the regression line in part (c).