Question 5

Intent of Question

The primary goals of this question were to assess students' ability to (1) determine the equation of the least squares regression line from a computer output; (2) use the slope of the least squares line to compare expected values of the response variable for different values of the explanatory variable; (3) recognize how to determine the proportion of variability in the response variable explained by the least squares line; (4) use computer output to determine whether the linear relationship between two quantitative variables is statistically significant.

Solution

Part (a):

The equation of the least squares regression line is

predicted electricity production = 0.137 + 0.240 × wind velocity.

Part (b):

The slope coefficient of 0.240 indicates that for each additional mph of wind speed, the expected electricity production increases by 0.240 amperes. Thus, the expected electricity production is $10 \times 0.240 = 2.40$ amperes higher on a day with 25 mph wind velocity as compared to a day with 15 mph wind velocity.

Part (c):

The proportion of variation in electricity production that is explained by the linear relationship with wind speed is R^2, which the regression output reports to be 0.873.

Part (d):

Yes, there is very strong statistical evidence that the population slope differs from zero, so electricity production is linearly related to wind speed. For testing the hypotheses $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$, where β represents the population slope, the output reveals that the test statistic is $t = 12.63$ and the *p*-value (to three decimal places) is 0.000. Because the *p*-value is so small (much less than both 0.05 and 0.01), the sample data provide very strong statistical evidence that electricity production is linearly related to wind speed.

Scoring

Parts (a), (b), (c) and (d) are scored as essentially correct (E), partially correct (P) or incorrect (I).

Question 5 (continued)

Part (a) is scored as follows:

Essentially correct (E) if the response gives the correct equation *AND* includes the following two components:

1. Provides correct variable names (with context).
2. Uses a modifier such as "expected" or "predicted" or "estimated" (or a "hat" symbol) with the response variable, electricity production.

Partially correct (P) if the response gives the correct equation *AND* includes exactly one of the two components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

Part (b) is scored as follows:

Essentially correct (E) if the response identifies and uses the correct slope value (0.240) *OR* the slope value identified in part (a) of the response

AND

the response includes the following three components:

1. Shows work (correct multiplication or correct substitution into an appropriate expression).
2. Arrives at an answer.
3. Provides correct measurement units (amperes).

Note: Calculating predicted values for both wind speeds and taking their difference is sufficient, as long as measurement units are provided.

Partially correct (P) if the response identifies and uses the correct slope value (0.240) or the slope value identified in part (a) of the response *AND* includes exactly two of the three components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

Part (c) is scored as follows:

Essentially correct (E) if response is 0.873.

Note: No work needs to be shown to earn an E, because the answer is read from the computer output.

Partially correct (P) if the response gives the value of adjusted R^2, rather than R^2, *OR* the response approximates (or rounds) the value of R^2.

Incorrect (I) if the response gives neither R^2 nor adjusted R^2, or if the response reports the square root of R^2.

Question 5 (continued)

Part (d) is scored as follows:

Essentially correct (E) if the response includes the following three components:

1. Gives the correct conclusion based on a test for the population slope.
2. Reports the correct p-value and/or t-statistic.
3. Provides linkage/justification between the p-value (or t-statistic) and the conclusion.

Partially correct (P) if the response provides exactly two of the three components listed above.

Note: If the wrong p-value is chosen, but the conclusion is consistent with that p-value and linkage or justification is provided, the response earns a P.

Incorrect (I) if the response fails to meet the criteria for E or P.

Each essentially correct (E) part counts as 1 point. Each partially correct (P) part counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, 2½ points), use a holistic approach to determine whether to score up or down, depending on the overall strength of the response and communication.

5A

5. Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237   R-Sq = 0.873   R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

$\widehat{\text{Electricity production}}$ (in amperes) = 0.137 + 0.240 wind velocity (mph)

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

$\widehat{\text{electricity produced}}$ = 0.137 + 0.240 (25 mph) = 6.137 amperes

$\widehat{\text{electricity production}}$ = 0.137 + 0.240 (15 mph) = 3.737 amperes

6.137 amperes - 3.737 amperes = 2.4 amperes

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

87.3% of the variation in electricity production is accounted by its linear relationship with wind velocity.

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

Since the p-value (0.000) < α (0.05) we reject the claim that there is no relationship between the electricity production and wind velocity. There is sufficient evidence to conclude that there is a relationship between electricity production and wind velocity

GO ON TO THE NEXT PAGE.

5B

5. Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237   R-Sq = 0.873   R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

$\hat{y} = 0.137 + 0.240(x)$

$\widehat{\text{electricity produced}} = 0.137 + 0.24(\text{wind velocity})$

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at ~~your answer~~.

$\hat{y} = 0.137 + 0.24(25) = 6.137$

$\hat{y} = 0.137 + 0.24(15) = 3.737$

6.137 − 3.737 = ~~2.4~~

2.4 more electricity

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity? R^2 – 87.3% of variation in electricity produce can be explained by wind velocity

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain. yes, because the p-value is low, therefore the null hypothesis should be rejected.

GO ON TO THE NEXT PAGE.

5C

5. Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

$\hat{y} = .019 + .240x$

$\hat{y}$ = predicted electricity production in amperes
x = wind velocity

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

$\hat{y} = .019 + .240(25) = 6.019$ amperes

$\hat{y} = .019 + .240(15) = 3.619$ amperes

$6.019 - 3.619 = 2.4$

The windmill would be expected to produce 2.4 amperes more when the wind is 25mph compared to it being 15mph.

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

According to the output, it is ≈0

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

$\alpha = .05 > p = 0$

We do have convincing evidence that electricity production by the windmill is related to velocity b/c $p < \alpha$.

GO ON TO THE NEXT PAGE.

Question 5

Overview

The primary goals of this question were to assess students' ability to (1) determine the equation of the least squares regression line from a computer output; (2) use the slope of the least squares line to compare expected values of the response variable for different values of the explanatory variable; (3) recognize how to determine the proportion of variability in the response variable explained by the least squares line; (4) use computer output to determine whether the linear relationship between two quantitative variables is statistically significant.

Sample: 5A
Score: 4

In part (a) the student gives the correct equation with correct variable names, in context, and uses the hat symbol as a modifier to indicate that the response variable is predicted electricity production. Although not required, the definitions include measurement units for each of the variables. The three required components are present, and part (a) was scored as essentially correct. In part (b) the student correctly calculates predicted values for both wind velocities (25 mph and 15 mph), takes their difference, arrives at the correct answer of 2.4, and reports the result with correct measurement units (amperes). The four required components are present, and part (b) was scored as essentially correct. In part (c) the student recognizes that the R-Sq value on the computer output is the proportion of variation in electricity production that is explained by its linear relationship with wind velocity and reports the correct value of 87.3 percent. Part (c) was scored as essentially correct. In part (d) the student correctly identifies the *p*-value (0.000), compares it to an α level of 0.05, and correctly rejects "the claim that there is no relationship between the electricity production and wind velocity" based on the comparison of $p < \alpha$. This is followed by a clearly communicated conclusion, in context. The three required components are included in this well-written response, and part (d) was scored as essentially correct. Because each of the four parts was scored as essentially correct, the response earned a score of 4.

Sample: 5B
Score: 3

In part (a) the student gives the correct equation, repeats the equation with the correct variable names substituted, and uses the hat symbol as a modifier on *y* and on electricity production. Part (a) includes the three required components and was scored as essentially correct. In part (b) the student correctly calculates predicted values for both wind speeds (25 mph and 15 mph), takes their difference, and arrives at the correct answer of 2.4. However, correct measurement units are not included. Given this omission, part (b) was scored as partially correct. In part (c) the student recognizes that the R-Sq value on the computer output is the proportion of variation in electricity production that is explained by its linear relationship with wind velocity and reports the correct value of 87.3 percent. Part (c) was scored as essentially correct. In part (d) the student reaches a correct conclusion ("yes") and justifies it by explaining that "the p-value is low." However the response does not identify which *p*-value is being referenced, so part (d) was scored as partially correct. Because two parts were scored as essentially correct and two parts were scored as partially correct, the response earned a score of 3.

Question 5 (continued)

Sample: 5C
Score: 2

In part (a) the student correctly identifies both the response variable (predicted electricity production) and the explanatory variable (wind velocity). The slope coefficient is correctly read from the output, but the standard error of the slope (0.019) is given as the y-intercept of the regression line. Because the equation is incorrect, part (a) was scored as incorrect. In part (b), using the incorrect equation from part (a), the student calculates predicted values for both wind speeds (25 mph and 15 mph) correctly, takes their difference, arrives at the correct answer (2.4) based on the equation given in part (a), and indicates the correct units (amperes). Because the answer is consistent with the equation given in part (a) and correct units are provided, part (b) was scored as essentially correct. In part (c) the student identifies the proportion of variation in electricity production that is explained by wind velocity to be approximately zero. This is an incorrect response, and part (c) was scored as incorrect. In part (d) the statement "$\alpha = .05 > p = 0$" identifies α, identifies the correct p-value, and compares the two. The student correctly concludes that the variables are related and justifies the conclusion because "$p < \alpha$." The three required components are included, and part (d) was scored as essentially correct. Because two parts were scored as essentially correct and two parts were scored as incorrect, the response earned a score of 2.