AP® STATISTICS
2006 SCORING GUIDELINES

Question 2

Intent of Question

The primary goal of this question is to assess a student's ability to identify the estimated regression line and to identify and interpret important statistics from regression output provided by statistical software in the context of a practical problem.

Solution

Part (a):

The regression line is $\hat{y} = -2.679 + 9.5x$, where $\hat{y}$ represents the estimated (or predicted) mean height of the soapsuds and x represents the amount of detergent added to the pan.

Part (b):

The value $s = 1.99821$ mm is the standard deviation of the residuals. This statistic measures a typical amount of variability in the vertical distances from the observed height of the soapsuds to the regression line.

OR

The value $s = 1.99821$ mm is a measure of variation in the height of soapsuds for a given amount of detergent.

Part (c):

The standard error of the estimated slope parameter is 0.7553 mm per gram. Thus, the standard deviation of the estimated slope for predicting the height of soapsuds by using an amount of detergent is estimated to be 0.7553 mm per gram. This value estimates the variability in the sampling distribution of the estimated slope (i.e., how much we would expect sample slopes to vary from experiment to experiment).

Scoring

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

Part (a) is essentially correct (E) if the least squares regression line is correctly identified and the variables are correctly defined.

Part (a) is partially correct (P) if:

the least squares regression line is correctly identified and either of the two variables are not correctly defined;

OR

the least squares regression line is not presented using estimated or predicted notation, or $\hat{y}$, AND both variables are correctly defined;

OR

only one of the two values is correctly identified from the table and both variables are correctly defined.

Part (a) is incorrect (I) if the least squares regression line is incorrectly identified or not identified, and the variables are not correctly defined.

Question 2 (continued)

Notes:

- If y is identified as the height of the soapsuds and x is identified as the amount of detergent, then the student should get credit for defining the variables. However, y must be identified as an estimated height somewhere in the student response in order to get this part essentially correct.
- If $\hat{y}$ (or *estimated height*) is used to specify the regression line and y is identified as the height, the response should be scored as essentially correct. For example, a response of $\text{estimated height} = -2.679 + 9.5\text{amount}$, where y represents the height of suds an x represents the amount of detergent should be scored essentially correct.
- If the regression line is specified using y (or height) instead of $\hat{y}$ (or estimated height), the response is scored as partially correct. For example, a response of $y = -2.679 + 9.5x$ where y = height of suds and x = amount of detergent should be scored as partially correct.
- If the estimates of the intercept and slope are reversed ($\hat{y} = 9.5 - 2.679x$), then the response should be scored as incorrect.

Part (b) is essentially correct (E) if the standard deviation is correctly interpreted in the context of this study.

Part (b) is partially correct (P) if the standard deviation is correctly interpreted in general terms without context.

Part (b) is incorrect (I) if the response indicates that s is any other standard deviation (e.g., univariate distribution of y), or slope.

Notes:

- If s is interpreted as the estimated standard deviation of the differences between the observed values for the height of soapsuds and the values predicted from the regression line, the response should be scored essentially correct.
- If s is interpreted as a "typical" prediction error for estimating height from the amount of detergent, then the response should be scored essentially correct.

Part (c) is essentially correct (E) if the standard error is identified and interpreted correctly.

Part (c) is partially correct (P) if standard error is identified but interpretation is weak (e.g., the standard error is a standard deviation of the slope). The major idea of sampling variability is not included.

Part (c) is incorrect (I) if the standard error is not correctly identified, identified with no interpretation, or an incorrect interpretation is provided.

Question 2 (continued)

4 Complete Response

All three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct and either zero or one part partially correct

OR

No parts essentially correct and two parts partially correct

(a) Write the equation of the fitted regression line. Define any variables used in this equation.

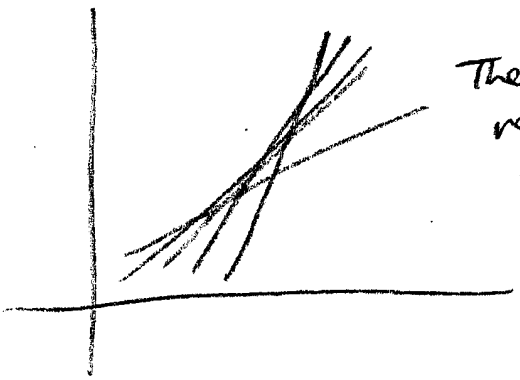
$\hat{y} = 9.5000x - 2.679$, where $\hat{y}$ represents the predicted height(mm) of soapsuds and x represents the amount of detergent (g).

(b) Note that $s = 1.99821$ in the computer output on the preceding page. Interpret this value in the context of this study.

This is the standard deviation of the distribution of heights for a fixed amount of detergent. In other words, if this experiment were repeated many times, s would be the standard deviation of the different heights (from each experiment) for a common amt. of detergent.

(c) Identify and interpret the standard error of the slope.

Standard error of the slope is the SE Coef of Amount, or 0.7553. This takes into account all possible slopes that can result from deviations in height for a given amount of detergent.



The slope may change for repeated experiments because the heights aren't always the same.

GO ON TO THE NEXT PAGE.

(a) Write the equation of the fitted regression line. Define any variables used in this equation.

$\hat{y} = -2.679 + 9.50x$

$\hat{y}$: the height (in mm) of the suds

x: the amount (in grams) of detergent added

(b) Note that $s = 1.99821$ in the computer output on the preceding page. Interpret this value in the context of this study.

This is the standard deviation of the residuals. It is a measures of the spread of the set of differences between the predicted and the observed values of y. In the context of this study, it is a measure of the spread of the set of differences (observed−predicted) between the predicted and observed values for the height of the suds.

(c) Identify and interpret the standard error of the slope.

The standard error of the slope (SE_b) is this problem is 0.7553. It is a measure, based on the standard deviation of the residuals and the spread of the x-values, of the variability of the coefficient of the slope of the least squares regression line. Such a low SE_b (0.7553), compared with a high b (9.500), signifies an extraordinarily high likelihood of an association between the amount of detergent and the height of the suds.

GO ON TO THE NEXT PAGE.

2C

(a) Write the equation of the fitted regression line. Define any variables used in this equation.

$\hat{y} = -2.679 + 9.50x$

y = height of resulting suds (millimeters)
x = amount of detergent (milliliters)

(b) Note that $s = 1.99821$ in the computer output on the preceding page. Interpret this value in the context of this study.

The s value is the standard deviation of the points on the least square regression line. In this problem you would expect 68% of the data to fall within ±1S of the LSRL, 95% of the data to fall within ±2S of the LSRL, and 99.7% of the data to fall within ±3S of the LSRL.

(c) Identify and interpret the standard error of the slope.

The standard error of the slope is .7593. This means that the standard deviation of the slope is this value.

GO ON TO THE NEXT PAGE.

AP® STATISTICS
2006 SCORING COMMENTARY

Question 2

Overview

The primary goal of this question was to assess a student's ability to identify the estimated regression line and to identify and interpret important statistics from regression output provided by statistical software in the context of a practical problem.

Sample: 2A
Score: 4

In part (a) the correct equation for the fitted regression line is given with the variables in the equation defined. In part (b) the student indicates knowledge of the s being the estimate of the standard deviation of the distribution of the heights for a fixed amount of detergent. The context is given by using the heights of the soapsuds and the amount of detergent. In part (c) the standard error of the slope is identified and knowledge of the concept of sampling variability is indicated. This essay earned a score of 4.

Sample: 2B
Score: 3

In part (a) the correct equation for the fitted regression line is given with the variables in the equation defined. In part (b) knowledge of the variation in the actual heights from the predicted heights is indicated and given in the context of the study. In part (c) the standard error of the slope is correctly identified as a measure of the variability of the slope. However, a sense of the recognition of sampling variability is not provided. This essay earned a score of 3.

Sample: 2C
Score: 2

In part (a) the correct equation for the fitted regression line is given with the variables in the equation defined. In part (b) an adequate interpretation of the standard deviation is given. The first sentence is somewhat awkward, but the next sentence indicates that there is some understanding of the variation about the fitted line. The response is not in context; hence it is only partially correct. In part (c) 0.7553 is identified as the standard deviation of the slope. This is considered a weak interpretation since there is no notion of sampling variability mentioned. This essay earned a score of 2.