STATISTICS
SECTION II
Part B
Question 6
Spend about 25 minutes on this part of the exam.
Percent of Section II grade—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A study was designed to explore subjects' ability to judge the distance between two objects placed in a dimly lit room. The researcher suspected that the subjects would generally overestimate the distance between the objects in the room and that this overestimation would increase the farther apart the objects were.

   The two objects were placed at random locations in the room before a subject estimated the distance (in feet) between those two objects. After each subject estimated the distance, the locations of the objects were rerandomized before the next subject viewed the room.

   After data were collected for 40 subjects, two linear models were fit in an attempt to describe the relationship between the subjects' perceived distances (y) and the actual distance, in feet, between the two objects.

   Model 1: $\hat{y} = 0.238 + 1.080 \times$ (actual distance)

   The standard errors of the estimated coefficients for Model 1 are 0.260 and 0.118, respectively.

   Model 2: $\hat{y} = 1.102 \times$ (actual distance)

   The standard error of the estimated coefficient for Model 2 is 0.393.

   (a) Provide an interpretation in context for the estimated slope in Model 1.

   (b) Explain why the researcher might prefer Model 2 to Model 1 in this context.

   (c) Using Model 2, test the researcher's hypothesis that in dim light participants overestimate the distance, with the overestimate increasing as the actual distance increases. (Assume appropriate conditions for inference are met.)

   The researchers also wanted to explore whether the performance on this task differed between subjects who wear contact lenses and subjects who do not wear contact lenses. A new variable was created to indicate whether or not a subject wears contact lenses. The data for this variable were coded numerically (1 = contact wearer, 0 = noncontact wearer), and this new variable, named "contact," was included in the following model.

   Model 3: $\hat{y} = 1.05 \times$ (actual distance) $+ 0.12 \times$ (contact) $\times$ (actual distance)

   The standard errors of the estimated coefficients for Model 3 are 0.357 and 0.032, respectively.

© 2007 The College Board. All rights reserved.
Visit apcentral.collegeboard.com (for AP professionals) and www.collegeboard.com/apstudents (for students and parents).

AP® STATISTICS
2007 SCORING GUIDELINES

Question 6

Intent of Question

This question was designed to evaluate a student's ability to make inferences for simple linear regression models. Interpreting model parameters and comparing and contrasting different models are important skills that are also being assessed. Finally, a multiple regression model with a special variable, an indicator variable, is introduced to investigate whether the relationship between the predictor and response variable differs for two different groups of people. Students are asked to sketch the estimated line for both groups and interpret the estimated parameters in the multiple regression model.

Solution

Part (a):

The value 1.080 estimates the *average* increase (in feet) in the perceived distance for each additional foot in actual distance between the two objects.

Part (b):

The model with zero intercept makes more intuitive sense in this particular situation. If the two objects are placed side by side (so the actual distance is zero), then we would expect the subjects to say that the distance between the objects is zero.

Part (c):

Let β denote the true slope between the perceived distances and the actual distances. The researcher's hypothesis is equivalent to $\beta > 1$. Thus, we want to conduct a hypothesis test for the slope parameter.

Step 1: States a correct pair of hypotheses:

$H_0 : \beta = 1$
$H_a : \beta > 1$

Step 2: Correct mechanics, including the value of the test statistic and *p*-value (or rejection region).

This is a *t*-test of a slope.

$$t = \frac{b - \beta}{s_b} = \frac{1.102 - 1}{0.393} = 0.260$$

$$\text{df} = 40 - 1 = 39$$

p-value = $P(t > .260) = 0.398$

© 2007 The College Board. All rights reserved.
Visit apcentral.collegeboard.com (for AP professionals) and www.collegeboard.com/apstudents (for students and parents).

Question 6 (continued)

Step 3: States a correct conclusion in the context of the problem, using the result of the statistical test.

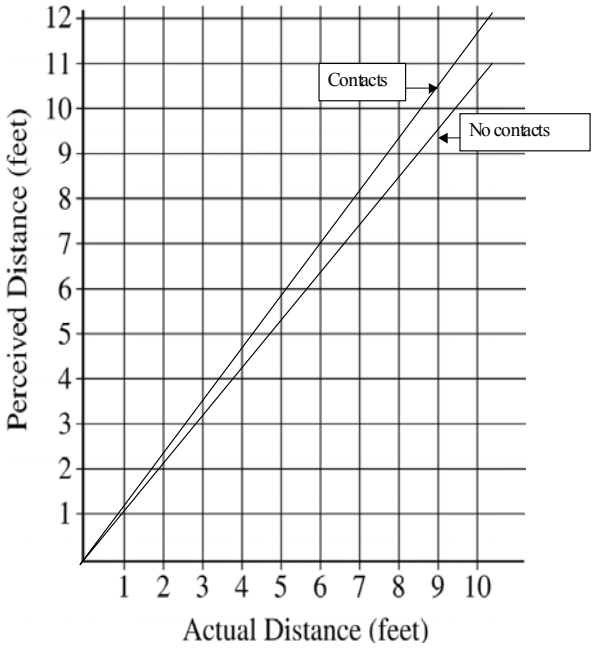
Since the *p*-value 0.398 is greater than 0.05, we cannot reject H_0. That is, we do not have statistically significant evidence to conclude that the subjects overestimate the distance with the magnitude of the overestimation increasing as the actual distance increases.

Part (d):

According to Model 3, the estimated models for the two groups are:

Contact wearers (*contact* = 1):
perceived distance = 1.05 (*actual distance*) + 0.12 (*actual distance*)
= 1.17 (*actual distance*)

Noncontact wearers (*contact* = 0):
perceived distance = 1.05 (*actual distance*)



Part (e):

Model 3 allows prediction of perceived distance separately for contact wearers and for noncontact wearers. The value of 1.05 estimates the average increase (in feet) in the perceived distance for each one-foot increase in actual distance for the population of noncontact wearers. The value of 0.12 estimates the *additional* increase (in feet) in the average perceived distance for each one-foot increase in actual distance for the contact wearers.

© 2007 The College Board. All rights reserved.
Visit apcentral.collegeboard.com (for AP professionals) and www.collegeboard.com/apstudents (for students and parents).

AP® STATISTICS
2007 SCORING GUIDELINES

Question 6 (continued)

Scoring

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

Parts (a) and (b) combined is scored as essentially correct (E) if both parts are correct.

Parts (a) and (b) combined is scored as partially correct (P) if:
one part is correct and the other part is incorrect;
OR
one part is correct and the other part is partially correct;
OR
both parts are partially correct.

Part (a) and (b) combined is scored as incorrect (I) if one part is partially correct.

Notes:

Part (a) is scored as partially correct if there is no word that makes it clear that 1.080 is not a deterministic increase.

Part (a) is scored as incorrect if the response:
- ignores the intercept and implies proportionality: for each foot of actual distance between the two objects, the subject perceives about 1.080 feet;
- consists of the equation rewritten in words.

Part (b)

Additional correct statement:
- The intercept is clearly not statistically significant, so the simpler model that includes only the slope is reasonable.

Partially correct statements:
- The SE for Model 2 is so large that Model 2 does not seem reasonable.
- The interpretation of the slope is straightforward if there is a 0 intercept: the percentage error is *slope* – 1 or 10.2 percent.
- The slope for Model 2 is farther above 1 than the slope for Model 1 and so more in line with the researcher's hypothesis.

Incorrect statements:
- Having one SE is better than having two.
- It is simpler/easier/shorter/more accurate to have just one coefficient.

© 2007 The College Board. All rights reserved.
Visit apcentral.collegeboard.com (for AP professionals) and www.collegeboard.com/apstudents (for students and parents).

AP® STATISTICS
2007 SCORING GUIDELINES

Question 6 (continued)

Part (c) is scored as:

Essentially correct (E) if three steps are correct.

Partially correct (P) if two steps are correct.

Incorrect (I) if one step is correct.

Notes:

- Hypotheses: the hypotheses step is incorrect if the alternative hypothesis is two-sided, or if the null hypothesis is $\beta = 0$. (It is not necessary to define β.)
- Computation: if the computation includes division by $\sqrt{40}$, the computation step is incorrect.
- Conclusion: a conclusion with no context is incorrect.

Part (d) is scored as essentially correct (E) if both estimated regression lines are graphed correctly and at least one is labeled.

Part (d) is scored as partially correct (P) if:

- the lines are graphed correctly but neither is labeled;
  OR
- the graphs consist of unconnected dots.

Part (d) is scored as incorrect (I) if:

- the two lines on the grid have the same slope;
  OR
- one line is plotted correctly and one line is not.

Part (e) is scored as essentially correct (E) if the response includes a correct interpretation of the estimated coefficients, 1.05 and 0.12. Unlike in part (a) there is no *y*-intercept, so this statement is correct: "For each foot of actual distance between the two objects, a noncontact wearer perceives about 1.05 feet, and a contact wearer will perceive about an additional 0.12 feet."

Part (e) is scored as partially correct (P) if:

- the response includes a correct interpretation of just one of the two coefficients;
  OR
- the response includes a correct interpretation of 1.05 and 1.05 + 0.12 = 1.17 but doesn't include a separate interpretation of 0.12;
  OR
- no numbers are mentioned, but it is made clear that both groups overestimate the distance *AND* that contact wearers overestimate more than do noncontact wearers.

Part (e) is scored as incorrect (I) if:

- the response says only that 1.05 and 0.12 are "slopes of regression lines";
  OR
- only the SEs of the coefficients, 0.357 and 0.032, are interpreted.

© 2007 The College Board. All rights reserved.
Visit apcentral.collegeboard.com (for AP professionals) and www.collegeboard.com/apstudents (for students and parents).