AP® Statistics
2002 Scoring Guidelines

These materials were produced by Educational Testing Service® (ETS®), which develops and administers the examinations of the Advanced Placement Program for the College Board. The College Board and Educational Testing Service (ETS) are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

The College Board is a national nonprofit membership association dedicated to preparing, inspiring, and connecting students to college and opportunity. Founded in 1900, the association is composed of more than 4,200 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 22,000 high schools, and 3,500 colleges, through major programs and services in college admission, 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 equity and excellence, and that commitment is embodied in all of its programs, services, activities, and concerns.

Question 1

Solution

Part (a):

The precision of the estimates of γ has gotten better over time. This is indicated by the fact that the intervals

$$\text{value} \pm \text{(margin of error)}$$

shown in the figure become narrower over time, indicating that the margin of error is getting smaller.

Part (b):

The value of $\gamma = 0$ is not included in any of the 21 intervals. This indicates that 0 is not a plausible value for γ. There is no support for Newton's theory.

Part (c):

The support for Einstein's theory that $\gamma = 1$ is quite strong. Most of the intervals contain the value 1, and the more recent intervals, where the precision is greater, suggest that the value of γ is at least very close to 1.

Scoring

Part (a) is scored as incorrect (I), partially correct (P), or essentially correct (E). The response is essentially correct if the student indicates that precision is increasing. (If the student also explains how he or she can tell this from the figure, this is a plus.)

If the student incorrectly says that precision is decreasing, but gives a good explanation that is tied to the figure, the response is scored as partially correct.

Part (b) is scored as incorrect (I), partially correct (P), or essentially correct (E). To be scored as essentially correct, the response must say that there is no (or very weak) support *and* give a valid reason for this conclusion based on the intervals in the figure.

If the student states only that there is no (or weak) support, but does not say how this follows from the intervals in the figure, the response is scored as partially correct.

Part (c) is scored as incorrect (I), partially correct (P), or essentially correct (E). To be scored as essentially correct, the response must say that there is strong support *and* give a valid reason for this conclusion based on the intervals in the figure.

If the student states only that there is strong support, but does not say how this follows from the intervals in the figure, the response is scored as partially correct.

Question 1 (cont'd.)

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 parts partially correct.
OR
No parts essentially correct and two parts partially correct.

Question 2

Solution

Part (a):

A **paired design** is used in which each subject receives a pair of boots where one boot is treated with the new method and the other with the current method.

Subjects should be randomly assigned to one of two groups. Group 1 would have the new method applied to the right boot; group 2 would have the new method applied to the left boot.

OR

For each subject, whether the new method is applied to the right or left boot is determined at random.

OR

A **crossover design** is used in which each subject receives a pair of boots, both of which were treated with one treatment. The boots are used for three months and then exchanged for a second pair of boots, both of which were treated with the other treatment. These boots are then used for the next three months.

Subjects should be randomly assigned to one of two groups. One group receives boots with the new treatment first and the other group receives boots with the current method first.

NOTE: Additional appropriate blocking schemes are considered extraneous.

Part (b):

The design could be double blind, as long as both the *subjects* and the person *evaluating* the boots for water damage do not know which boots were treated with the new method and which were treated with the current method.

NOTE: If the student does something unexpected in part (a) and gives a design that actually cannot be double blind, then part (b) could be considered correct provided the response explains why the design could not be double blind.

Scoring

A student response is scored as **E** (essentially correct), **P** (partially correct), or **I** (incorrect) for each of the following key elements:

1. **Design**
   - **E** - paired design (may be described as blocking on individual) or crossover design
   - **P** - 2 or more groups (e.g., Completely Randomized Design)
   - **I** - no grouping or grouping with no treatments specified
2. **Implementation**: Randomization appropriate to the design
   - **E** - Written description of appropriate randomization
   - **P** - Incomplete or incorrect description of randomization
   - **I** - No description of randomization

   NOTE: (1) Diagram alone can be scored at most a **P**.
   (2) The randomization must apply to the allocation or assignment of subjects to the treatment groups or the allocation of treatments to the subjects.
   (3) Randomization to select the 100 volunteers without assignment to the treatment groups is scored an **I**.

Question 2 (cont'd.)

3. **Double blind**: Explanation in parts (a) and/or (b) that shows understanding of what it means for an experiment to be double blind.
   - **E** - response indicates that blinding applies to both the evaluator and subjects.
   - **P** - response recognizes that blinding applies to the subjects and at least one other party, whether or not they think that this can be accomplished; the other party may not be correctly identified.
   - **I** - response fails to recognize that both the subject and another party must be blinded or is missing or irrelevant.

Score as Design - Randomization - Double Blind

4 Complete Response

E E E

3 Substantial Response

Any one of the following combinations:

E E P	P E E	P E P *
E E I		
E P E		

2 Developing Response

Any one of the following combinations:

E P P	P E I	I E E	P E P *
E P I	P P E	I P E	
E I E	P P P		
E I P	P I E		

1 Minimal Response

Any one of the following combinations:

E I I	P P I	I E P
I E I	P I P	
I I E	I P P	

Question 2 (cont'd.)

0 **No Credit**

P I I   I I I
I P I
I I P

*** P E P** may be scored as either a **2** or a **3**:

(1) If the description of the randomization *only* says, "Randomly allocate", then score P E P a 2.

(2) If the description of the randomization says, "Randomly allocate", but also contains greater detail about the randomization or the inclusion of blocking in the design or other statistical thinking, then score P E P a 3.

Question 3

Solution

Part (a):

For runner 3

$$P(time < 4.2) = P\left(z < \frac{4.2 - 4.5}{.14}\right) = P(z < -2.14) = 0.0162 \quad (\textit{from table})$$

OR

$$P(time < 4.2) = 0.0160622279 \quad (\textit{from} \text{ Calculator})$$

It is possible but unlikely that runner 3 will run a mile in less than 4.2 minutes on the next race. Based on his running time distribution, we would expect that he would have times less than 4.2 minutes less than 2 times in 100 races in the long run.

OR

It is possible but unlikely that runner 3 will run a mile in less than 4.2 minutes on the next race because 4.2 is more than 2 standard deviations below the mean. Since the running time has a normal distribution, it is unlikely to be more than 2 standard deviations below the mean.

Part (b):

$$\mu_T = \mu_1 + \mu_2 + \mu_3 + \mu_4 = 4.9 + 4.7 + 4.5 + 4.8 = 18.9$$

The runners' times are independently distributed, therefore

$$\sigma_T^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 = (.15)^2 + (.16)^2 + (.14)^2 + (.15)^2 = 0.0902$$

$$\sigma_T = \sqrt{.0902} = 0.3003$$

Part (c):

$$P(team\ time < 18.4) = P\left(z < \frac{18.4 - 18.9}{.3003}\right) = P(z < -1.67) = 0.0475 \quad (\textit{from table})$$

OR

$$P(team\ time < 18.4) = 0.0479561904 \quad (\textit{from} \text{ Calculator})$$

Question 3 (cont'd.)

Scoring

Each part is scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is *essentially correct* if:

the probability is calculated correctly, it is correctly assessed as unlikely, and the justification is acceptable
OR
the student does not compute the probability, but appeals to the fact that a time of 4.2 minutes is more than 2 standard deviations below the mean of a normal distribution and then uses this information to reach a conclusion with appropriate communication.

Part (a) is *partially correct* if:

the probability computed is not correct (for example, $P(z > -2.14)$ or $P(z < +2.14)$ might be computed), but the given probability is correctly assessed
OR
an argument is based on the number of standard deviations from the mean without invoking normality.

Part (b) is *essentially correct* if both the mean and the standard deviation of the team time distribution are correctly computed (except for purely arithmetic mistakes).

Part (b) is *partially correct* if only one of these is correctly computed (except for purely arithmetic mistakes).

CAUTION: A standard deviation of .3 (numerically correct) can arise from this incorrect calculation: $\dfrac{(.15+.16+.14+.15)}{\sqrt{4}} = 0.3$

Part (c) is *essentially correct* if the probability is correctly calculated using a mean which is either correct or carried from (b) as well as a standard deviation which is either correct or carried from (b).

Part (c) is *partially correct* if:

both the mean and standard deviation are correct or carried from (b), but the computed probability is incorrect
OR
the mean or standard deviation is incorrectly derived from (b) but the subsequent probability calculation is correct.

Question 3 (cont'd.)

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
One part essentially correct and one part partially correct.
OR
Three parts partially correct.

1 Minimal Response
One part essentially correct and zero parts partially correct.
OR
No parts essentially correct and two parts partially correct.

Question 4

Solution

Part (a):

Predicted cost = 1136 + 14.673 (number of passenger seats)

OR

$\hat{y} = 1136 + 14.673x$ where y = operating cost per hour and x = number of passenger seats

Part (b):

- The value of the correlation coefficient

  $r = +\sqrt{0.570} = 0.755$ (r is positive because the scatterplot shows a positive association)

- The interpretation of correlation

  There is a moderate (or strong) positive linear relationship between operating costs per hour and number of passenger seats.

  OR

  Fifty-seven percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats AND the relationship is positive.

Part (c):

No. The equation of the least-squares regression line is influenced by the three points in the upper right-hand corner and the two points in the lower left-hand corner of the scatterplot. The seven remaining points (with number of seats in the 250 to 350 range) would have a negative correlation. Hence, the slope of the recalculated least-squares regression line is negative.

Scoring

The student response should include the following elements:

1. the correct equation of the least squares regression line with variables correctly defined;
2. the correct value for the correlation coefficient;
3. a correct interpretation of the given correlation coefficient; and
4. a complete explanation of why the given least-squares line would not be appropriate for describing the relationship over the restricted range.

Each element is scored as essentially correct (E), partially correct (P), or incorrect (I).

Question 4 (cont'd.)

Part (a) addresses the first element.

Element one is:

essentially correct if the solution has the correct equation and variables are defined correctly.

partially correct if only the equation is correct.

incorrect if the equation is not stated correctly.

Part (b) addresses the second and third elements.

Element two is:

essentially correct if the student's solution states that $r = 0.755$.

partially correct if the student's solution only states that $r = \pm\, 0.755$.

incorrect if the student states any other value of r including $r = 0.726$ (square root of R-Sq (adj)).

Element three is:

essentially correct if the student's solution

addresses, based on a correct understanding of the correlation coefficient, three or four of the following:

- type of relationship
- strength
- direction
- context

OR

states, based on a correct understanding of r^2:

- that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats

  AND

- that the relationship is positive.

Note: If the student gives a correct interpretation of r but then incorrectly explains r^2, this is considered a parallel solution and is incorrect.

Question 4 (cont'd.)

partially correct if the student's solution

addresses exactly two of the following – type of relationship (linear), strength, direction, and context (based on a correct understanding of the correlation coefficient).

OR

only states that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats (based on a correct understanding of r^2) – BUT – does not state that the relationship is positive.

NOTE: Element three may be scored essentially or partially correct if the student uses a reasonable r (between 0 and 1) or R-Sq (adj) value.

Part (c) addresses the fourth element.

Element four is **essentially correct** if the student's solution

states that the existing line is not a good fit for the remaining seven points and correctly explains that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

Element four is **partially correct** if the student's solution

explains why the existing line is not a good fit for the remaining seven points but does not communicate that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

OR

removes fewer than the specified five points, but gives a correct interpretation of the effect on the correlation or slope of the least-squares regression line.

Question 4 (cont'd.)

For elements 1 through 4, essentially correct responses count as one element and partially correct responses count as one-half of an element.

4 Complete Response

Four elements correct.

3 Substantial Response

Three elements correct.

2 Developing Response

Two elements correct.

1 Minimal Response

One element correct.

If a paper is between two scores (for example, 2 1/2 elements) use a holistic approach to determine whether to score up or down depending on the strength of the response and quality of communication.

Question 5

Solution

Part (a):

Possibilities include:

1. p_E = proportion of early birds who recall dreams
p_N = proportion of night owls who recall dreams

H_0: $p_E - p_N = 0$ vs. H_a: $p_E - p_N < 0$ OR H_0: $p_E = p_N$ vs. H_a: $p_E < p_N$

OR

p_E = proportion of early birds who do not recall dreams
p_N = proportion of night owls who do not recall dreams

H_0: $p_E - p_N = 0$ vs. H_a: $p_E - p_N > 0$ OR H_0: $p_E = p_N$ vs. H_a: $p_E > p_N$

NOTE: Either of these, BUT NOT BOTH, can be used as one of the possibilities for part (a).

2. p_E = proportion of early birds who recall 5 or more dreams
p_N = proportion of night owls who recall 5 or more dreams

H_0: $p_E - p_N = 0$ vs. H_a: $p_E - p_N < 0$ OR H_0: $p_E = p_N$ vs. H_a: $p_E < p_N$

OR

p_E = proportion of early birds who do not recall 5 or more dreams
p_N = proportion of night owls who do not recall 5 or more dreams

H_0: $p_E - p_N = 0$ vs. H_a: $p_E - p_N > 0$ OR H_0: $p_E = p_N$ vs. H_a: $p_E > p_N$

NOTE: Either of these, BUT NOT BOTH, can be used as one of the possibilities for part (a).

3. M_E = median number of dreams early birds recall
M_N = median number of dreams night owls recall

H_0: $M_E - M_N = 0$ vs. H_a: $M_E - M_N < 0$ OR H_0: $M_E = M_N$ vs. H_a: $M_E < M_N$

NOTE:
1. A complete response for part (a) requires two pairs of hypotheses.
2. Hypotheses for a chi-square test of homogeneity are not correct since this is a one-sided test.

Question 5 (cont'd.)

Part (b):

Part 1: States a correct pair of hypotheses, identifies a correct test (by name or by formula) and checks appropriate requirements.

μ_E = mean number of dreams early birds recall
μ_N = mean number of dreams night owls recall

$H_0 : \mu_E = \mu_N$
$H_a : \mu_E < \mu_N$

OR

$H_0 : \mu_E - \mu_N = 0$
$H_a : \mu_E - \mu_N < 0$

Two-sample *t*-test

$$t = \frac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\frac{s_E^2}{n_E} + \frac{s_N^2}{n_N}}}$$

Requirements:
1. Problem states that independent random samples were taken.
2. Normal population distributions or large samples. Since these are not normal, we need to note that $n_E = 100$ and $n_N = 100$ are both large in order to perform the *t*-test.

OR

Two-sample *z*-test

$$z = \frac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\frac{s_E^2}{n_E} + \frac{s_N^2}{n_N}}}$$

Requirements:
1. Problem states that independent random samples were taken.
2. Since $n_E = 100$ and $n_N = 100$ are both large, it is OK to perform the approximate *z*-test.

OR

Pooled *t*-test

$$t = \frac{\bar{x}_E - \bar{x}_N - 0}{s_p\sqrt{\frac{1}{n_E} + \frac{1}{n_N}}}$$ where s_p^2 is the pooled variance.

Requirements:
1. Problem states that independent random samples were taken.
2. Normal population distributions or large samples. Since these are not normal, we need to note that $n_E = 100$ and $n_N = 100$ are both large in order to perform the *t* test.
3. The sample standard deviations $s_E = 6.94$ and $s_N = 5.88$ are reasonably close, so it is OK to assume that the two population variances are equal, i.e. $\sigma_E^2 = \sigma_N^2$.

Question 5 (cont'd.)

Part 2: Correct mechanics, including the value of the test statistic, df, and P-value (or rejection region)

$$t = \frac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\frac{s_E^2}{n_E} + \frac{s_N^2}{n_N}}} = \frac{7.26 - 9.55}{\sqrt{\frac{(6.94)^2}{100} + \frac{(5.88)^2}{100}}} = -2.52$$

So, t = -2.52 df = 192 P-value = 0.006

- It is OK to use conservative df of 99.
- Using t-tables: P-value < 0.01
- Using calculator: $t = -2.517578$, P-value = 0.006304, df = 192.799
- Using z-test: P-value = 0.005908
- Using z-table: 0.0059 < P-value < 0.0060
- The pooled t-test results in the same value of t, but a df of 198.

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

Because the P-value is small (or less than an α selected and stated by the student), reject H_0. There is convincing evidence that the mean number of dreams night owls recall is greater for than the mean number of dreams early birds recall.

Question 5 (cont'd.)

Scoring

Part (a):
Essentially correct (E) if two distinct pairs of hypotheses are given and the parameters in the hypotheses are defined.

Partially correct (P) if only one pair of correct hypotheses is given and the parameters in these hypotheses are defined or if two "approved" pairs of hypotheses are given but the parameters are poorly defined.

Part (b):
Each of the 3 parts of the hypothesis test is scored either as correct (E) or incorrect (I).

- Because the hypotheses are given in the statement of part (a), they need not be restated here. However, if wrong hypotheses are stated, then part 1 is scored as incorrect.
- Because the problem states that samples are random, it is OK if a student doesn't repeat this.
- Some reference to both samples being large is essential.
- For the pooled *t*- test, some comment on the reasonableness of such an assumption is necessary. It is not sufficient just to say that population variances are equal.

4 Complete Response
Part (a) essentially correct and all three parts of the hypothesis test correct.

3 Substantial Response
Part (a) essentially correct and two parts of the hypothesis test correct.
OR
Part (a) partially correct or incorrect and three parts of the hypothesis test correct.

Part (a) partially correct and two parts of the hypothesis test correct may be scored either as a 2 or a 3, depending on the overall strength of the paper.

2 Developing Response

Part (a) essentially correct and one part of the hypothesis test correct.
OR
Part (a) incorrect and two parts of the hypothesis test correct.

Part (a) partially correct and two parts of the hypothesis test correct may be scored either as a 2 or a 3, depending on the overall strength of the paper.

1 Minimal Response

Part (a) essentially correct.
OR
Part (a) partially correct and zero or one parts of the hypothesis test correct.
OR
Part (a) is incorrect and one part of the hypothesis test correct.

Question 6

Solution

Part (a):

p = proportion of students at this university who would respond S to the question, "Do you prefer S or F?"

Large sample confidence interval for a population proportion.

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

State and Check Assumptions and Conditions:

Simple random sample (given in the problem stem—need not be mentioned in solution). Also need large sample with $n\hat{p} \geq 5$ and $n\left(1-\hat{p}\right) \geq 5$. Here,

$n\hat{p} = 185$ $\quad\quad$ $n(1-\hat{p}) = 139$ are both greater than 5 (or 10)

or $\hat{p} \pm 3\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$ is entirely in the interval (0,1).

We assume the university has at least 10(324) = 3240 students ($N \geq 10n$).

Calculations:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = 0.571 \pm 1.96\sqrt{\frac{(.571)(.429)}{324}} = 0.571 \pm 0.054 = (0.517,\ 0.625)$$

Calculator solution: (0.5171, 0.62488). The procedure is specified in the stem, but students still need to check assumptions and conditions.

Interpretation:

Based on this sample, we can be 95 percent confident that the proportion of students at this university who would respond S to the question, "Do you prefer S or F?" is between 0.517 and 0.625.
OR
We have 95 percent confidence that the interval (0.5171, 0.62488) captures the proportion of students who would respond "S" to the question, "Do you prefer S or F?"

Question 6 (cont'd.)

Part (b):

Meaning of Confidence Level:

In repeated sampling, 95 percent of the intervals produced using this method will contain the proportion of students at this university who would respond S to the question "Do you prefer S or F?"

Part (c):

Approach 1: Hypothesis Test – Two Proportion *Z*- test

States a Correct Pair of Hypotheses:

H_0: $p_1 - p_2 = 0$
H_a: $p_1 - p_2 \neq 0$

where
p_1 = proportion of students at this university who would respond S with the original question wording
p_2 = proportion of students at this university who would respond S with the revised question wording

Note: A one-sided test is incorrect.

Name Test and State and Check Assumptions and Conditions:

Identifies a correct test (by name or by formula), and checks appropriate assumptions.

Two sample *z*-test for proportions

Note: Problem states that samples are random samples, so this does not need to be addressed in the assumptions.

Large samples: $n_1\hat{p}_1 = 185$; $n_1(1-\hat{p}_1) = 139$; $n_2\hat{p}_2 = 68$; $n_2(1-\hat{p}_2) = 88$
All are greater than 5 (or 10), so the sample sizes are large enough.

Question 6 (cont'd.)

Calculations:

Correct mechanics, including P-value or rejection region (except for minor arithmetic errors)

For two sample proportion z-test:

$\hat{p}_1 = \frac{185}{324} = 0.571$ $\qquad\qquad$ $\hat{p}_2 = \frac{68}{156} = 0.436$

$$\hat{p} = \frac{185 + 68}{324 + 156} = \frac{253}{480} = 0.527$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{0.571 - 0.436}{\sqrt{\frac{(0.527)(0.473)}{324} + \frac{(0.527)(0.473)}{156}}} = \frac{0.135}{0.0487} = 2.77$$

P-value = 2(.0028) = .0056 from tables

(Calculator: $z = 2.776554085$, P-value = .0054939656)

If the proportions are not pooled, then $z = 2.795$ and $p = 0.00518$.

Conclusion:

Since the P-value (0.0055) is so small, we reject the null hypothesis that the proportions of this university's students who would respond S to the two survey questions are equal. We believe the order in which the choices are given affects the students' response.

Stating a correct conclusion in the context of the problem, using the result of the statistical test (i.e., linking the conclusion to the result of the hypothesis test).

If both an α and a P-value are given, the linkage 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.

If the P-value in part 3 is incorrect but the conclusion is consistent with the computed P-value, part 4 should be considered as correct.

Question 6 (cont'd.)

Approach 2: Hypothesis Test – Chi-square Test for Homogeneity

States a correct pair of hypotheses:

H_0: population response proportions are the same for the two question wordings
H_a: population response proportions are not the same for the two question wordings

NOTE: Although the computations are the same, stating the hypotheses in terms of independence is not correct.

Name Test and State and Check Assumptions and Conditions:

Identifies a correct test (by name or by formula), and checks appropriate assumptions.

Chi-square test for homogeneity

Observed counts

	S	F	row total
Original wording	185	139	324
Revised wording	68	88	156
Column total	253	227	480

Expected counts

	S	F
Original wording	170.775	153.225
Revised wording	82.225	73.775

NOTE: Problem states that samples are random samples, so this does not need to be addressed in the assumptions. All expected counts are greater than five, so the sample sizes are large enough.

Calculations:

Correct mechanics, including P-value or rejection region (except for minor arithmetic errors)

For chi-square test:

$$\chi^2 = \sum\frac{(O-E)^2}{E} = \frac{(185\ -\ 170.775)^2}{170.775} + \ldots + \frac{(88\ -\ 73.775)^2}{73.775} = 7.709253$$

df = 1 from tables $0.005 < \text{P-value} < 0.01$

(Calculator: P-value = 0.0054938481)

Question 6 (cont'd.)

Conclusion:

Since the P-value is so small, we reject the null hypothesis that the proportions of students at this university who would respond S are the same for the two survey questions. We believe the order in which the choices are given affects the students' response.

Stating a correct conclusion in the context of the problem, using the result of the statistical test (i.e., linking the conclusion to the result of the hypothesis test).

If both an α and a P-value are given, the linkage 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.

If the P-value in part 3 is incorrect but the conclusion is consistent with the computed P-value, part 4 should be considered as correct.

Approach 3: Two sample confidence interval

Name Test and State and Check Assumptions and Conditions:

Two-sample confidence interval.

Problem states that samples are random samples, so this does not need to be addressed in the assumptions.

Large samples:

$n_1\hat{p}_1 = 185; \quad n_1(1-\hat{p}_1) = 139; \quad n_2\hat{p}_2 = 68; \quad n_2(1-\hat{p}_2) = 88$

All are larger than 5 (or 10), so the sample sizes are large enough.

Calculations:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

90 percent CI: (0.05565, 0.21453)
95 percent CI: (0.04044, 0.22974)
99 percent CI: (0.01069, 0.25949)

Question 6 (cont'd.)

Conclusion:

Since the confidence interval does not include 0, there is evidence that the proportions of students at this university who respond S is not the same for the two question wordings.

Approach 4: (This approach can score at most partially correct for part (c))

One sample confidence interval for p_2.

All checks and assumptions must be included (same as in section (a)).

95 percent CI: (0.35808, 0.51371)

Since this interval does not overlap with the interval computed in part (a), $(0.517, 0.625)$, conclude that the proportions of this university's students who would respond S is not the same for the two question wordings.

Part (d):

If the sample sizes had been equal, it would be reasonable to combine the data from the two samples by pooling, which would be equivalent to averaging the two proportions in this case. But, since the wording of the question makes a difference, and more people were asked the original version than were asked the revised version, we cannot just pool.

OR

It is only reasonable to pool estimates if they are estimating the same population parameter. Here the proportion who would respond S differs with the survey question so the estimates should not be pooled.

Approach 1:

One reasonable approach would be to scale sample 2 up to a sample size of 324 while maintaining the same sample proportion. To do this, the number of S's in sample 2 would be multiplied by a factor of 2.076923 (It is OK if the student uses a factor of 2 for simplicity). This would result in two samples of sizes 324 with 185 S's in sample 1 and 141 S's in sample 2. This would result in an estimate of those who prefer S of

$$\hat{p} = \frac{185 + 141}{648} = \frac{326}{648} = 0.503$$

Question 6 (cont'd.)

Note: A comparable approach would be to scale sample 1 down to a sample size of 156 by using a multiplier of 0.481481 to obtain

$$\hat{p} = \frac{89 + 68}{312} = \frac{157}{312} = 0.503$$

This is very close to 0.5.

Approach 2:

The approach described above is equivalent to just averaging the two proportions, and so averaging the two given proportions is also an acceptable approach.

$$\hat{p} = \frac{\frac{185}{324} + \frac{68}{156}}{2} = \frac{0.571 + 0.436}{2} = 0.503$$

Note: A weighted average of the two proportions (with weights proportional to sample size) is equivalent to the given pooled value. If the student rejects the pooled value and proposes a weighted average of the two sample proportions as an alternative, part (d) is incorrect.

Scoring for Question 6

Parts (a) and (b) should be read together.

Part (a) is scored as essentially correct (E) if the assumptions are checked, the interval is computed correctly (except for minor arithmetic errors), and a correct interpretation of the interval is given.

It is partially correct (P) if the interval is computed correctly (except for minor arithmetic errors) but either the assumptions or the interpretation is not correct.

Otherwise, part (a) is scored as incorrect (I).

Part (b) is scored as either essentially correct (E) or incorrect (I). It is not possible to score partially correct on this part.

Part (c) is scored as essentially correct (E) if all four parts of the hypothesis test are correct. It is scored as partially correct (P) if two or three of the components of the test are correct. Otherwise, it is scored as incorrect (I).

Part (d) is scored as essentially correct (E) if the student produces a reasonable estimate that takes the different sample sizes into account, the explanation is correct and communication is good.

Question 6 (cont'd.)

It is partially correct (P) if

- a reasonable estimate is produced but the explanation is incorrect or weak
  OR
- a good explanation of why not to use the pooled estimate but no reasonable alternative is given

For parts (a), (b), and (c), essentially correct responses count as 1 part and partially correct responses count as ½ part.

4 **Complete Response**
Four parts correct.

3 **Substantial Response**
Three parts correct.

2 **Developing Response**
Two parts correct.

1 **Minimal Response**
One part correct.

If a paper is between two scores (for example, 2 ½ parts) use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

Note: If the paper is between two scores and (a) or (c) has the interpretation correct, then round up. If neither is correct, round down.