(b) Write a few sentences summarizing what the display in part (a) reveals about the association between gender and job experience for the students in the sample.

(c) Which test of significance should be used to test if there is an association between gender and job experience for the population of high school seniors in the district?

State the null and alternative hypotheses for the test, but do not perform the test.

---

2. A tire manufacturer designed a new tread pattern for its all-weather tires. Repeated tests were conducted on cars of approximately the same weight traveling at 60 miles per hour. The tests showed that the new tread pattern enables the cars to stop completely in an average distance of 125 feet with a standard deviation of 6.5 feet and that the stopping distances are approximately normally distributed.

(a) What is the 70th percentile of the distribution of stopping distances?

(b) What is the probability that at least 2 cars out of 5 randomly selected cars in the study will stop in a distance that is greater than the distance calculated in part (a) ?

(c) What is the probability that a randomly selected sample of 5 cars in the study will have a mean stopping distance of at least 130 feet?

---

3. Before beginning a unit on frog anatomy, a seventh-grade biology teacher gives each of the 24 students in the class a pretest to assess their knowledge of frog anatomy. The teacher wants to compare the effectiveness of an instructional program in which students physically dissect frogs with the effectiveness of a different program in which students use computer software that only simulates the dissection of a frog. After completing one of the two programs, students will be given a posttest to assess their knowledge of frog anatomy. The teacher will then analyze the changes in the test scores (score on posttest minus score on pretest).

(a) Describe a method for assigning the 24 students to two groups of equal size that allows for a statistically valid comparison of the two instructional programs.

(b) Suppose the teacher decided to allow the students in the class to select which instructional program on frog anatomy (physical dissection or computer simulation) they prefer to take, and 11 students choose actual dissection and 13 students choose computer simulation. How might that self-selection process jeopardize a statistically valid comparison of the changes in the test scores (score on posttest minus score on pretest) for the two instructional programs? Provide a specific example to support your answer.

GO ON TO THE NEXT PAGE.

4. One of the two fire stations in a certain town responds to calls in the northern half of the town, and the other fire station responds to calls in the southern half of the town. One of the town council members believes that the two fire stations have different mean response times. Response time is measured by the difference between the time an emergency call comes into the fire station and the time the first fire truck arrives at the scene of the fire.

   Data were collected to investigate whether the council member's belief is correct. A random sample of 50 calls selected from the northern fire station had a mean response time of 4.3 minutes with a standard deviation of 3.7 minutes. A random sample of 50 calls selected from the southern fire station had a mean response time of 5.3 minutes with a standard deviation of 3.2 minutes.

   (a) Construct and interpret a 95 percent confidence interval for the difference in mean response times between the two fire stations.

   (b) Does the confidence interval in part (a) support the council member's belief that the two fire stations have different mean response times? Explain.

---

5. For many years, the medically accepted practice of giving aid to a person experiencing a heart attack was to have the person who placed the emergency call administer chest compression (CC) plus standard mouth-to-mouth resuscitation (MMR) to the heart attack patient until the emergency response team arrived. However, some researchers believed that CC alone would be a more effective approach.

   In the 1990s a study was conducted in Seattle in which 518 cases were randomly assigned to treatments: 278 to CC plus standard MMR and 240 to CC alone. A total of 64 patients survived the heart attack: 29 in the group receiving CC plus standard MMR, and 35 in the group receiving CC alone. A test of significance was conducted on the following hypotheses.

   H_0: The survival rates for the two treatments are equal.

   H_a: The treatment that uses CC alone produces a higher survival rate.

   This test resulted in a p-value of 0.0761.

   (a) Interpret what this p-value measures in the context of this study.

   (b) Based on this p-value and study design, what conclusion should be drawn in the context of this study? Use a significance level of $\alpha = 0.05$.

   (c) Based on your conclusion in part (b), which type of error, Type I or Type II, could have been made? What is one potential consequence of this error?

AP® STATISTICS
2009 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe a randomization process required for comparing two groups in a randomized experiment; and (2) describe a potential consequence of using self-selection instead of randomization.

Solution

Part (a) (completely randomized design):

Each student will be assigned a unique random number using a random number generator on a calculator, statistical software, or a random number table. The assigned numbers will be listed in ascending order. The students with the lowest 12 numbers in the ordered list will receive the instructional program that requires physically dissecting frogs. The students with the highest 12 numbers will receive the instructional program that uses computer software to simulate the dissection of a frog.

Part (a) *alternative* (randomized block design):

Students will be paired or placed into blocks of size two, based on having similar pretest scores. So, the first block will contain the two students with the two lowest pretest scores, the second block will contain the two students with the third- and fourth-lowest pretest scores, and so on, with the last block containing the two students with the two highest pretest scores. In each block, the students will be assigned a unique random number using a random number generator on a calculator, statistical software, or a random number table. The student in each block with the lower random number will receive the instructional program that requires physically dissecting frogs, and the student with the higher random number will receive the instructional program that uses computer software to simulate the dissection of a frog.

Part (b):

By not randomizing and allowing the students to self-select, there is a potential for changes to occur in the differences between pretest and posttest scores for a particular group because of the characteristics of students who choose a particular instructional method, not because of the instructional method itself. For example, suppose frog-loving students already know a lot about frog anatomy; one would therefore expect these students to be less likely to show a large change between the pretest and posttest scores. Suppose the frog-loving students tend to select the computer simulation method (perhaps because they do not like the notion of dissecting the frogs they love). The possible low change between pretest and posttest scores for the computer simulation group might then be attributed to the students' already knowing a lot about frog anatomy beforehand, not to the instructional method itself. The frog dissection group might see a larger change in scores because the students entering this group are those with the lower pretest scores (less prior knowledge) and who are thus more likely to show greater improvement between pretest and posttest scores.

Scoring

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

AP® STATISTICS
2009 SCORING GUIDELINES

Question 3 (continued)

Part (a) is scored as follows:

Essentially correct (E) if a proper method of randomization is described that (1) creates two groups of equal size; *AND* (2) assigns the named treatments to the groups in a manner that knowledgeable statistics users would employ to assign the students to the two instructional groups.

Partially correct (P) if only one of the two criteria above is met.

Incorrect (I) if neither criterion is met.

Notes:

- Coin tossing (or equivalent method) using a stopping rule to obtain equal sample sizes requires placing the students in the class in a random order. If this method does not include a random order, at best, part (a) is scored as partially correct.
- In using a random number table, if numbers are specified, the student must work with two-digit numbers. For example, if using the first 24 integers, the student must use 01–24, not 1–24. If the student uses numbers such as 1–24, a solution that would otherwise be essentially correct becomes partially correct, and a partially correct response becomes incorrect.

Part (a) *alternative* is scored as follows:

Essentially correct (E) if (1) blocks are formed based on students' having similar pretest scores; *AND* (2) the two students in each block are assigned to different treatments; *AND* (3) the method of randomization used to assign the students in each block to the treatments is correct and can be implemented after reading the student's response (in a manner that knowledgeable statistics users would employ to assign the students to the two instructional groups).

Partially correct (P) if two of the three components above are presented correctly.

Incorrect (I) if no more than one of the three components is presented correctly.

Part (b) is scored as follows:

Essentially correct (E) if (1) the example gives a reasonable characteristic of the self-selected students in the study; *AND* (2) explains how this characteristic could be associated with changes in the differences between the pretest and posttest scores.

Partially correct (P) if (1) the example gives a reasonable characteristic of the self-selected students in the study; *AND* (2) a weak explanation is provided of how this characteristic could be associated with changes in the differences between pretest and posttest scores.

Note: A *weak* explanation of how a characteristic could be associated with changes in the differences between pretest and posttest scores must at least mention test scores or state that one group will perform better than the other. (Simply mentioning a behavioral difference is not sufficient.)

Question 3 (continued)

Incorrect (I) if an incorrect or no explanation is provided of how a characteristic could be associated with changes in the differences between pretest and posttest scores

OR

the example does not give a reasonable characteristic of the self-selected students in the study

OR

a student says that there must be an equal number of students in the class assigned to each treatment.

4 Complete Response

Both parts essentially correct

3 Substantial Response

One part essentially correct and the other part partially correct

2 Developing Response

One part essentially correct and the other part incorrect

OR

Both parts partially correct

1 Minimal Response

No part essentially correct and only one part partially correct