CollegeBoard AP

AP® Statistics
2011 Scoring Guidelines

The College Board

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 College Board is composed of more than 5,700 schools, colleges, universities and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,800 colleges through major programs and services in college readiness, college admission, guidance, assessment, financial aid and enrollment. Among its widely recognized programs are the SAT®, the PSAT/NMSQT®, the Advanced Placement Program® (AP®), SpringBoard® and ACCUPLACER®. 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.

Visit the College Board on the Web: www.collegeboard.org.
AP Central is the official online home for the AP Program: apcentral.collegeboard.com

Question 1

Intent of Question

The primary goals of this question were to assess students' ability to (1) relate summary statistics to the shape of a distribution; (2) calculate and interpret a z-score; (3) make and justify a decision that involves comparing variables that are recorded on different scales.

Solution

Part (a):

No, it is not reasonable to believe that the distribution of 40-yard running times is approximately normal, because the minimum time is only 1.33 standard deviations below the mean $\left(z = \dfrac{4.4 - 4.6}{0.15} \approx -1.33 \right)$. In a normal distribution, approximately 9.2 percent of the z-scores are below -1.33. However, there are no running times less than 4.4 seconds, which indicates that there are no running times with a z-score less than -1.33. Therefore, the distribution of 40-yard running times is not approximately normal.

Part (b):

The z-score for a player who can lift a weight of 370 pounds is $z = \dfrac{370 - 310}{25} = 2.4$. The z-score indicates that the amount of weight the player can lift is 2.4 standard deviations above the mean for all previous players in this position.

Part (c):

Because the two variables — time to run 40 yards and amount of weight lifted — are recorded on different scales, it is important not only to compare the players' values but also to take into account the standard deviations of the distributions of the variables. One reasonable way to do this is with z-scores.

The z-scores for the 40-yard running times are as follows:

$$\text{Player A: } z = \frac{4.42 - 4.60}{0.15} = -1.2$$
$$\text{Player B: } z = \frac{4.57 - 4.60}{0.15} = -0.2$$

The z-scores for the amount of weight lifted are as follows:

$$\text{Player A: } z = \frac{370 - 310}{25} = 2.4$$
$$\text{Player B: } z = \frac{375 - 310}{25} = 2.6$$

AP® STATISTICS
2011 SCORING GUIDELINES

Question 1 (continued)

The z-scores indicate that both players are faster than average in the 40-yard running time and both are well above average in the amount of weight lifted. Player A is better in running time, and Player B is better in weight lifting. But the z-scores also indicate that the difference in their weight lifting (a difference of 0.2 standard deviation) is quite small compared with the difference in their running times (a difference of 1.0 standard deviation). Therefore, Player A is the better choice, because Player A is much faster than Player B and only slightly less strong.

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 the answer is "no" *AND* the response provides a reasonable explanation, based on the relationship between the mean, standard deviation, and minimum value of a data set whose distribution can be approximated by a normal distribution.

Partially correct (P) if the answer is "no" but the explanation is weak.

Incorrect (I) if the answer is "no" without an explanation or with an unreasonable explanation, *OR* if the response concludes that it is reasonable to believe that the distribution is approximately normal.

Notes
- A reasonable explanation should describe a characteristic of a normal distribution that is substantially contradictory to the information given for the running time data so that the running time distribution cannot be reasonably approximated by a normal distribution.
- Plausible comments about the distribution of running times are considered extraneous.
- Incorrect comments about the distribution of running times can lower the score one level (that is, from E to P or from P to I), depending on the severity of the comment.

Part (b) is scored as follows:

Essentially correct (E) if the response calculates the *z*-score correctly *AND* provides a correct interpretation that includes direction.

Partially correct (P) if the response has only one of the two components (calculation and interpretation) correct.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- Calculating a probability from a normal distribution for the weights is considered extraneous and is not a sufficient interpretation of a *z*-score.
- Percentiles are extraneous and cannot be used to indicate direction from the mean, because the distribution cannot be determined from the information provided.
- Context is provided in the stem of problem and is not required for the response to be considered correct.

- Either the formula with correct symbols or with correct numerical values is needed in addition to the value 2.4 in the calculation of the z-score.
- A diagram can show direction from the mean, if the quantities are appropriately labeled.

Part (c) is scored as follows:

Essentially correct (E) if the response addresses the following three components:
1. Correct selection of Player A.
2. Numerical adjustments of the scales so that the players' values can be compared for *BOTH* variables: time to run 40 yards and amount of weight lifted.
3. Justification of the selection in component 1 by using the players' values on both variables with respect to the adjusted scales.

Partially correct (P) if the response has exactly two of the three components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- It is not necessary to calculate z-scores. For example, the following response is scored as essentially correct (E): "Players A and B are close in weight lifting, because the difference of 5 pounds is much less than 1 standard deviation (25 pounds), but much less close in running time because the difference is 0.15 seconds, which is exactly one standard deviation. Therefore, player A should be selected since he is considerably faster and almost as strong as player B."
- Component 3 is not satisfied by the statement, "Player A should be selected since the weights lifted are close and running times are less close," because the adjusted scales are not mentioned. Such a statement could apply to the original data, where the values are on different scales.
- The justification in component 3 must reference the adjusted scale for at least one variable *AND* at least be implied for the other variable.
- Normal probability calculations can be used in establishing the numerical scale adjustments for component 2 and for justifying the selection of the players in component 3. However, this results in a lowering of scores (that is, from E to P or from P to I) unless the student has concluded in part (a) that it was reasonable to believe that the distribution of running times was approximately normal.
- Conceptual miscalculation of z-scores or probabilities (for example, using the wrong mean, reversing the order of subtraction, or multiplying probabilities) results in the loss of credit for component 2, whereas minor arithmetic mistakes are overlooked.

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 one part incorrect

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 two parts incorrect

OR

        Two parts partially correct and one part incorrect

Question 2

Intent of Question

The primary goals of this question were to assess students' ability to (1) determine a conditional probability from a table of data; (2) use a table of data to determine whether or not two events are independent; (3) demonstrate an understanding of the concept of independence by constructing a graph that displays independence between two variables.

Solution

Part (a):

Of the 200 male registered voters in Franklin Township, 48 are registered for Party Y. Therefore the conditional probability that a randomly selected voter is registered for Party Y, given that the voter is a male, is $\frac{48}{200} = 0.24$.

Part (b):

No, the events "is a male" and "is registered for Party Y" are not independent. One justification of this conclusion is to note that the conditional probability of the event "is registered for Party Y" given the event "is a male" — which was computed in part (a) — is not equal to the probability of the event "is registered for Party Y," as shown below.

$$P(\text{is registered for Party Y} \mid \text{is a male}) = 0.24$$
$$P(\text{is registered for Party Y}) = \frac{168}{500} = 0.336$$

Because $0.24 \neq 0.336$, the two events are not independent.

Part (c):

The marginal proportions of voters registered for each of the three political parties (without regard to gender) are given below.
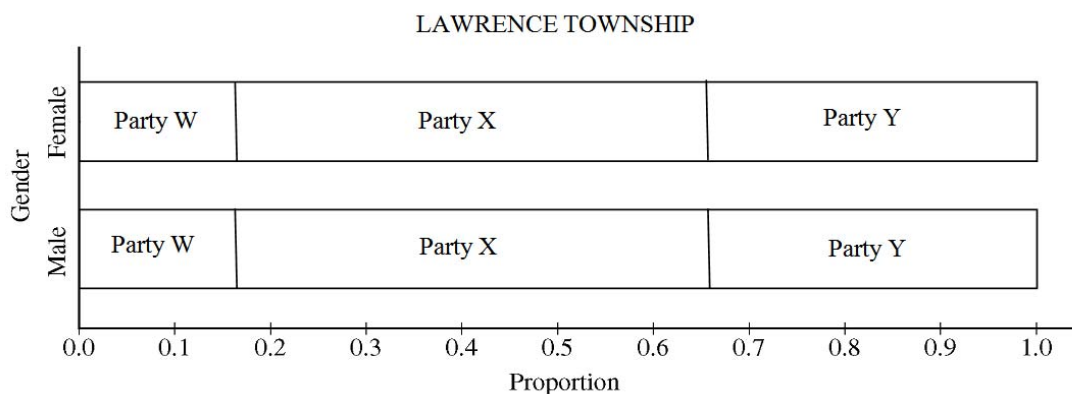
$$\text{Party W: } \frac{88}{500} = 0.176$$
$$\text{Party X: } \frac{244}{500} = 0.488$$
$$\text{Party Y: } \frac{168}{500} = 0.336$$

Because party registration is independent of gender in Lawrence Township, the proportions of males and females registered for each party must be identical to each other and also identical to the marginal proportion of voters registered for that party. Using the order Party W, Party X, and Party Y, the graph for Lawrence Township is displayed below.

AP® STATISTICS
2011 SCORING GUIDELINES

Question 2 (continued)

LAWRENCE TOWNSHIP

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 the response has the correct conditional probability *AND* shows the work.

Partially correct (P) if the response has the correct reverse conditional probability (of being a male given that he is registered for Party Y),
    OR
if the response has the correct conditional probability *BUT* does not show work.

Incorrect (I) if the response fails to meet the criteria for E or P.

Part (b) is scored as follows:

Essentially correct (E) if the response identifies two values whose inequality implies a lack of independence between the events *AND* includes the following three components:
1. Correct computations of the two values.
2. An explicit statement of whether the two values are equal or unequal.
3. An appropriate conclusion about the independence of the events.

Partially correct (P) if the response identifies two values whose inequality implies a lack of independence between the events but includes only two of the three components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Part (c) is scored as follows:

Essentially correct (E) if the response shows the same conditional distribution of party registration for both males and females *AND* includes the following two components:
1. Correct proportions for each party.
2. Correct labels (Party W, Party X, Party Y).

Partially correct (P) if the response shows the same conditional distribution of party registration for both males and females *AND* includes only one of the two components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Note: For all three parts, an incorrect statement that indicate a serious misunderstanding of statistical concepts, even if unrelated to the rest of the response, lowers the score one level (that is, from E to P, or from P to I). An example of this is a response that indicates confusion between independent events and disjoint events.

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 one part incorrect
>
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 two parts incorrect
>
OR
> Two parts partially correct and one part incorrect

Question 3

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe a process for implementing cluster sampling; (2) describe a statistical advantage of stratified sampling over cluster sampling in a particular situation.

Solution

Part (a):

The following two-step process can be used to select the eight apartments.

Step 1: Generate a random integer between 1 and 9, inclusive, using a calculator, a computer program, or a table of random digits. Select all four apartments on the floor corresponding to the selected integer.

Step 2: Generate another random integer between 1 and 9, inclusive. If the generated integer is the same as the integer generated in step 1, continue generating random integers between 1 and 9 until a different integer appears. Again select all four apartments on the floor corresponding to the second selected integer.

The cluster sample consists of the eight apartments on the two randomly selected floors.

Part (b):

Because the amount of wear on the carpets in apartments with children could be different from the wear on the carpets in apartments without children, it would be advantageous to have apartments with children represented in the sample. The cluster sampling procedure in part (a) could produce a sample with no children in the selected apartments; for example, a cluster sample of the apartments on the third and sixth floors would consist entirely of apartments with no children. Stratified random sampling, where the two strata are apartments with children and apartments without children, guarantees a sample that includes apartments with and without children, which, in turn, would yield sample data that are representative of both types of apartments.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response correctly addresses the following two components:
1. Indication that two floors are randomly selected, with all four apartments on each of the selected floors forming the sample (or that the entire floors should be carpeted).
2. Description of a valid random sampling procedure for selecting two floors that could be implemented after reading the response (so that two knowledgeable statistics users would use the same method to select the floors).

Partially correct (P) if the response includes exactly one of the two components listed above.

Incorrect (I) if the response includes neither of the two components listed above *OR* the response does not involve taking a random sample of two floors out of the nine.

Note: Some possible errors in component 2 include the following:
- Using 10 random digits rather than nine
- Failing to explicitly deal with the issue of potentially repeated random numbers

Part (b) is scored as follows:

Essentially correct (E) if the response indicates the following two components:

1. The amount of carpet wear could be different for apartments with and without children.
2. The stratified random sample ensures that some apartments with children will be selected.

Partially correct (P) if the response includes exactly one of the two components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- If the response in part (b) says that this stratified sampling method guarantees proportional representation of apartments with and without children, then the second component is satisfied.
- If the sampling procedure in part (a) divides the floors into two groups — those that have apartments with children and those that do not ("prestratification") — and then selects one floor from each group, score part (b) based on the degree to which a statistical advantage of the stratified sampling in part (b) is addressed.

4   Complete Response

> Both parts essentially correct

3   Substantial Response

> One part essentially correct and one part partially correct

2   Developing Response

> One part essentially correct and one part incorrect
>
> *OR*
>
> Two parts partially correct

1   Minimal Response

> One part partially correct and one part incorrect

Question 4

Intent of Question

The primary goal of this question was to assess students' ability to set up, perform and interpret the results of a hypothesis test. More specific goals were to assess students' ability to: (1) state hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) compute the test statistic and p-value; (4) draw a conclusion, with justification, in the context of the problem.

Solution

Step 1: States a correct pair of hypotheses.

Let μ_A represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet and also receive a placebo.

Let μ_B represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet but receive the drug instead of a placebo.
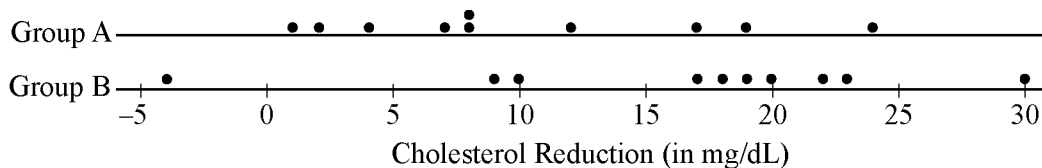
The hypotheses to be tested are $H_0: \mu_A = \mu_B$ versus $H_a: \mu_A < \mu_B$.

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

The appropriate procedure is a two-sample t-test.

When comparing two experimental treatments using a two-sample t-test, the subjects must be randomly assigned to the treatments. This condition is stated in the question (10 men were randomly assigned to group A and the remaining 10 men to group B).

The second condition is that the two populations are approximately normally distributed or the sample sizes are sufficiently large. Because of the small sample sizes (10 in each treatment group), we need to check whether it is reasonable to assume that the samples came from populations that are normally distributed. The following dotplots reveal slight skewness and a possible outlier for group B, but it appears reasonable to proceed with the two-sample t-test.

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

The test statistic is: $t = \dfrac{\overline{x}_A - \overline{x}_B}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} = \dfrac{10.20 - 16.40}{\sqrt{\dfrac{7.66^2}{10} + \dfrac{9.40^2}{10}}} \approx -1.62$

With df = 17.3, p-value ≈ 0.062.

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

Because the *p*-value is greater than the significance level of $\alpha = 0.01,$ we fail to reject $H_0.$ The data do not provide enough evidence at the 0.01 level of significance to conclude that the drug is effective in producing a mean cholesterol reduction beyond that provided by exercise and dietary advice.

Scoring

Steps 1, 2, 3 and 4 are each scored as essentially correct (E), partially correct (P) or incorrect (I).

Step 1 is scored as follows:

Essentially correct (E) if the response states hypotheses with correct comparisons between the means and defines the population means as the parameters.

Partially correct (P) if the response states hypotheses with correct comparisons between the means *OR* correctly defines the population means as the parameters, but not both.

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

Note: Defining the parameter symbols in context or simply using μ_A and $\mu_B,$ with subscripts clearly relevant to the context is sufficient for defining parameters.

Step 2 is scored as follows:

Essentially correct (E) if the response correctly includes the following three components:
1. Identifies the correct test procedure (by name or by formula).
2. Checks for random assignment to treatments.
3. Checks for normality.

Partially correct (P) if the response correctly includes exactly two of the three components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- Graphs of both distributions must be produced and described to check the normality condition.
- If the response calls for a *pooled* two-sample *t*-test, step 2 can be scored as E as long as the condition of equal variances is mentioned and checked by comparing the variability in the graphs or the sample standard deviations.
- If the response calls for applying a paired *t*-test, then step 2 is scored as I, but steps 3 and 4 can be scored as E if the test mechanics are correct in step 3 and the conclusion is correct in step 4.

Step 3 is scored as follows:

Essentially correct (E) if both the test statistic and p-value are correctly calculated.

Partially correct (P) if the test statistic is correctly calculated but not the p-value
   OR
if the test statistic is calculated incorrectly, but the correct p-value for the computed test statistic is given.

Incorrect (I) if the response fails to meet the criteria for E or P.

Step 4 is scored as follows:

Essentially correct (E) if the response provides a correct conclusion in context, also providing justification based on the linkage between the size of the p-value and the conclusion.

Partially correct (P) if the response provides a correct conclusion, including justification based on the size of the p-value, but not in context
   OR
if the response provides a correct conclusion, written in context, but without justification based on linkage to the p-value.

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

Notes
- If the conclusion is consistent with the p-value from step 3, and also in context with justification based the size of the p-value, then step 4 is scored as E (even if the p-value in step 3 is incorrect).
- A conclusion in step 4 that is equivalent to "accept H_0" (such as "we conclude that the drug is not effective") is not acceptable for an E. Such a response should be scored as P, provided that the conclusion is in context with justification based on the size of the p-value. Such a response should be scored as I if it lacks either context or linkage to the p-value.

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

Question 5

Intent of Question

The primary goals of this question were to assess students' ability to (1) determine the equation of the least squares regression line from a computer output; (2) use the slope of the least squares line to compare expected values of the response variable for different values of the explanatory variable; (3) recognize how to determine the proportion of variability in the response variable explained by the least squares line; (4) use computer output to determine whether the linear relationship between two quantitative variables is statistically significant.

Solution

Part (a):

The equation of the least squares regression line is

$$\text{predicted electricity production} = 0.137 + 0.240 \times \text{wind velocity}.$$

Part (b):

The slope coefficient of 0.240 indicates that for each additional mph of wind speed, the expected electricity production increases by 0.240 amperes. Thus, the expected electricity production is $10 \times 0.240 = 2.40$ amperes higher on a day with 25 mph wind velocity as compared to a day with 15 mph wind velocity.

Part (c):

The proportion of variation in electricity production that is explained by the linear relationship with wind speed is R^2, which the regression output reports to be 0.873.

Part (d):

Yes, there is very strong statistical evidence that the population slope differs from zero, so electricity production is linearly related to wind speed. For testing the hypotheses $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$, where β represents the population slope, the output reveals that the test statistic is $t = 12.63$ and the p-value (to three decimal places) is 0.000. Because the p-value is so small (much less than both 0.05 and 0.01), the sample data provide very strong statistical evidence that electricity production is linearly related to wind speed.

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 the response gives the correct equation *AND* includes the following two components:
1. Provides correct variable names (with context).
2. Uses a modifier such as "expected" or "predicted" or "estimated" (or a "hat" symbol) with the response variable, electricity production.

Partially correct (P) if the response gives the correct equation *AND* includes exactly one of the two components listed above.

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

Part (b) is scored as follows:

Essentially correct (E) if the response identifies and uses the correct slope value (0.240) *OR* the slope value identified in part (a) of the response
    AND
the response includes the following three components:
1. Shows work (correct multiplication or correct substitution into an appropriate expression).
2. Arrives at an answer.
3. Provides correct measurement units (amperes).

Note: Calculating predicted values for both wind speeds and taking their difference is sufficient, as long as measurement units are provided.

Partially correct (P) if the response identifies and uses the correct slope value (0.240) or the slope value identified in part (a) of the response *AND* includes exactly two of the three components listed above.

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

Part (c) is scored as follows:

Essentially correct (E) if response is 0.873.

Note: No work needs to be shown to earn an E, because the answer is read from the computer output.

Partially correct (P) if the response gives the value of adjusted R^2, rather than R^2, *OR* the response approximates (or rounds) the value of R^2.

Incorrect (I) if the response gives neither R^2 nor adjusted R^2, or if the response reports the square root of R^2.

Part (d) is scored as follows:

Essentially correct (E) if the response includes the following three components:
1. Gives the correct conclusion based on a test for the population slope.
2. Reports the correct p-value and/or t-statistic.
3. Provides linkage/justification between the p-value (or t-statistic) and the conclusion.

Partially correct (P) if the response provides exactly two of the three components listed above.

Note: If the wrong p-value is chosen, but the conclusion is consistent with that p-value and linkage or justification is provided, the response earns a P.

Incorrect (I) if the response fails to meet the criteria for E or P.

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

Question 6

<u>**Intent of Question**</u>

The primary goals of this question were to assess students' ability to (1) construct and interpret a confidence interval for a population proportion; (2) create a probability tree to represent a particular random process; (3) use a probability tree to calculate a probability; and (4) integrate provided information to create a confidence interval for an atypical parameter.

<u>**Solution**</u>

Part (a):

The appropriate inference procedure is a one-sample z-interval for a population proportion p, where p is the proportion of all United States twelfth-grade students who would answer the question correctly.

The conditions for this inference procedure are satisfied because:
1. The question states that the students are a random sample from the population, and
2. $n \times \hat{p} = 9{,}600 \times 0.28 = 2{,}688$ and $n \times (1 - \hat{p}) = 9{,}600 \times 0.72 = 6{,}912$ are both much larger than 10.
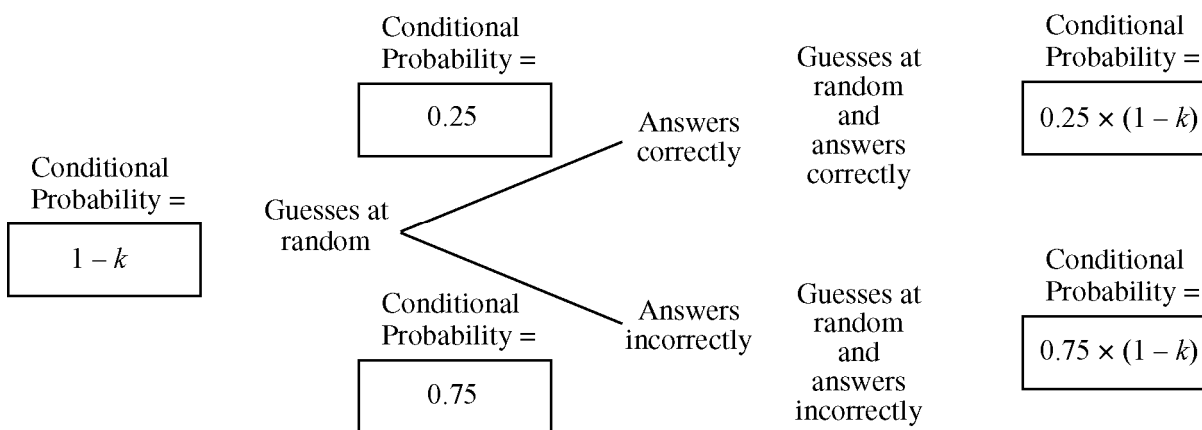
A 99 percent confidence interval for the population proportion p is constructed as follows:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.28 \pm 2.576 \sqrt{\frac{0.28(0.72)}{9{,}600}}$$
$$= 0.28 \pm 0.012$$
$$\rightarrow (0.268, 0.292)$$

We are 99 percent confident that the interval from 0.268 to 0.292 contains the population proportion of all United States twelfth-grade students who would answer this question correctly.

Part (b):

The five probabilities to be filled in the boxes are shown below.

Part (c):

P(answers correctly) =
P(knows correct answer and answers correctly) + P(guesses at random and answers correctly) =

$k + 0.25 \times (1 - k)$, which simplifies to $0.25 + 0.75k$, or $\dfrac{3k + 1}{4}$.

Part (d):

We want to estimate k, the proportion of all United States twelfth-grade students who actually know the answer to the history question.

From part (c) the probability that a randomly selected student correctly answers the question is $0.25 + 0.75k$. From part (a) we are 99 percent confident that this probability is between 0.268 and 0.292. Thus the endpoints for a confidence interval for k can be found by equating the expression $0.25 + 0.75k$ from part (c) to the endpoints of the interval from part (a) as follows:

$$0.25 + 0.75k = 0.268 \qquad\qquad 0.25 + 0.75k = 0.292$$
$$k = 0.024 \qquad\qquad\qquad k = 0.056$$

We are 99 percent confident that the interval from 0.024 to 0.056 contains the proportion of all United States twelfth-grade students who actually know the answer to the history question.

<u>Scoring</u>

This question is scored in four sections. Sections 1 and 2 are based on part (a), section 3 consists of parts (b) and (c) and section 4 consists of part (d). Each section is scored as essentially correct (E), partially correct (P) or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response correctly includes the following three components:
1. Identifies the correct inference procedure.
2. Checks the randomness condition.
3. Checks the large sample size condition.

Partially correct (P) if the response correctly includes exactly two of the three components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- The identification of the procedure must include "*z*," "proportion," and "interval."
- Stating the correct formula for a confidence interval for a proportion is sufficient for the first component.
- "Random sample given" is sufficient for the second component.

- To satisfy the third component, the response:
    - o Must check both the number of successes and the number of failures.
    - o Must use a reasonable criterion (for example, ≥ 5 or ≥ 10).
    - o Must provide numerical evidence (for example, $2{,}688 \geq 10$ and $6{,}912 \geq 10$, or $9{,}600 \times 0.28 \geq 10$ and $9{,}600 \times 0.72 \geq 10$).
- Any statement of hypotheses, definitions of parameters, statements of populations, etc. should be considered extraneous. However, if such statements are included and incorrect, this should be considered poor communication in terms of holistic scoring.
- Any checks of reasonable conditions, such as independence of observations, sample size less than 10 percent of population size, $9{,}600 > 30$, etc. should be considered extraneous. However, if a response includes an incorrect condition, such as population normality, reduce the score in section 1 from E to P or from P to I.
- Any reference to the central limit theorem should be treated as extraneous and not sufficient for the large sample size condition.

Section 2 is scored as follows:

Essentially correct (E) if the response correctly includes the following two components:
1. Calculates the interval.
2. Interprets the interval, including a confidence statement and correct parameter, in context.

Notes
- The critical value for the confidence interval must be for 99 percent confidence.
- If the response includes an incorrect formula or has incorrect values substituted into the formula, then the response does not earn credit for the calculation component, even if the final interval is correct.
- A response that makes minor arithmetic mistakes in the calculation of the interval is considered correct, as long as the resulting interval is reasonable.
- A correct interval that is stated only in the interpretation is considered sufficient for the first component.
- To identify the parameter, the response must refer to the proportion "who would answer the question correctly" or include a modifier for the proportion such as "population" or "true." An interpretation about the sample proportion (for example, "the proportion of students who answered correctly") is not sufficient for the second component.
- If the response provides only an interpretation of the confidence level instead of the confidence interval, the second component is considered incorrect. If an interpretation of the confidence level is given along with an interpretation of the confidence interval, both must be correct to be considered sufficient.
- A correct interpretation with an incorrect interval is sufficient for the second component.

Partially correct (P) if the response correctly includes exactly one of the two components.

Incorrect (I) if the response fails to meet the criteria for E or P.

Section 3 is scored as follows:

Essentially correct (E) if the response correctly includes the following two components:
1. In part (b) completes the tree diagram in terms of k.
2. In part (c) adds the correct results from the tree diagram.

Notes
*   If a response states "not k" or " k^{C} " in the first box, the first component is considered incorrect.
*   If the response to part (b) is incorrect, then part (c) is considered correct if the response is consistent with the response from part (b) or if the response to part (c) is correct.
*   The response to part (c) does not need to show a simplified expression.
*   The response to part (c) can be expressed as a fraction with the sum of the four branches in the denominator.
*   A response to part (c) that adds the appropriate probabilities from the tree but has an error in the simplification of the sum is still considered correct.
*   If the response to part (c) is expressed as $P(0.25 + 0.75k)$ or equivalent, the second component is considered incorrect.
*   If the tree diagram includes numbers only, adding the appropriate values is sufficient for the second component, provided that the sum is between 0 and 1.

Partially correct (P) if the response correctly includes exactly one of the two components.

Incorrect (I) if the response fails to meet the criteria for E or P.

Section 4 is scored as follows:

Essentially correct (E) if the response correctly includes the following three components:
1. Equates the expression from part (c) to a numerical estimate from part (a).
2. Uses the endpoints from part (a) to calculate a reasonable interval.
3. Interprets the resulting interval, including a confidence statement and correct parameter, in context.

Partially correct (P) if the response correctly includes exactly two of the three components.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
*   Using the point estimate $\hat{p} = 0.28$ from part (a) or the endpoints of the interval $(0.268, 0.292)$ from part (a) is sufficient for the first component.
*   A response that makes minor arithmetic mistakes in the calculation of the interval is considered correct, as long as the resulting interval is reasonable.
*   For the third component, the parameter must be the proportion of students who actually know the answer to the history question.
*   A response that creates a correct interval using linear transformations (of the point estimate and standard error/margin of error) is equivalent to transforming the endpoints and therefore is sufficient for the first two components.

Each essentially correct (E) section counts as 1 point. Each partially correct (P) section 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 decide whether to score up or down, depending on the overall strength of the response and communication, particularly in parts (a) and (d). However, a response that earns a P or an I in section 4 cannot receive a score of 4.