CollegeBoard
Advanced Placement Program

AP® Statistics
2007 Scoring Guidelines

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,000 schools, colleges, universities, and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

Intent of Question

The goals of this question are to assess a student's ability to: (1) explain how a commonly used statistic measures variability; (2) use a graphical display to address the research question of interest in a simple comparative experiment; and (3) use a confidence interval to make an appropriate inference.

Solution

Part (a):

Roughly speaking, the standard deviation $(s = 2.141)$ measures a "typical" or "average" distance between the individual discoloration ratings and the mean discoloration rating for the strawberries in the control group.

Part (b):

The preservative does appear to have been effective in lowering the amount of discoloration in strawberries. The discoloration ratings for the strawberries that received the preservative, shown in the top dotplot, are clearly centered at a value that is lower than the center of the discoloration rating distribution for the control group, shown in the bottom dotplot. In addition, the dotplots can be used to find all five statistics in the five-number summary (min, Q1, median, Q3, and max) for both groups. In fact, four of the five statistics (the maximum is the only exception) are lower for the strawberries that received the preservative.

Part (c):

Since zero is not contained in the 95 percent confidence interval for the difference $\mu_c - \mu_t$, we can conclude that there is a significant difference between the mean ratings for the two groups at the $\alpha = 0.05$ level. The population mean discoloration rating for untreated strawberries is estimated to be between 0.16 and 2.72 units higher than the population mean discoloration rating for treated strawberries. Thus, we think there would be a difference in the population mean discoloration ratings for treated and untreated strawberries.

Scoring

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

Part (a) is scored as essentially correct (E) if the standard deviation is interpreted correctly in the context of this experiment.

Part (a) is scored as partially correct (P) if:
  a standard textbook description of the standard deviation is provided without any reference to context, e.g., the standard deviation is described as the square root of an average squared deviation, or a "typical" or "average" deviation from the mean;
  OR
  the student provides evidence that the distribution of discoloration ratings in the control group is approximately normal, then correctly applies the 68-95-99.7 rule.

Part (a) is scored as incorrect (I) if:
>   the formula for the standard deviation is copied from the formula sheet and no further explanation is provided;
>   *OR*
>   the student uses the 68-95-99.7 rule without justifying that the distribution of discoloration ratings in the control group is approximately normal.

Part (b) is scored as essentially correct (E) if the student indicates that the preservative appears to be effective and explicitly links this decision to comparison of a characteristic of relative standing from the dotplots for the two groups.

Part (b) is scored as partially correct (P) if:
>   the student says that the preservative appears to be effective because the discoloration ratings appear to be lower for the treatment group, but the student does not explicitly link this decision to comparison of a characteristic of relative standing for the two groups;
>   *OR*
>   the student correctly compares one or more characteristics of relative standing for the two groups but never states that the preservative was effective in lowering discoloration.

Part (b) is scored as incorrect (I) if:
>   the student says that the preservative is not effective because the centers of the two distributions are roughly the same;
>   *OR*
>   the student says that the preservative is effective, with incorrect or no justification.

Part (c) is scored as essentially correct (E) if the student indicates that zero is not included in the confidence interval, so there is a difference (in population means), AND states the conclusion in the context of this experiment.

Part (c) is scored as partially correct (P) if:
>   the student indicates that zero is not included in the confidence interval, so there is a difference (in population means), but does not state the conclusion in the context of this experiment;
>   *OR*
>   the student correctly interprets the 95 percent confidence interval in context and indicates that there is a difference (in population means), without indicating that zero is not included in the confidence interval.

Part (c) is scored as incorrect (I) if the student concludes that the preservative is not effective *OR* says that no conclusion can be made based on the confidence interval, *OR* the student states a conclusion that refers to sample means instead of population means.

Notes:
*   The student is not required to specify the significance level in part (c), but if it is specified, it must be correct.

- An adjustment could be made to formally conduct a one-sided test, but in general, confidence intervals are used to conduct two-sided tests. The fact that the lower endpoint of the confidence interval is positive does provide evidence that the preservative is effective in lowering the amount of discoloration in strawberries. The correct formal statement is: The 97.5 percent lower confidence bound for the difference in the means is above zero (0.16), so at the 0.025 level we would conclude that the mean rating for the treated berries is significantly lower than the mean for the untreated strawberries.

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

AP® STATISTICS
2007 SCORING GUIDELINES

Question 2

Intent of Question

The three primary goals of this question are to assess a student's ability to: (1) clearly explain the importance of a control group in the context of an experiment; (2) describe the randomization process required for three groups; and (3) reduce variability by grouping experimental units as homogeneously as possible.

Solution

Part (a):

A control group gives the researchers a comparison group to be used to evaluate the effectiveness of the treatments. The control group allows the impact of the normal aging process on joint and hip health to be measured with appropriate response variables. The effects of glucosamine and chondroitin can be assessed by comparing the responses for these two treatment groups with those for the control group.

Part (b):

Each dog will be assigned a unique random number, 001–300, using a random number generator on a calculator, statistical software, or a random number table. The numbers will be sorted from smallest to largest. The dogs assigned the first 100 numbers in the ordered list will receive glucosamine. The dogs with the next 100 numbers in the ordered list will be assigned to the control group. Finally, the dogs with the numbers 201–300 will receive chondroitin.

Part (c):

The key question is which variable has the strongest association with joint and hip health. The goal of blocking is to create groups of homogeneous experimental units. It is reasonable to assume that most clinics will see all kinds and breeds of dogs so there is no reason to suspect that joint and hip health will be strongly associated with a clinic. On the other hand, different breeds of dogs tend to come in different sizes. The size of a dog is associated with joint and hip health, so it would be better to form homogeneous groups of dogs by blocking on breed.

Scoring

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

Part (a) is scored as essentially correct (E) if an advantage of using a comparison group is described in the context of this study.

Part (a) is scored as partially correct (P) if an advantage of using a control group is described but not in the context of this study.

Part (a) is scored as incorrect (I) if the student says that control groups should always be used but gives no further explanation *OR* an incorrect explanation.

Note: Since "treatment" and "control" are standard terms in design, a comparison of specific aspects of the study is needed to establish context.

Part (b) is scored as essentially correct (E) if randomization is used correctly, and the method of randomization can be implemented after reading the student response (so that two knowledgeable statistics users would use the same method to assign dogs to treatment groups).

Part (b) is scored as partially correct (P) if randomization or chance is used, but the method could not be implemented after reading the student response.

Part (b) is scored as incorrect (I) if randomization or chance is not used in a planned way *OR* the solution does not yield a completely randomized design.

Part (c) is scored as essentially correct (E) if:
> the student argues that the variable with the stronger relationship to joint and hip health should be used as the blocking variable;
>> *OR*
> the student states that the variable with the larger anticipated variability in the response measure should be used as the blocking variable so that units within blocks are as homogeneous as possible. A rationale is required, but a variable does not have to be selected.

Part (c) is scored as partially correct (P) if:
> the student indicates that the purpose of blocking is to create groups of homogeneous experimental units but makes an error in the application to this experiment;
>> *OR*
> the student does not acknowledge that there may be more variability associated in the response variable with one of the variables (breed or clinic) than the other;
>> *OR*
> the student does not recognize that both variables are associated with variation in the response variable.

Part (c) is scored as incorrect (I) if the student does not exhibit an understanding of the purpose of blocking.

> **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

Intent of Question

This question was developed to assess a student's understanding of the sampling distribution of the sample mean: in particular, a student's ability to: (1) compare probabilities concerning sample means from different sample sizes; (2) compute an appropriate probability; and (3) recognize that an application of the Central Limit Theorem is being evaluated.

Solution

Part (a):

The random sample of $n = 15$ fish is more likely to have a sample mean length greater than 10 inches. The sampling distribution of the sample mean $\bar{x}$ is normal with mean $\mu = 8$ and standard deviation $\sigma/\sqrt{n}$. Thus, both sampling distributions will be centered at 8 inches, but the sampling distribution of the sample mean when $n = 15$ will have more variability than the sampling distribution of the sample mean when $n = 50$. The tail area ($\bar{x} > 10$) will be larger for the distribution that is less concentrated about the mean of 8 inches when the sample size is $n = 15$, as shown in the following graph.
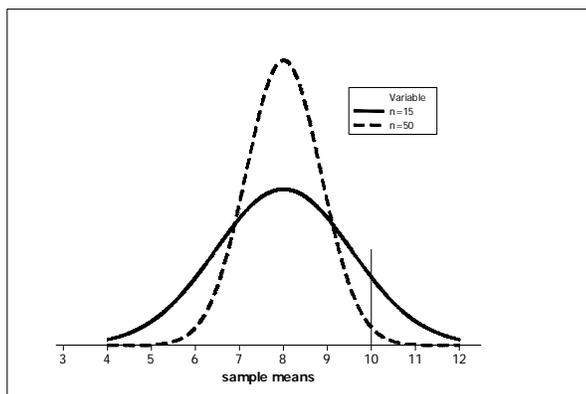
Part (b):

$$P(\bar{x} < 7.5) = P\left(z < \frac{7.5 - 8}{0.3}\right) = P\left(z < -\frac{5}{3}\right) = P(z < -1.67) = 0.0475$$
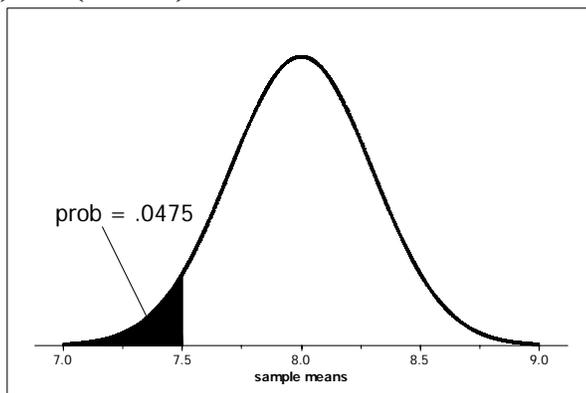
Part (c):

Yes. The Central Limit Theorem says that the sampling distribution of the sample mean will become approximately normal as the sample size *n* increases. Since the sample size is reasonably large (*n* = 50), the calculation in part (b) will provide a good approximation to the probability of interest even though the population is nonnormal.

Scoring

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

Part (a) is scored as essentially correct (E) if the student says that the sample of 15 fish is more likely to have a mean length that is greater than 10, AND the justification is based on *variability* in the *sampling* distributions.

Part (a) is scored as partially correct (P) if:
    the student makes correct statements about the sampling distribution of the sample mean or the probabilities but does not specifically refer to the variability in these two sampling distributions;
        OR
    the student remarks that the sample mean approaches the population mean as the sample size increases (an argument based on the Law of Large Numbers).

    Some examples of partially correct statements are:
    - With the smaller sample size we will be more likely to get an extreme value for the sample mean.
    - Variability in the smaller sample is larger.
    - Variability in the larger sample is smaller.
    - The sample mean approaches the population mean as the sample size increases.

Part (a) is scored as incorrect (I) if an answer is provided with no justification or incorrect justification.

Note: If a student chooses a particular value for a standard deviation and goes through the correct calculations, or shows the result algebraically based on a generic standard deviation, then the response should be scored essentially correct.

Part (b) is scored essentially correct (E) if the probability is calculated correctly and a reasonable sketch or evidence of calculation is shown.

Part (b) is scored partially correct (P) if:
    an incorrect but plausible calculation is shown. Examples include using an incorrect standard deviation (such as $0.3/\sqrt{50}$) to obtain the probability;
        OR
    the student switches the sample mean and the population mean to obtain a standardized *z* value of 1.67.

Part (b) is scored incorrect (I) if an answer is provided with no justification or incorrect justification.

Note: Normalcdf (…) with no additional work is at best partially correct. If an appropriate sketch accompanies the calculator command, *OR* if the components of the calculator command are clearly identified/labeled, then the solution should be scored essentially correct.

Part (c) is scored as essentially correct (E) if the student says that the probability is a reasonable approximation because of the Central Limit Theorem and also refers to the large sample size in this case.

Part (c) is scored partially correct (P) if the student indicates that the response in part (b) would not change but provides a weak justification. Examples of a weak justification include mentioning CLT without reference to sample size, and mentioning sample size without reference to CLT.

Part (c) is scored incorrect (I) if an answer is provided with no justification or incorrect justification.

Note: An E counts for 2 points in part (a), and an E counts for 1 point in each of parts (b) and (c). Similarly, a P counts for 1 point in part (a), and a P counts for ½ point in parts (b) and (c). When the total number of points earned is not an integer, the final score earned will be rounded down to the integer value.

> **4**   **Complete Response**
>
>   4 points earned
>
> **3**   **Substantial Response**
>
>   3 or 3½ points earned
>
> **2**   **Developing Response**
>
>   2 or 2½ points earned
>
> **1**   **Minimal Response**
>
>   1 or 1½ points earned

Question 4

Intent of Question

This statistical inference question was developed to assess a student's ability to distinguish paired-data procedures from two-sample procedures and to execute the selected procedure. The ability to provide a complete statistical justification is an important skill that can be evaluated with this standard inference problem.

Solution

A hypothesis test for the mean difference in the level of *E. coli* bacteria contamination in beef detected by the two methods will be conducted.

Part 1: States a correct pair of hypotheses:

$$H_0 : \mu_d = 0$$
$$H_a : \mu_d \neq 0$$

where μ_d is the mean difference (method A − method B) in the level of *E. coli* bacteria contamination in beef detected by the two methods

Part 2: Identifies a correct test (by name or by formula) and checks appropriate conditions:

Paired *t*-test $t = \dfrac{\overline{x}_d - 0}{s_d / \sqrt{n_d}}$

Conditions:
1. Since the observations are obtained on 10 randomly selected specimens, it is reasonable to assume that the 10 data pairs are independent of one another.
2. The population distribution of differences is normal.

The computed differences are:
      -0.3   0.5   0.3   0.6   0.8   0.7   1.2   0.2 -0.1 -1.0

Histogram of the differences (A-B):
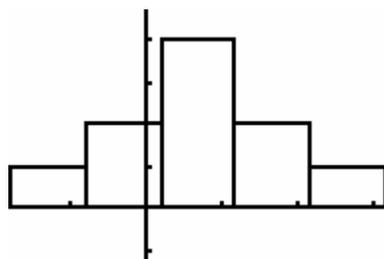
This histogram of differences is symmetric with no apparent outliers. Even though it is hard to judge the overall shape of a distribution with only 10 observations, it appears that the normal distribution is a reasonable option in this case.

Part 3: Correct mechanics, including the value of the test statistic, d.f., and *P*-value (or rejection region):

$$\overline{x}_d = 0.29 \qquad\qquad s_d = 0.629727$$

$$t = \frac{0.29 - 0}{0.629727/\sqrt{10}} = \frac{0.29}{0.199137} = 1.46 \qquad\qquad \text{d.f.} = 9 \qquad\qquad P\text{-value} = 0.179$$

OR

Calculator: t = 1.4563, *P*-value = 0.1793, d.f. = 9.

Part 4: States a correct conclusion in the context of the problem, using the result of the statistical test.

Since the *P*-value is greater than 0.05, we cannot reject H_0. We do NOT have statistically significant evidence to conclude that there is a difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef. In other words, there does not appear to be a significant difference in these two methods for measuring the level of *E. coli* contamination in beef.

Scoring

Parts 1, 2, 3, and 4 are scored as essentially correct (E) or incorrect (I).

Part 1 is scored as essentially correct (E) if the student states a correct pair of hypotheses. The hypotheses may be stated in terms of μ_A and μ_B. With any nonstandard notation used, the parameters must be identified in context clearly indicating the *population* mean.

Part 2 is scored as essentially correct (E) if the student identifies a correct test (by name or formula) and checks appropriate conditions. The conditions for the paired *t*-test are about the differences. If the student says that the 10 differences can be viewed as a SRS of all differences, the answer is acceptable. However, the student does not need to repeat the fact that these specimens can be viewed as a random sample.

It is not acceptable to view all 20 observations as a random sample or two independent samples. If conditions are stated and checked using the two separate samples, part 2 is scored as incorrect (I) for a paired *t*-test.

For part 2, a graphical check of normality is required. Graph(s) should be consistent with the data, *AND* students must comment linking the graph to the condition.

Histogram of the differences (B-A):
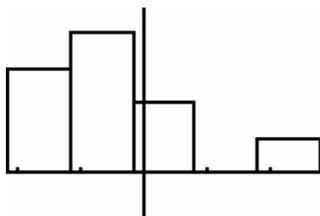
This histogram of differences is roughly symmetric with no apparent outliers. Even though it is hard to judge the overall shape of a distribution with only 10 observations, it appears that the normal distribution is a reasonable option in this case.

Boxplot of differences (A-B):

Boxplot of differences (B-A):

The boxplot of the differences shows that the distribution is approximately symmetric with no outliers, so it is reasonable to proceed with the paired *t*-test.
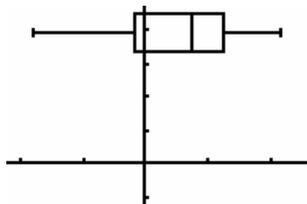
Normal Probability Plot of differences (A-B) (data on *x*-axis):
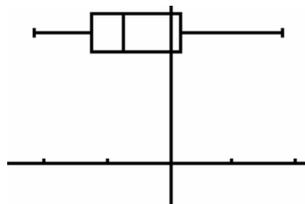
Normal Probability Plot of differences (B-A) (data on *x*-axis):

The normal probability plot shows linear trend with no obvious departures from linear trend, so the normal model is reasonable.

Part 3 is scored essentially correct (E) if the student performs correct mechanics when calculating the value of the test statistic and correctly calculates the *p*-value for the rejection region.

Part 4 is scored essentially correct (E) if the student states a correct conclusion in the context of the problem, using the result of the statistical test.

If the *p*-value in part 3 is incorrect but the conclusion is consistent with the computed *p*-value, part 4 can be considered correct.

In part 4, if both an α and a *p*-value are given together, the linkage between the *p*-value and the conclusion is implied.

If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the *p*-value or explaining how the conclusion follows from the *p*-value.

Scoring Confidence Interval approach:
    A confidence interval may be used to make the inference but must include all four parts to get full credit.

    The confidence level must be stated to get credit for part 3.

    A 95 percent confidence interval for μ_d is (-0.16, 0.74).

Question 4 (continued)

Since zero is included in the 95 percent confidence interval, we cannot reject the null hypothesis at the 0.05 level. We do NOT have statistically significant evidence to conclude that there is a difference in the mean amount of *E.coli* bacteria detected by the two methods for this type of beef. In other words, there does not appear to be a significant difference in these two methods for measuring the level of *E.coli* contamination in beef.

Scoring independent samples *t*-test or confidence interval approach:
 If an independent samples *t*-test or confidence interval is done, the maximum score is 3, provided all four parts for independent samples *t*-test are done correctly.

 For the independent samples *t*-test or confidence interval, the condition of normality must be checked using two samples separately.

 $t = 0.079$ $p = 0.9377$ df = 18 (pooled) or 17.97 (unpooled)

 A 95 percent independent-samples (two-sample) confidence interval for $\mu_A - \mu_B$ is (-7.40, 7.98).

Each part is scored as correct or incorrect.

 4 **Complete Response**

   Four parts correct

 3 **Substantial Response**

   Three parts correct

 2 **Developing Response**

   Two parts correct

 1 **Minimal Response**

   One part correct

Intent of Question

The primary goals of this statistical inference question are to assess a student's ability to: (1) distinguish an observational study from an experiment; (2) state the appropriate hypotheses for a research problem; (3) check the appropriate conditions for an inference procedure; and (4) interpret standard results for an inference procedure that is unfamiliar to students.

Solution

Part (a):

This is an experiment because the researchers imposed treatments by randomly assigning drivers to the two different conditions (simulated driving while talking on a cell phone versus simulated driving while talking to a passenger).

Part (b):

Let p_{cell} denote the proportion of drivers who miss an exit while using a cell phone and p_{pass} denote the proportion of drivers who miss an exit while talking to a passenger.

H_0: $p_{cell} = p_{pass}$
H_A: $p_{cell} > p_{pass}$

Part (c):

The conditions required for a two-sample z-test of equal proportions are:
(1) independent random samples or random assignment, and
(2) large sample sizes $\left[n_1 \hat{p}_1 \geq 10, n_1(1-\hat{p}_1) \geq 10, n_2 \hat{p}_2 \geq 10, n_2(1-\hat{p}_2) \geq 10 \right]$.

Random assignment is stated in the stem so the first condition is met. However, the numbers of successes ($n_{cell} \hat{p}_{cell} = 7$ and $n_{pass} \hat{p}_{pass} = 2$) are both smaller than 10, so the large sample condition is not met in this situation. Note: If the student uses the rule of thumb with 10 replaced by 5, then the number of successes for the second sample is still too small.

Part (d):

Interpretation: Assuming that talking on a cell phone and talking to a passenger are equally distracting (there is no difference in the two population proportions of drivers who will miss the exit), the p-value measures the chance of observing a difference in the two sample proportions as large as or larger than the one observed.

Conclusion: Since the p-value 0.0683 is larger than 0.05, we cannot reject the null hypothesis. That is, we do not have statistically significant evidence to conclude that using a cell phone is more distracting to drivers than talking to another passenger in the car.

Notice that if we increase the significance level to 0.1, then we could reject the null hypothesis and conclude that drivers are significantly more distracted when using a cell phone.

Scoring

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

Part (a) is scored as essentially correct (E) if the student indicates that this is an experiment because treatments were imposed.

Part (a) is scored as incorrect (I) if no explanation is provided, or the student says that this is an observational study.

Part (b) is scored as essentially correct (E) if the student correctly identifies the two population proportions with the correct hypotheses. Nonstandard notation must indicate reference to population proportions.

Part (b) is scored as incorrect (I) if the student is *clearly* referring to the sample proportions.

Part (c) is scored as essentially correct (E) if the student provides both conditions and correctly comments on both.

Part (c) is scored as partially correct (P) if the student provides and correctly comments on only one of the conditions.

Part (c) is scored as incorrect (I) if conditions are provided but no correct comments are given.

Part (d) is scored as essentially correct (E) if the *p*-value is correctly interpreted *AND* the correct conclusion is provided *AND* context is given.

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

   i)   either the *p*-value is correctly interpreted OR the correct conclusion is provided
     AND
   ii)   context is given.

Part (d) is scored as incorrect (I) if *neither* a correct interpretation of the *p*-value in context *NOR* a correct conclusion in context is provided.

In part (d) if both an α and a *p*-value are given together, the linkage between the *p*-value and the conclusion is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the *p*-value or explaining how the conclusion follows from the *p*-value.

Note: Any choice of an α could have been made as long as the appropriate interpretation is made relative to that choice of α.

Question 5 (continued)

Each essentially correct (E) response counts as 1 point; each partially correct (P) response 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 strength of the response and communication.

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

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

$$p\text{-}value = P(t > .260) = 0.398$$

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*)
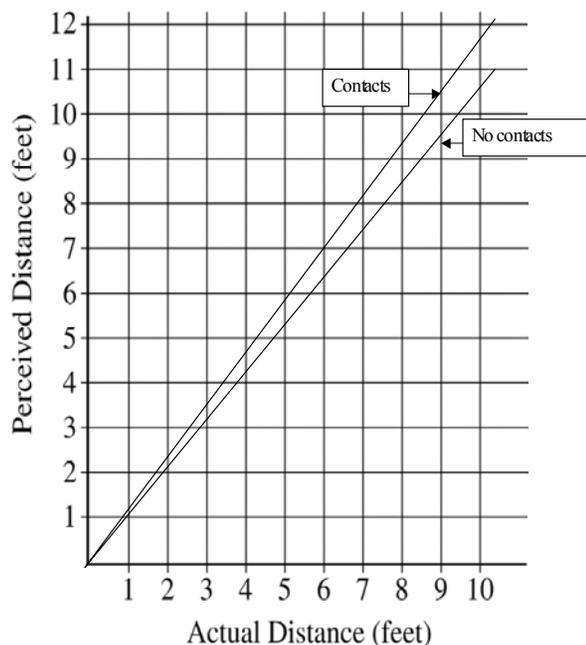
Part (e):

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.

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.

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.

Each essentially correct (E) response counts as 1 point; each partially correct (P) response 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 strength of the response and communication.