CollegeBoard
Advanced Placement Program

# AP® Statistics
# 2008 Scoring Guidelines

## Question 1

### Intent of Question

The primary goals of this question were to assess a student's ability to (1) compare two distributions; (2) reevaluate shape, center, and spread for comparing the two distributions after one of the distributions is transformed by multiplying each of the data points by a constant; and (3) make a prediction about the means of the two distributions based on information derived about the behavior of the distributions from the boxplots.

### Solution

**Part (a):**

The cereals that list a serving size of one cup have a median sugar amount larger than the median for the cereals that list a serving size of three-quarters of a cup. There is more variability (larger range and larger IQR) for the one-cup cereals. The shapes of the two distributions differ. The distribution of sugar content for three-quarter-cup cereals is reasonably symmetric: notice that the median is in the middle of the box. The distribution of sugar content for one-cup cereals is clearly skewed to the left (skewed toward the lower values): notice that the median is pulled to the right side of the central box closer to the third quartile.

**Part (b):**

The distribution of sugar content in the cereals that list one-cup serving sizes remains the same as in part (a) because no transformations were applied to this distribution. There is a noticeable shift toward higher sugar content for the cereals that list three-quarter-cup servings after the transformation was applied to this distribution. The two types of cereals (one-cup and three-quarter-cup) now have similar medians, and the two distributions now show similar maximum values. In addition, the variability in the sugar content for cereals with a three-quarter-cup serving size increased by a factor of $\frac{4}{3}$ after the transformation was applied to the data in this distribution.

**Part (c):**

Judging from the boxplots in part (b), we would expect the mean amounts of sugar per serving to be different. By the symmetry of the boxplot for the three-quarter-cup cereals, we would expect the mean and median to be similar. Because the boxplot for the one-cup cereals is skewed to the left, we would expect the mean to be lower than the median. Thus, because both types of cereal have similar medians, we would expect the mean amount of sugar per cup for cereals with a one-cup serving size to be lower than the mean amount of sugar per cup for cereals with a three-quarter-cup serving size.

### Scoring

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

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

Essentially correct (E) if the student correctly <u>compares</u> center, shape, and spread of the two distributions. Specific numerical values are not required.

### Question 1 (continued)

Partially correct (P) if the student correctly compares any two of the three characteristics (center, shape, and spread) of the two distributions.

Incorrect (I) if the student compares no more than one of the three characteristics.

*Note:* If the student uses "normal," "mound-shaped," or "uniform" for "symmetric," then no credit is given for shape. If the student uses "mean" for "median," then no credit is given for center.

**Part (b)** is scored as follows:

Essentially correct (E) if the student correctly indicates that the three-quarter-cup cereals shift to higher values *AND* that the three-quarter-cup distribution becomes more variable.

Partially correct (P) if the student recognizes one of the two changes (shift to higher values or increase in variability) for the distribution of the three-quarter-cup cereals.

Incorrect (I) if the student identifies neither the shift to higher values nor the increased variability for the three-quarter-cup distribution.

**Part (c)** is scored as follows:

Essentially correct (E) if the student predicts that the mean for cereals with a one-cup serving size would be lower than the mean for cereals with a three-quarter-cup serving size *AND* provides a reasonable justification based on the left skewness of the distribution for the one-cup serving size.

Partially correct (P) if the student predicts that the mean will be lower for the one-cup serving size but provides a weak justification *OR* if the student correctly compares mean and median for each serving size with a justification based on the shapes of the distributions but fails to compare the means of the two serving sizes.

Incorrect (I) if the student predicts that the means will be different with no justification *OR* predicts that the means will be similar *OR* makes no prediction.

**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 one or two parts partially correct

*OR*

Three parts partially correct

## Question 1 (continued)

**1** **Minimal Response**

One part essentially correct and no parts partially correct

*OR*

No parts essentially correct and two parts partially correct

## Question 2

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) identify a potential source of nonresponse bias and recognize a possible consequence for interpreting the results of a survey; (2) recognize that increasing sample size does not remove bias; and (3) recommend an appropriate course of action to solve a practical problem with the use of a survey.

**Solution**

**Part (a):**

Responses were received from only 98 of the 500 (or 19.6 percent) of the randomly selected families. In other words, 80.4 percent of the randomly selected families did not respond to the survey. To obtain a nearly unbiased estimate of the proportion of families with at least one child in elementary school in this school district who support year-round school, we would need to assume that the families that did not respond would have a similar level of support for year-round school as those who did respond. This would not be the case, for example, if families who support year-round school were more likely to respond than families who do not support year-round school. In such a case, the estimate of the proportion of families who support year-round school calculated from the responses would tend to be higher than the population proportion of families who favor year-round school.

**Part (b):**

No, the nonresponse bias still exists. Combining the results from the original sample with a new random sample of 500 families will not solve the problem. Regardless of what happens in the second sample, the problem of nonresponse bias will still exist in the combined sample because there would be at least 402 nonresponses included from the original sample.

**Part (c):**

Contact the 402 families from whom responses were not received and ask their opinion on the proposal. This may require additional mailings or telephone calls, but it will provide better information about support for year-round school among all families in the school district with at least one child in elementary school.

*OR*

Take a new random sample or take a census and use an alternative strategy, such as telephone calls or in-person interviews, to help increase the response rate.

**Scoring**

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

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

Essentially correct (E) if the student clearly links the effect of nonresponse to biased estimation by explaining why population support for year-round school would be overestimated (or underestimated) from the sample results.

## Question 2 (continued)

Partially correct (P) if the student describes a reasonable consequence in the context of the study *OR* describes a difference related to the issue between families who are likely to respond and families who are not likely to respond.

Incorrect (I) otherwise.

**Part (b)** is scored as follows:

Essentially correct (E) if the student says no, taking another sample will not solve the problem with nonresponse bias *AND* explains that nonresponse bias in the original sample would result in nonresponse bias in the combined sample regardless of the results from the second sample.

Partially correct (P) if the student says no, taking another sample will not solve the problem with nonresponse bias, but provides a weak explanation *OR* says that the second sample will produce similar results to the first sample.

Incorrect (I) if the student says yes, combining results from the original sample and a new random sample will solve the problem *OR* says no but provides no explanation or an incorrect argument *OR* does not respond to the question.

**Part (c)** is scored as follows:

Essentially correct (E) if the student provides an explicit description of a reasonable strategy for reducing nonresponse in a new survey or census or by following up with families who did not respond to the original study.

Partially correct (P) if a student suggests that nonresponse should be reduced or response should be mandatory without providing an explicit description of a reasonable strategy.

Incorrect (I) if the student does not provide a strategy to increase response rates or suggests a strategy that would result in other biases.

**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 one or two parts partially correct

*OR*

Three parts partially correct

## Question 2 (continued)

1 **Minimal Response**

One part essentially correct and no parts partially correct

*OR*

No parts essentially correct and two parts partially correct

## Question 3

**Intent of Question**

The primary goals for this question were to assess a student's ability to (1) recognize and calculate the mean as the expected value of a probability distribution; (2) demonstrate how to use two distributions to form all possible ways a specific difference may occur; (3) calculate a probability for this specific difference occurring; and (4) calculate a probability from the probability distribution of all possible differences.

**Solution**

**Part (a):**

The expected scores are as follows:

Josephine
$\mu_J = 16(0.1) + 17(0.3) + 18(0.4) + 19(0.2) = 17.7$
Crystal
$\mu_C = 17(0.45) + 18(0.4) + 19(0.15) = 17.7$

**Part (b):**

| J | C | Probability |
|---|---|---|
| 16 | 17 | (0.1)(0.45) = 0.045 |
| 17 | 18 | (0.3)(0.40) = 0.12 |
| 18 | 19 | (0.4)(0.15) = 0.06 |

**Part (c):**

The probability is
$0.045 + 0.12 + 0.06 = 0.225$

**Part (d):**

P(difference = –1) = 0.225 (from part c)
P(difference = –2) = 1 – 0.015 – 0.225 – 0.325 – 0.260 – 0.90 = 0.085

**Distribution of Josephine – Crystal**

| **Differences** | –3 | –2 | –1 | 0 | 1 | 2 |
|---|---|---|---|---|---|---|
| **Probability** | 0.015 | **0.085** | **0.225** | 0.325 | 0.260 | 0.090 |

The probability that Crystal's score is higher than Josephine's score is
P(difference < 0) = 0.015 + 0.085 + 0.225 = 0.325

**Scoring**

This problem is scored in three sections. Section 1 consists of part (a). Section 2 consists of parts (b) and (c). Section 3 consists of part (d). Each section is scored as essentially correct (E), partially correct (P), or incorrect (I).

## Question 3 (continued)

**Section 1** [part (a)] is scored as follows:

Essentially correct (E) if correct expected scores (means) are calculated for both Josephine and Crystal with appropriate calculations or formulas shown for at least one of the players.

Partially correct (P) if the student makes one of the following errors:

- Rounds both expected values to integers (e.g., approximately 18 or 17–18)
- Calculates only one player's score correctly with appropriate calculations or formula
- Uses nonuniversal calculator syntax with linkage to the values in the table to describe how the correct expected values for both players are calculated
- Shows correct work for the expected values but gives answers of 17.5 and 18 (the unweighted averages)
- Gives correct expected values but does not show the multiplications or does not show the additions

Incorrect (I) if two or more of the errors above are made *OR* if no justification is given for correct answers *OR* if both expected scores are calculated using an incorrect method *OR* if the expected values are not calculated.

*Note*: If the student shows correct work but has at most one minor arithmetic error and/or copies at most one probability incorrectly from the table, the student should not be penalized for these types of errors.

**Section 2** [parts (b) and (c)] is scored as follows:

Essentially correct (E) if all five of the components below are correctly completed by the student:

- Lists all the score combinations that result in a difference of –1 in part (b)
- Calculates the probabilities correctly in part (b)
- Shows appropriate work or formula in part (b)
- Calculates the correct probability for the difference of –1 in part (c)
- Shows appropriate work or formula in part (c)

Partially correct (P) if three or four of the previous components are correct.

Incorrect (I) if at most two of the previous components are correct.

*Notes*:

- If a student gets incorrect answers for the three combinations that result in a difference of –1 but uses them correctly in part (c), the student can still get credit for the last two components if the resulting probability is between 0 and 1.
- If the student shows correct work but has at most one minor arithmetic error and/or copies at most one probability incorrectly from the table, the student should not be penalized for these types of errors.

**Section 3** [part (d)] is scored as follows:

Essentially correct (E) if both of the components below are successfully done by the student:

- Completes the table correctly
- Calculates the correct probability that Crystal's score is higher than Josephine's score *AND* shows appropriate work or formula

## Question 3 (continued)

Partially correct (P) if only one of the components is correct.

Incorrect (I) if both components are incorrect.

*Notes*:

- It is possible to calculate P(difference = –2) = 0.085 by listing the two combinations that result in a difference of –2.

| J | C | Probability |
|---|---|---|
| 16 | 18 | (0.1)(0.4) = 0.04 |
| 17 | 19 | (0.3)(0.15) = 0.045 |

- If a student has an incorrect answer in part (c) but uses it correctly in part (d), then the P(difference = –2) must be 0.085 *OR* the probabilities in the table must add up to 1 to get credit for the first component.
- If any of the values in the table are less than 0 or greater than 1, then no credit will be given for the first component.
- If the student shows correct work but has at most one minor arithmetic error and/or copies at most one probability incorrectly from the table, the student should not be penalized for these types of errors.

**4 Complete Response**

All three sections essentially correct

**3 Substantial Response**

Two sections essentially correct and one section partially correct

**2 Developing Response**

Two sections essentially correct and no sections partially correct

*OR*

One section essentially correct and one or two sections partially correct

*OR*

Three sections partially correct

**1 Minimal Response**

One section essentially correct and no parts partially correct

*OR*

No sections essentially correct and two sections partially correct

## Question 4

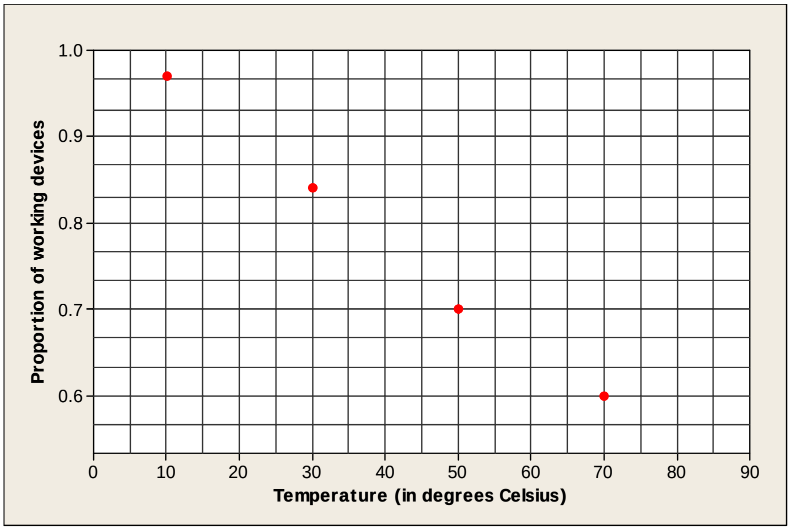
**Intent of Question**

The primary goals for this question were to assess a student's ability to (1) create and interpret a scatterplot and (2) estimate a proportion and the associated standard error.

**Solution**

**Part (a):**

Create a new variable *p* that indicates the proportion (number working / number tested) of working devices. A scatterplot of the proportion of working devices and temperature in degrees Celsius is shown below.




**Part (b):**

The scatterplot clearly shows a very strong, negative, linear association between the temperature and proportion of working devices at the end of the 5,000 hours. As the temperature increases, the proportion of working devices decreases.

**Part (c):**

The estimated proportion is the linear combination $\hat{p}_{40^\circ} = \frac{1}{2}\left(\frac{X_{30^\circ}}{n_{30^\circ}}\right) + \frac{1}{2}\left(\frac{X_{50^\circ}}{n_{50^\circ}}\right) = \frac{1}{2}\left(\frac{42}{50}\right) + \frac{1}{2}\left(\frac{21}{30}\right) = 0.77.$

Because the results for the two devices are independent, the variance of the estimated proportion is $\text{Var}(\hat{p}_{40^\circ}) = \left(\frac{1}{2}\right)^2 \left(\frac{\left(\frac{42}{50}\right)\left(\frac{8}{50}\right)}{50}\right) + \left(\frac{1}{2}\right)^2 \left(\frac{\left(\frac{21}{30}\right)\left(\frac{9}{30}\right)}{30}\right).$

Thus, the standard error is given by $\sqrt{\text{Var}(\hat{p}_{40^\circ})} = \left(\frac{1}{2}\right)\sqrt{0.0027 + 0.007} = 0.0492.$

**Question 4 (continued)**

**Scoring**

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

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

Essentially correct (E) if a scatterplot of the proportions is provided with appropriate scales and labels.

Partially correct (P) if a scatterplot of the proportions is provided with inadequate scales or inadequate labeling.

Incorrect (I) if an incorrect plot (e.g., a plot of the original counts or a nonscatterplot) is provided.

*Notes*:

- Students must give labels and at least two values on each scale to get full credit. One exception is that if the correct unit (°C, or similar) is given with values on the temperature axis, then this will suffice as a temperature label. (Note that the ° symbol alone does not suffice as the temperature label.)
- Labels that use "probability" instead of "proportion" or "percentage" are inadequate labels. Labels that use "estimated probability" or "sampled probability" are acceptable.

**Part (b)** is scored as follows:

Essentially correct (E) if the strength, direction, and type of relationship between these variables are correctly described in context.

Partially correct (P) if only two or three of the four (strength, direction, type, or context) are correctly described.

Incorrect (I) if none or one of the four (strength, direction, type, or context) is correctly described.

*Note*: Correlation is considered to be synonymous with linear association. Also, a student could give a correct interpretation of an incorrect plot (e.g., a plot of counts).

**Part (c)** is scored as follows:

Essentially correct (E) if an appropriate estimate and its standard error are calculated correctly with work shown.

Partially correct (P) if one of the two calculations is correct with work shown.

Incorrect (I) if correct answers are provided with no work shown *OR* correct answers are provided using inappropriate methods.

**Question 4 (continued)**

*Notes*:

- Students can give two possible approaches to "averaging the estimates at 30°C and 50°C." The first is the arithmetic mean $\hat{p}_{40^\circ} = \dfrac{\frac{42}{50} + \frac{21}{30}}{2} = 0.77$ with its standard error as given in the model solution. The second is the pooled estimate $\hat{p}_{40^\circ} = \dfrac{42 + 21}{50 + 30} = 0.7875.$ The standard error of this pooled estimate is calculated as follows:

$$\text{Var}\left(\frac{X_{30^\circ} + X_{50^\circ}}{n_{30^\circ} + n_{50^\circ}}\right) = \frac{1}{80^2}\left(n_{30^\circ} p_{30^\circ} q_{30^\circ} + n_{50^\circ} p_{50^\circ} q_{50^\circ}\right) = \frac{1}{80^2}\left(50\left(\frac{42}{50}\right)\left(\frac{8}{50}\right) + 30\left(\frac{21}{30}\right)\left(\frac{9}{30}\right)\right) = 0.00203.$$

$$\text{SE} = \sqrt{\text{Var}\left(\frac{X_{30^\circ} + X_{50^\circ}}{n_{30^\circ} + n_{50^\circ}}\right)} = \sqrt{0.00203} = 0.0451.$$

- Any standard error of a single proportion is not acceptable.

**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 one or two parts partially correct

*OR*

Three parts partially correct

**1 Minimal Response**

One part essentially correct and no parts partially correct

*OR*

No parts essentially correct and two parts partially correct

## Question 5

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) state the appropriate hypotheses; (2) identify and compute the appropriate test statistic; (3) make a conclusion in the context of the problem; and (4) compare two sets of proportions to identify the preferred habitat.

**Solution**

**Part (a):**

Step 1: States a correct pair of hypotheses.

$H_0$ : Moose have no preference for habitat type.
$H_a$ : Moose have a preference for habitat type.

*OR*

$H_0$ : The number of moose in each habitat type is proportional to the amount of acreage of that habitat type.
$H_a$ : The number of moose in at least one habitat type is not proportional to the amount of acreage of that habitat type.

*OR*

$H_0: p_1 = 0.340, p_2 = 0.101, p_3 = 0.104, p_4 = 0.455$, where $p_i$ = the proportion of moose in habitat type *i*.
$H_a$ : At least one of these proportions is incorrect.

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

- Chi-square goodness-of-fit test (or test for more than two proportions)

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

- The stem of the problem stated that conditions for inference are met.

Step 3: Correct mechanics, including the value of the test statistic, df, and *p*-value (or rejection region).

- The test statistic, with df = 4 − 1 = 3, is

$$\chi^2 = \frac{(25-39.780)^2}{39.780} + \frac{(22-11.817)^2}{11.817} + \frac{(30-12.168)^2}{12.168} + \frac{(40-53.235)^2}{53.235} = 43.6893.$$

- The *p*-value is $P(\chi_3^2 \geq 43.6893) < 0.0005$ (a calculator gives the *p*-value as $1.7569 \times 10^{-9}$ ).

**Question 5 (continued)**

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

The data are not consistent with the researchers' expectation. Because the *p*-value is less than $\alpha = 0.05$, we reject $H_0$. There is strong evidence that moose have a preference for habitat type.

*OR*

The data are not consistent with the researchers' expectation. If the null hypothesis is true and the number of moose in each of the habitat types is proportional to the acreage in that habitat type, then we would observe a test statistic of 43.69 or one more extreme less than 0.05 percent of the time. There is strong evidence that moose have a preference for habitat type.

**Part (b):**

The moose seem to prefer habitat types 2 and 3. Relative to the proportion of total acreage, a higher proportion of moose were observed in each of these habitat types than expected. In habitat types 1 and 4, the observed proportion of moose was less than the expected proportion of moose, indicating that these two habitat types are less desirable.

*OR*

Habitat type 3 seems to be the most preferred—it has a positive difference between the observed (30) and expected (12.168) counts of moose and the largest contribution to the chi-square statistic (26.1325). Alternatively, habitat type 3 has the largest positive difference between the observed proportion of moose (0.256) and the expected proportion of moose (0.104).

**Scoring**

This problem is scored in four sections. Section 1 consists of part (a), step 1. Section 2 consists of part (a), steps 2 and 3. Section 3 consists of part (a), step 4. Section 4 consists of part (b). Sections 1, 2, and 3 are scored as essentially correct (E) or incorrect (I), and section 4 is scored as essentially correct (E), partially correct (P), or incorrect (I).

If an inappropriate inference procedure is used in part (a), then all three sections must be scored as incorrect (I).

**Section 1** [part (a), step 1]: States a correct pair of hypotheses.

- Hypotheses must be given in context—which includes some reference to moose and the different habitat types—to earn an E. Hypotheses that clearly address sample data (like "observed number of moose") are incorrect.

## Question 5 (continued)

**Section 2** [part (a), steps 2 and 3]: Identifies a correct test and checks appropriate conditions. Mechanics are correct.

- A discussion of conditions for inference should generally be treated as extraneous. However, if the response includes inappropriate conditions—like normality or independent samples—the response cannot receive a score of 4.
- An inappropriate method of calculating df will result in these combined steps being scored incorrect.

**Section 3** [part (a), step 4]: States a correct conclusion in the context of the problem.

- If an incorrect *p*-value in steps 2 and 3 is obtained from a chi-square goodness-of-fit test, but the conclusion is consistent with this *p*-value, step 4 can be considered correct.
- If both an $\alpha$ and a *p*-value are given together, the linkage between the *p*-value and the conclusion is implied. If no $\alpha$ 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.

**Section 4** [part (b)] is scored as follows:

Essentially correct (E) if habitat types 2 and 3 are identified as the preferred habitat types with a justification that indicates there is a higher proportion (or a higher number) of moose than expected relative to the proportion of total acreage in those areas. One way to do this is to compare the observed density of moose across the four habitat types. Note that habitat types 2 and 3 also happen to make the largest contribution to the chi-square statistic.

*OR*

Habitat type 3 is identified as the most preferred because it has a higher proportion (or higher number) of moose than expected <u>and</u> the largest chi-square contribution *OR* the largest positive difference in observed and expected proportions *OR* the highest density of moose.

Partially correct (P) if habitat types 2 and 3 (or habitat type 3 alone) are identified with an incomplete justification. For example, a student might select habitat type 3 as most preferred based on the fact that it yields the largest contribution to the chi-square statistic but not indicate that there is a higher proportion (or higher number) of moose than expected in these areas.

Incorrect (I) if habitat types 2 and 3 (or just habitat type 3) are identified with no or incorrect justification *OR* habitat types 1 or 4 are identified.

Each essentially correct (E) response counts as 1 point, and a partially correct (P) response in part (b) counts as $\frac{1}{2}$ point.

**4** **Complete Response**

**3** **Substantial Response**

**2** **Developing Response**

**Question 5 (continued)**

**1 Minimal Response**

If a response is between two scores (for example, $2\frac{1}{2}$ points), use a holistic approach to determine whether to score up or down, depending on the strength of the response and communication.

## Question 6

**Intent of Question**

The primary goals of this investigative task were to assess a student's ability to (1) identify and conduct an appropriate inference based on the differences in the posttest and pretest scores; (2) identify and interpret appropriate information from statistical software; (3) make an inference based on separate regression analyses; and (4) recognize and explain the additional information provided from the different analyses.

**Solution**

**Part (a):**

Component 1: States a correct pair of hypotheses.

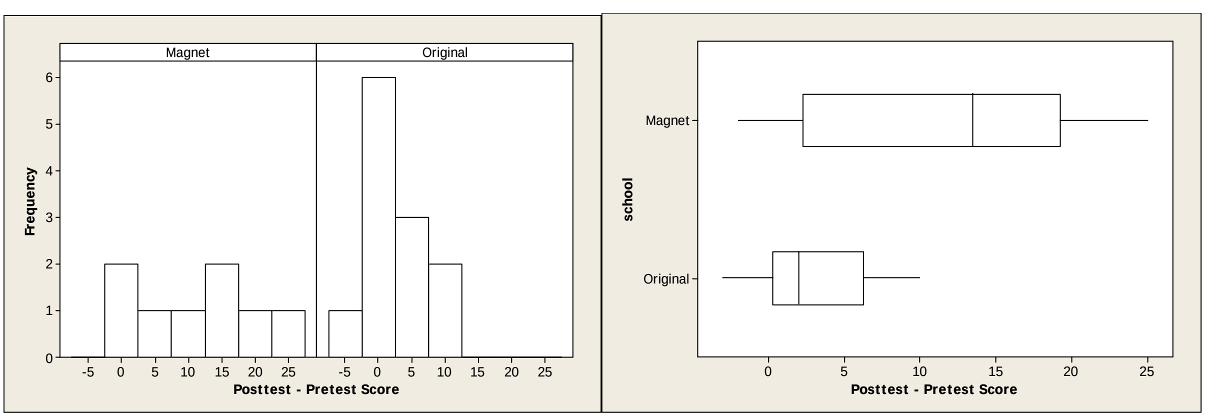
We want to test $H_0 : \mu_{DiffM} = \mu_{DiffO}$ versus $H_a : \mu_{DiffM} > \mu_{DiffO}$, where $\mu_{DiffM}$ is the mean difference (posttest − pretest) for all students at the magnet school and $\mu_{DiffO}$ is the mean difference (posttest − pretest) for all students who applied to attend the magnet school but were not selected and then attended the original school.

Component 2: Identifies a correct test (by name or formula) and checks the conditions.

A two-sample *t*-test for means, or $t = \dfrac{\overline{x}_M - \overline{x}_O}{\sqrt{\dfrac{s_M^2}{n_M} + \dfrac{s_O^2}{n_O}}}$

1. We need to assume randomness of the sampling used. It was stated in the stem that the students from the two different schools were randomly selected.

2. We need to check the assumption that the distributions of differences (posttest – pretest) for each of the two schools are normally distributed. Based on histograms and boxplots of these differences, there are no outliers or extreme skewness. Because these graphs reveal no obvious departures from normality, it appears reasonable to proceed with the *t*-test.



## Question 6 (continued)

Component 3: Performs correct mechanics, which include the value of the test statistic and *p*-value (or rejection region): $t = \dfrac{\overline{x}_M - \overline{x}_O}{\sqrt{\dfrac{s_M^2}{n_M} + \dfrac{s_O^2}{n_O}}} = \dfrac{11.75 - 3}{\sqrt{\dfrac{88.55}{8} + \dfrac{15.84}{12}}} \approx 2.487$

with a (one-sided) *p*-value ≈ 0.0177, df ≈ 8.69.

Component 4: Draws an appropriate conclusion in context and with linkage to the *p*-value (or rejection region):

Using $\alpha = 0.05$, we reject $H_0$ because 0.0177 < 0.05. We conclude that the sample data provide convincing evidence that students who attend the magnet school have a higher mean difference in test scores than students who attend the original school.

**Part (b):**

Let *y* = posttest score and *x* = pretest score.

(i). The predicted regression equation for the magnet school is $\hat{y} = 73.27 + 0.1811x$. For students at the magnet school, a 1-point increase in the pretest score is associated with a predicted increase of 0.1181 points on the posttest (i.e., the slope is positive but close to zero).

(ii). The predicted regression equation for the original school is $\hat{y} = 9.24 + 0.9204x$. For students at the original school, a 1-point increase in the pretest score is associated with a predicted increase of 0.9204 points on the posttest (i.e., the slope is positive and close to 1).

**Part (c):**

(i). The test statistic is $t = 0.40$ with a ***p*-value of 0.706**. Because the *p*-value is greater than any reasonable significance level, say 0.05, we fail to reject $H_0$. We conclude that there is insufficient evidence to state that pretest score is a significant predictor of posttest score at the magnet school. The data do not support a conclusion that a correlation exists between pretest and posttest scores at the magnet school.

(ii). The test statistic is $t = 6.09$ with a ***p*-value of 0.000**. Because the *p*-value is less than any reasonable significance level, say 0.05, we reject $H_0$ and conclude that there is sufficient evidence to state that pretest score is a significant predictor of posttest score at the original school. The data support a conclusion that a correlation exists between pretest and posttest scores at the original school.

**Part (d):**

Unlike the two-sample analysis of differences in part (a), the regression analyses allow us to explore the relationship between pretest and posttest scores at each school. From the regression output and graph, we see that students with low pretest scores benefit more from attending magnet schools, as compared with students with low pretest scores at the original school. Also at the magnet school, students with low pretest scores benefit more than students with high pretest scores. In other words, students at the magnet school all score high on the posttest, regardless of how they scored on the pretest. But at the original school, only students who scored high on the pretest scored high on the posttest.

## Question 6 (continued)

**Scoring**

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

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

Essentially correct (E) if all four components are correct.

Partially correct (P) if two or three components are correct.

Incorrect (I) if at most one component is correct.

**Part (b)** is scored as follows:

Essentially correct (E) if all four components—both equations and both interpretations in (i) and (ii)—are correct.

Partially correct (P) if two or three components are correct.

Incorrect (I) if at most one component is correct.

**Part (c)** is scored as follows:

Essentially correct (E) if all four components—both *p*-values and both conclusions in (i) and (ii)—are correct.

Partially correct (P) if two or three components are correct.

Incorrect (I) if at most one component is correct.

**Part (d)** is scored as follows:

Essentially correct (E) if the response clearly explains how the regression analyses provide additional information in this context **by addressing the impact of the magnet school on students with low pretest scores.**

Partially correct (P) if the response clearly describes how the regression analyses provide additional information in context but does not explain the impact of the magnet school on students with low pretest scores.

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

### Question 6 (continued)

Each essentially correct (E) response counts as 1 point. Each partially correct (P) response counts as $\frac{1}{2}$ point.

**4** **Complete Response**

**3** **Substantial Response**

**2** **Developing Response**

**1** **Minimal Response**

If a response is between two scores (for example, $2\frac{1}{2}$ points) use a holistic approach to determine whether to score up or down, depending on the overall strength of the response and communication.