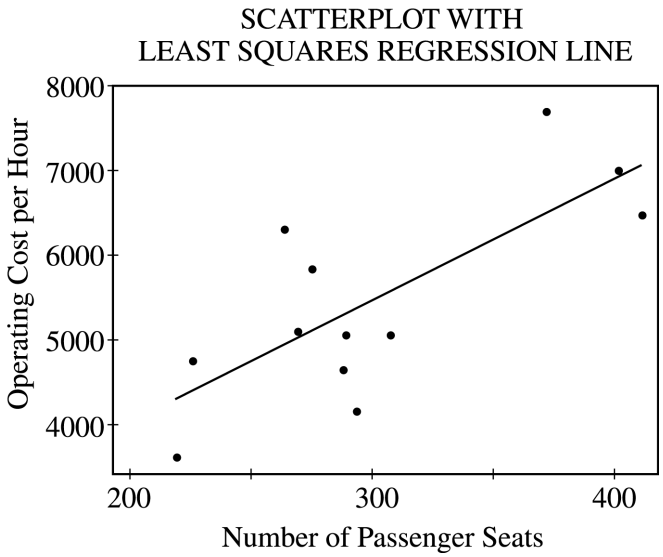
4. Commercial airlines need to know the operating cost per hour of flight for each plane in their fleet. In a study of the relationship between operating cost per hour and number of passenger seats, investigators computed the regression of operating cost per hour on the number of passenger seats. The 12 sample aircraft used in the study included planes with as few as 216 passenger seats and planes with as many as 410 passenger seats. Operating cost per hour ranged between $3,600 and $7,800. Some computer output from a regression analysis of these data is shown below.



Predictor	Coef	StDev	T	P
Constant	1136	1226	0.93	0.376
Seats	14.673	4.027	3.64	0.005
S = 845.3	R-Sq = 57.0%		R-Sq (adj) = 52.7%	

(a) What is the equation of the least squares regression line that describes the relationship between operating cost per hour and number of passenger seats in the plane? Define any variables used in this equation.

(b) What is the value of the correlation coefficient for operating cost per hour and number of passenger seats in the plane? Interpret this correlation.

(c) Suppose that you want to describe the relationship between operating cost per hour and number of passenger seats in the plane for planes only in the range of 250 to 350 seats. Does the line shown in the scatterplot still provide the best description of the relationship for data in this range? Why or why not?

Copyright © 2002 by College Entrance Examination Board. All rights reserved.
Advanced Placement Program and AP are registered trademarks of the College Entrance Examination Board.

5. Sleep researchers know that some people are early birds (E), preferring to go to bed by 10 P.M. and arise by 7 A.M., while others are night owls (N), preferring to go to bed after 11 P.M. and arise after 8 A.M. A study was done to compare dream recall for early birds and night owls. One hundred people of each of the two types were selected at random and asked to record their dreams for one week. Some of the results are presented below.

	Number of Dreams Recalled During the Week			Proportion Who Recalled	
Group	Mean	Median	Standard Deviation	No dreams	5 or more dreams
Early birds	7.26	6.0	6.94	0.24	0.55
Night owls	9.55	9.5	5.88	0.11	0.69

(a) The researchers believe that night owls may have better dream recall than do early birds. One parameter of interest to the researchers is the mean number of dreams recalled per week with μ_E representing this mean for early birds and μ_N representing this mean for night owls. The appropriate hypotheses would then be $H_0: \mu_E - \mu_N = 0$ and $H_a: \mu_E - \mu_N < 0$. State two other pairs of hypotheses that might be used to test the researchers' belief. Be sure to define the parameter of interest in each case.

(b) Use the data provided to carry out a test of the hypotheses about the mean number of dreams recalled per week given in the statement of part (a). Do the data support the researchers' belief ?

Copyright © 2002 by College Entrance Examination Board. All rights reserved.
Advanced Placement Program and AP are registered trademarks of the College Entrance Examination Board.

AP® STATISTICS
2002 SCORING GUIDELINES

Question 4

Solution

Part (a):

Predicted cost = 1136 + 14.673 (number of passenger seats)

OR

$\hat{y} = 1136 + 14.673x$ where y = operating cost per hour
and x = number of passenger seats

Part (b):

- The value of the correlation coefficient

  $r = +\sqrt{0.570} = 0.755$ (r is positive because the scatterplot shows a positive association)

- The interpretation of correlation

  There is a moderate (or strong) positive linear relationship between operating costs per hour and number of passenger seats.

  OR

  Fifty-seven percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats AND the relationship is positive.

Part (c):

No. The equation of the least-squares regression line is influenced by the three points in the upper right-hand corner and the two points in the lower left-hand corner of the scatterplot. The seven remaining points (with number of seats in the 250 to 350 range) would have a negative correlation. Hence, the slope of the recalculated least-squares regression line is negative.

Scoring

The student response should include the following elements:

1. the correct equation of the least squares regression line with variables correctly defined;
2. the correct value for the correlation coefficient;
3. a correct interpretation of the given correlation coefficient; and
4. a complete explanation of why the given least-squares line would not be appropriate for describing the relationship over the restricted range.

Each element is scored as essentially correct (E), partially correct (P), or incorrect (I).

Copyright © 2002 by College Entrance Examination Board. All rights reserved.

Question 4 (cont'd.)

Part (a) addresses the first element.

Element one is:

essentially correct if the solution has the correct equation and variables are defined correctly.

partially correct if only the equation is correct.

incorrect if the equation is not stated correctly.

Part (b) addresses the second and third elements.

Element two is:

essentially correct if the student's solution states that $r = 0.755$.

partially correct if the student's solution only states that $r = \pm 0.755$.

incorrect if the student states any other value of r including $r = 0.726$ (square root of R-Sq (adj)).

Element three is:

essentially correct if the student's solution

addresses, based on a correct understanding of the correlation coefficient, three or four of the following:

- type of relationship
- strength
- direction
- context

OR

states, based on a correct understanding of r^2:

- that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats

AND

- that the relationship is positive.

Note: If the student gives a correct interpretation of r but then incorrectly explains r^2, this is considered a parallel solution and is incorrect.

Copyright © 2002 by College Entrance Examination Board. All rights reserved.

Question 4 (cont'd.)

partially correct if the student's solution

addresses exactly two of the following – type of relationship (linear), strength, direction, and context (based on a correct understanding of the correlation coefficient).

OR

only states that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats (based on a correct understanding of r^2) – BUT – does not state that the relationship is positive.

NOTE: Element three may be scored essentially or partially correct if the student uses a reasonable r (between 0 and 1) or R-Sq (adj) value.

Part (c) addresses the fourth element.

Element four is **essentially correct** if the student's solution

states that the existing line is not a good fit for the remaining seven points and correctly explains that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

Element four is **partially correct** if the student's solution

explains why the existing line is not a good fit for the remaining seven points but does not communicate that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

OR

removes fewer than the specified five points, but gives a correct interpretation of the effect on the correlation or slope of the least-squares regression line.

Copyright © 2002 by College Entrance Examination Board. All rights reserved.