Question 1

1. A student interested in purchasing her first vehicle wanted to research the typical fuel efficiency of some of the most popular vehicles. The following data show the city miles per gallon (mpg) of the twelve best-selling vehicles in the previous calendar year, as reported by one widely used car buying website that the student visited.

27 16 29 17 27 26 25 17 18 28 30 19

(a) Display these data in a dotplot.

(b) Use your dotplot from part (a) to describe the main features of this city mpg distribution.

(c) Why would it be misleading for this student to use only a measure of center for this city mpg distribution as an indication of the typical fuel efficiency for popular vehicles?

Intent of Question

The primary goals of this question are to assess a student's ability to (1) construct a dotplot from a given data set, (2) describe the important features of the plot, and (3) discuss how a single measure of centrality fails to convey important features of the plot.

Solution

Part (a):

![Dotplot](image)

Part (b):

The most striking feature of the plot is that the mpg values cluster into two groups, one concentrated in the upper teens and one the other in the upper twenties. There are no values between 20 and 24.

Part (c):

A measure of center might fall between the two groups (as does the mean of 23.25 here) where there are no data values and would not provide an accurate picture of the typical fuel efficiency for popular vehicles. It would not indicate that the most popular vehicles tend to fall into either a low fuel efficiency group (likely made up of trucks and SUVs) or a high fuel efficiency group (likely made up of sedans and coupes).
**2015 Clemson AP Statistics Practice Exam – Scoring Guidelines**

**Question 1**

**Scoring**

This question is scored in four sections: section 1 is part (a), and sections 2 to 4 consist of elements of parts (b) and (c).

Section 1 is scored as essentially correct (E), partially correct (P), or incorrect (I).

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

- Essentially correct (E) if the response includes a correctly constructed dotplot.
- Partially correct (P) if the dotplot is correct except for labels.
- Incorrect (I) for any other type of plot.

*Note:* One or two misplaced or omitted values can still be considered essentially correct as long as the important features of the display are not altered.

Parts (b) and (c) are scored together in three sections, each of which is scored as essentially correct (E), partially correct (P), or incorrect (I).

**Section 2** is scored as follows:

- Essentially correct (E) if in either part (b) or (c) the response clearly notes
  1. that there are two groups;
  2. that there is a gap in the middle of the distribution;
  3. the relative or specific positions of the two groups OR the location of the gap.

- Partially correct (P) if the response notes two out of the three points.

- Incorrect (I) if the response makes note of only one or none of the three points.

**Section 3** is scored as follows:

- Essentially correct (E) if in part (b) or part (c) the solution is given in the context of the problem and is communicated well.

- Partially correct (P) if the response mentions the context (for instance, using the abbreviation “mpg”), but communication of the context is weak.

- Incorrect (I) if the context is not mentioned at all.
Section 4 is scored as follows:

Essentially correct (E) if in part (c) a valid reason is given for why a measure of center is not sufficient for data of this type (with the two groups and a gap).

Partially correct (P) if a response gives a general reason for why a measure of center is not sufficient (for instance, by stating that center alone without some measure of spread is never sufficient) or if the response compares the mean and median and cites outliers or skewness as the reason why a measure of center is not sufficient.

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

Note: Section 4 can be at most partially correct if a student does not recognize the groups or gap.

Each essentially correct (E) section counts as 1 point, and a partially correct (P) section 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, 2½ points), score down.
2. The remains of five concentric agricultural terraces set in circles of increasing depths can be found in Peru, at the ancient site of Moray in the Sacred Valley of the Incas. These terraces were used by the Incas to grow crops of varying species. The current local inhabitants wish to use them to compare the yields of four varieties of corn: Kulli black, Oaxacan green, Chullpi yellow, and Incan giant.

Due to their design, each terrace differs greatly in terms of soil type, irrigation level, and amount of sunlight. Each terrace has been divided into eight sections, resulting in 40 sections total. The diagram below is an overhead view of the five terraces.

To study yields, the inhabitants plan to assign the four corn varieties completely at random to one of the 40 sections while ensuring that each corn variety is represented the same number of times.

(a) A second way to design the experiment is to use blocking while still ensuring that each corn variety is represented the same number of times within each block. Identify the factor to be used to create the blocks and justify your choice.

(b) Describe a process by which to assign the corn varieties to the sections in the randomized complete block design.

(c) In the context of this situation, describe one statistical advantage of selecting a randomized complete block design as opposed to the completely randomized design.
Intent of Question

The primary goals of this question are to assess a student’s ability to: (1) use blocking in designing an experiment, (2) describe a mechanism for randomly assigning treatments to experimental units in the context of a randomized block design, and (3) explain the statistical advantage of incorporating blocks in an experiment.

Solutions

Part (a):

The agricultural terraces should be used as blocks. Terraces should be used because the blocks differ greatly in terms of soil type, irrigation level, and amount of sunlight – all factors that would likely affect the corn yields, but which are not of interest to the current inhabitants.

Part (b):

Uniquely label the sections within the first terrace 1 through 8 (inclusive). Generate 8 random integers between 1 and 8 (inclusive) without replacement via software or a random number table. The sections corresponding to the first 2 random numbers will receive the first variety of corn, the sections corresponding to the next 2 random numbers will receive the second variety of corn, the sections corresponding to the next 2 random numbers will receive the third variety of corn, and the sections corresponding to the final 2 random numbers will receive the fourth variety of corn. Repeat this process for each of the four remaining terraces. The assignment results in the 8 sections within each terrace being assigned to one of the four corn varieties, with each variety appearing twice in each terrace.

Part (c):

Because the terraces differ in terms of soil type, irrigation level, and sunlight we expect the yields to be different between the terraces, regardless of the variety of corn. It is possible that all of one variety may be on the lowest terrace while all of another variety is on the highest terrace, resulting in confounding between the corn variety and the terrace. Blocking on terrace removes that confounding factor and allows for a comparison of corn varieties within each terrace.
2015 Clemson AP Statistics Practice Exam – Scoring Guidelines
Question 2

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 includes the following two components:
1. The agricultural terraces have been selected as the blocking factor.
2. The justification for the blocking factor demonstrates an understanding that blocks should consist of experimental units (plots of land) that are similar with respect to factors that affect the response (corn yield).

Partially correct (P) if the response includes only 1 of the 2 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 includes the following two components:
1. An appropriate method for assigning corn varieties within each block has been described that results in each corn variety appearing twice within each terrace.
2. A method of randomization has been described that can be implemented by the reader.

Partially correct (P) if the response correctly provides only 1 of the 2 components listed above.

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

Note: Simply saying “use a random number table” or “flip a coin” is not sufficient to get credit for the method of randomization.

Part (c) is scored as follows:

Essentially correct (E) if the response provides a valid description of the advantage of a block design in this experiment.

Partially correct (P) if the response provides an incomplete description that indicates an understanding of confounding in this experiment. For example, the response indicates that differences in the terrace conditions can affect corn yield but fails to link this to the inability to distinguish between terrace differences and corn variety effects.

Incorrect (I) if the response does not meet the criteria for E or P.
2015 Clemson AP Statistics Practice Exam – Scoring Guidelines

Question 2

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
3. A triathlon is an athletic event that consists of swimming, cycling and running. At the Lake Tahoe Ironman Triathlon competitors swim for 2.4 miles, cycle for 112 miles, and run for 26.2 miles. Competitors are timed for each individual event and then receive an overall time which is the sum of the three different event times. The winner of the triathlon is the competitor with the lowest total time. The times for all competitors for each event at the Lake Tahoe Ironman Triathlon are approximately normally distributed. Their means and standard deviations, in minutes, are summarized in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swimming</td>
<td>80</td>
<td>23</td>
</tr>
<tr>
<td>Cycling</td>
<td>390</td>
<td>60</td>
</tr>
<tr>
<td>Running</td>
<td>300</td>
<td>46</td>
</tr>
</tbody>
</table>

(a) What is the probability that a randomly chosen competitor has a swim time less than 1 hour?

(b) How fast would a competitor need to run in order to be in the fastest 2.5% of runners?

(c) On her last triathlon, Christine’s total time was 900 minutes. She would like to know how well she performed relative to the other competitors. She determines that the mean for the distribution of total times is equal to

$$\mu_T = \mu_S + \mu_C + \mu_R = 770 \text{ minutes}$$

and the standard deviation for the distribution of total times is equal to

$$\sigma_T = \sqrt{\sigma_S^2 + \sigma_C^2 + \sigma_R^2} = \sqrt{6245} \approx 79.03 \text{ minutes}.$$ 

What assumption did Christine make in her calculations? Comment on the validity of this assumption.
**Intent of Question**

The primary goals of this question were to assess students’ ability to (1) calculate a probability from a normal distribution, (2) calculate a percentile value from a normal distribution, and (3) recognize that the sum of variances formula applies only to independent random variables and evaluate the validity of the independence of a set of random variables.

**Solution**

**Part (a):**

Let $S$ denote the swim time of a randomly chosen competitor where $S$ is normally distributed with a mean of 80 minutes and a standard deviation of 23 minutes.

The $z$-score for a time of one hour (60 minutes) is $z = \frac{60 - 80}{23} \approx -0.87$.

The standard normal probability table reveals that $P(S < 60) = P(Z < -0.87) \approx 0.1922$. (Calculator answer: 0.1922689816)

**Part (b):**

Let $R$ denote the run time of a randomly chosen competitor where $R$ is normally distributed with a mean of 300 minutes and a standard deviation of 46 minutes. The $z$-score corresponding to a cumulative probability of 2.5 percent is $z = -1.96$. Thus, the run time corresponding to the fastest 2.5 percent of runners can be calculated as

$$r = \mu_R + z\sigma_R = 300 - 1.96(46) = 209.84 \text{ minutes.} \quad \text{(Calculator answer: 209.8416566)}$$

**Part (c):**

In Christine's calculations, she assumed that a competitor's swimming time, cycling time, and running time are independent of one another. This is most likely not a valid assumption because, for example, it seems reasonable that knowing a competitor had a fast running time makes them more likely to have also had fast swimming and cycling times.
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 includes the following three components:
1. Indicates use of a normal distribution and clearly identifies the correct parameter values.
2. Uses the correct boundary value of 60 minutes.
3. Reports the correct normal probability consistent with components 1 and 2.

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

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

Part (b) is scored as follows:

Essentially correct (E) if the response includes the following three components:
1. Indicates use of a normal distribution and clearly identifies the correct parameter values.
2. Uses the correct percentile rank.
3. Reports the correct percentile value consistent with components 1 and 2.

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

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

Part (c) is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. Indicates that the provided calculations assume independence of the three random variables.
2. Provides a valid argument for why independence is not a reasonable assumption in this case.

Partially correct (P) if the response correctly includes only one of the components listed above.

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

Notes: The following are all sufficient for satisfying component 1 in parts (a) and (b).
- The standard notation \( N(\mu, \sigma) \) with the values of \( \mu \) and \( \sigma \) substituted.
- The \( z \)-score formula with the correct values substituted.
- A sketch of a normal curve with the center and spread labeled.
- Calculator commands with the values for the mean and standard deviation substituted into the expression AND labeled. For example, in part (b) the expression \( \text{invNorm}(0.025, \mu = 300, \sigma = 46) \) satisfies component 1, but the expression \( \text{invNorm}(0.025, 300, 46) \) could earn at most a P.
2015 Clemson AP Statistics Practice Exam – Scoring Guidelines

Question 3

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
4. A psychologist conducted a study to investigate the theory that firstborn children are smarter than their younger siblings. The psychologist randomly selected 10 families, each family consisting of just two children who both attend a school within a local school district. An IQ test was administered to each of the two siblings in the 10 families. The results are presented in the table below.

<table>
<thead>
<tr>
<th>Family</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ Score for Firstborn Sibling</td>
<td>125</td>
<td>112</td>
<td>130</td>
<td>113</td>
<td>115</td>
<td>88</td>
<td>83</td>
<td>103</td>
<td>107</td>
<td>93</td>
</tr>
<tr>
<td>IQ Score for Younger Sibling</td>
<td>123</td>
<td>99</td>
<td>125</td>
<td>99</td>
<td>106</td>
<td>82</td>
<td>85</td>
<td>96</td>
<td>107</td>
<td>94</td>
</tr>
</tbody>
</table>

Do the data provide convincing evidence that in this school district firstborn children have higher IQ scores, on average, than their younger siblings?

**Intent of Question**

The primary goal of this question was to assess students’ ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess students’ ability to (1) state appropriate hypotheses, (2) identify the appropriate statistical test procedure and check appropriate conditions for inference, (3) calculate the appropriate test statistic and *p*-value, and (4) draw an appropriate conclusion, with justification, in the context of the study.

**Solution**

Step 1: States a correct pair of hypotheses.

Let \( \mu_D \) denote the population mean difference in IQ scores (firstborn – younger) for siblings from families with two children attending school in this school district.

The hypotheses to be tested are \( H_0: \mu_D = 0 \) versus \( H_a: \mu_D > 0 \).

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

The appropriate procedure is a paired *t*-test.

The conditions for the paired *t*-test are:

1. The sample is randomly selected from the population.
2. The population of IQ score differences (firstborn – younger) is normally distributed, or the sample size is large.

The first condition is met because the ten families were randomly selected. The sample size \((n = 10)\) is not large, so we need to investigate whether it is reasonable to assume that the population of IQ score differences is normally distributed.
The computed differences are: 2 13 5 14 9 6 -2 7 0 -1

Stem-and-leaf plot of differences:

```
-0  | 21
 0   | 02
 0   | 5679
 1   | 34
```

It is reasonable to assume that the population of differences is approximately normal since the stem-and-leaf plot is roughly symmetric with no apparent outliers.

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

The test statistic is

\[
t = \frac{5.3 - 0}{5.62} \approx 2.98.
\]

The \( p \)-value, based on a \( t \)-distribution with \( 10 - 1 = 9 \) degrees of freedom, is 0.0077.

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

Because the \( p \)-value is very small (for instance, smaller than \( \alpha = 0.05 \)), we reject the null hypothesis. The data provide convincing evidence that that in this school district firstborn children have higher IQ scores, on average, than their younger siblings.
Scoring

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

Step 1 is scored as follows:

Essentially correct (E) if the response identifies the correct parameter \textit{AND} states correct hypotheses.

Partially correct (P) if the response identifies the correct parameter \textit{OR} states correct hypotheses, but not both.

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

Note: Defining the parameter symbol in context or simply using common parameter notation is sufficient.

Step 2 is scored as follows:

Essentially correct (E) if the response identifies the correct test procedure (by name or formula) \textit{AND} checks both conditions correctly.

Partially correct (P) if the response correctly completes two of the three components (identification of procedure, check of randomness condition, check of normality condition).

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

Note: A graphical check of normality is required. Graphs should be consistent with the data \textit{AND} responses must link the graph to the condition.

Dotplot of differences:

The dotplot of sample IQ score differences reveals a fairly symmetric distribution. Thus, it is reasonable to assume the population of differences is approximately normal.

Histogram of differences:

The histogram of sample IQ score differences is roughly symmetric with no apparent outliers. Thus, it is reasonable to assume the population of differences is approximately normal.
The normal probability plot shows a general linear trend with no obvious departures from linearity. Thus, it is reasonable to assume the population of differences is approximately normal.

**Step 3** is scored as follows:

Essentially correct (E) if the response correctly calculates both the test statistic and the $p$-value.

Partially correct (P) if the response correctly calculates the test statistic but not the $p$-value; 

**OR**

if the response calculates the test statistic incorrectly but then calculates the correct $p$-value for the computed test statistic.

Incorrect (I) if the response does not 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 $p$-value and the conclusion.

Partially correct (P) if the response provides a correct conclusion 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 $p$-value in step 3 is incorrect but the conclusion is consistent with the computed $p$-value, step 4 can be considered essentially correct.
- In step 4, justification based on the $p$-value can come by stating a significance level and noting that the $p$-value is smaller than the significance level OR by simply stating that the $p$-value is small. If an interpretation of the $p$-value is given, it must be correct.
- A confidence interval may be used to make the inference but must include all four parts to get full credit. The confidence level must be stated to get credit for step 3. A 95 percent confidence interval for $\mu_D$ is (1.28, 9.32).
- If an independent samples $t$-test is done, the maximum score is 3, provided all four parts for the independent samples $t$-test are done correctly. The condition of normality must be checked using two samples separately. The independent sample $t$-test results are $t = 0.80, p = 0.2169, df = 17.9$. The resulting decision is to fail to reject $H_0$. A conclusion that is equivalent to “accept $H_0$” will lose credit for step 4.

Each essentially correct (E) section counts as 1 point, and a partially correct (P) section 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, 2½ points), score down.
5. The proficiency levels on a statewide *End of Course Exam* are shown in the table below for a random sample of high school students. The results are also classified by gender.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Below Basic</th>
<th>Basic</th>
<th>Proficient</th>
<th>Advanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2757</td>
<td>6552</td>
<td>9951</td>
<td>1734</td>
<td>20994</td>
</tr>
<tr>
<td>Female</td>
<td>5131</td>
<td>6878</td>
<td>8550</td>
<td>1083</td>
<td>21642</td>
</tr>
<tr>
<td>Total</td>
<td>7888</td>
<td>13430</td>
<td>18501</td>
<td>2817</td>
<td>42636</td>
</tr>
</tbody>
</table>

(a) If a student is to be selected at random, what is the probability that the student scored basic or above on this exam?

(b) If a female student is to be selected at random, what is the probability that she was at the advanced proficiency level?

If a chi-square test of homogeneity were to be performed the hypotheses would be:

- **H₀**: The proportions in each proficiency level category are the same for both genders.
- **Hₐ**: The proficiency level category proportions are not all the same for both genders.

The computer output below gives the results from performing this test. For each cell, the observed and expected counts are reported, as well as the contribution of each cell \([(\text{observed} – \text{expected})^2/\text{expected}]\) to the chi-square statistic.

<table>
<thead>
<tr>
<th></th>
<th>Below Basic</th>
<th>Basic</th>
<th>Proficient</th>
<th>Advanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>2757</td>
<td>6552</td>
<td>9951</td>
<td>1734</td>
<td>20994</td>
</tr>
<tr>
<td>Count</td>
<td>(3884.06)</td>
<td>(6612.94)</td>
<td>(9109.91)</td>
<td>(1387.09)</td>
<td></td>
</tr>
<tr>
<td>(Expected)</td>
<td>(327.04)</td>
<td>(0.56)</td>
<td>(77.66)</td>
<td>(86.76)</td>
<td></td>
</tr>
<tr>
<td>(Contribution)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>5131</td>
<td>6878</td>
<td>8550</td>
<td>1083</td>
<td>21642</td>
</tr>
<tr>
<td>Count</td>
<td>(4003.94)</td>
<td>(6817.06)</td>
<td>(9391.09)</td>
<td>(1429.91)</td>
<td></td>
</tr>
<tr>
<td>(Expected)</td>
<td>(317.25)</td>
<td>(0.54)</td>
<td>(75.33)</td>
<td>(84.16)</td>
<td></td>
</tr>
<tr>
<td>(Contribution)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square</td>
<td>3</td>
<td>969.31168</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

(c) Given the results of the test, is there statistically convincing evidence that the proficiency level proportions are not the same across genders? Justify your decision.

(d) The state's Department of Education would like to determine which one group needs more tutoring assistance. Which group would you recommend? Use the contributions to chi-square results shown in the table above to justify your choice.
2015 Clemson AP Statistics Practice Exam – Scoring Guidelines
Question 5

**Intent of Question**

The primary goals of this question were to assess students’ ability to (1) calculate appropriate probabilities, including conditional probabilities, from a two-way table; (2) determine the results of a statistical test using computer output; (3) interpret results from a statistical analysis in the context of a question.

**Solution**

**Part (a):**

Using the addition rule, the probability that the randomly selected student scored basic or above on this end-of-course test is:

\[
P(\text{basic or proficient or advanced}) = P(\text{basic}) + P(\text{proficient}) + P(\text{advanced})
\]

\[
= \frac{13430}{42636} + \frac{18501}{42636} + \frac{2817}{42636} = \frac{34748}{42636} = 0.8150
\]

Or, using the complement rule, \(1 - P(\text{below basic}) = 1 - \frac{7888}{42636} = 0.8150\)

**Part (b):**

Reading values from the table, the conditional probability that the selected student was at the advanced proficiency level given she was female is \(\frac{1083}{21642} = 0.0500\)

**Part (c):**

The \(p\)-value of <0.0001 is much smaller than the conventional significance levels such as \(\alpha = 0.10\) or \(\alpha = 0.05\) or \(\alpha = 0.01\). Therefore, the \(p\)-value indicates that the sample data do provide strong enough evidence to conclude that the proficiency levels are not at the same proportions across genders.

**Part (d):**

The contributions to the chi-square test statistic value are the largest for both genders in the below basic level. This shows a significant difference between the observed and expected values if the null hypothesis was true. However, the male group data shows that fewer males than expected were at the below basic level while the female group data revealed that more females than expected were at the below basic level. Therefore, I would recommend that the female students in the below basic group receive the extra tutoring assistance.
2015 Clemson AP Statistics Practice Exam – Scoring Guidelines
Question 5

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the probability is computed correctly and appropriate work is shown

OR

if the probability calculation is set up correctly but a minor computational error is made.

Partially correct (P) if the probability of scoring at only the basic proficiency level is calculated,

\[
\frac{13430}{42636} = 0.3150
\]

Incorrect (I) if the response does not meet the criteria for an E or P, or includes the correct decimal answer with no accompanying work or justification.

Part (b) is scored as follows:

Essentially correct (E) if the probability is correctly computed and appropriate work is shown

OR

if the calculation is set up correctly but a minor computational error is made.

Partially correct (P) if the reverse conditional probability of being a female given that the student was at the advanced proficiency level, resulting in

\[
\frac{1083}{2817} = 0.3845
\]

Incorrect (I) if the response otherwise fails to meet the requirements for an E or P.

Part (c) is scored as follows:

Essentially correct (E) if the response states a correct conclusion in the context of the study AND provides correct justification of that conclusion based on linkage to the \( p \)-value.

Partially correct (P) if the response provides no conclusion in context but does provide correct justification based on linkage to the \( p \)-value

OR

if the response provides a correct conclusion in context but with incorrect or missing justification based on linkage to the \( p \)-value.

Incorrect (I) if the response otherwise fails to meet the requirements for an E or P.

Note: Justification based on the \( p \)-value can come by stating a significance level and noting that the \( p \)-value is smaller than the significance level OR by simply stating that the \( p \)-value is small.
Part (d) is scored as follows:

Essentially correct (E) is the female below basic group is selected as the group requiring additional tutoring assistance based on the contribution to the chi-square test statistic.

Partially correct (P) if the male below basic group is selected as the group requiring additional tutoring based on it having the largest contribution to the chi-square test statistic OR if the female below basic group is selected without referring to its large contribution to the chi-square test statistic.

Incorrect (I) if the response otherwise fails to meet the requirements for an E or P.

Each essentially correct (E) section counts as 1 point, and a 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), score down.
6. Each flu season, medical researchers estimate the effectiveness of the flu vaccine that was administered that season. At the end of the most recent flu season, 2321 adults were randomly selected to participate in the U.S. Influenza Vaccine Effectiveness Study. Each participant was classified by whether or not they chose to receive the flu vaccine and whether or not they were diagnosed with the flu that flu season. The results are presented in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Flu</th>
<th>No Flu</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Vaccinated</td>
<td>239</td>
<td>846</td>
<td>1085</td>
</tr>
<tr>
<td>Vaccinated</td>
<td>124</td>
<td>1112</td>
<td>1236</td>
</tr>
<tr>
<td>Total</td>
<td>363</td>
<td>1958</td>
<td>2321</td>
</tr>
</tbody>
</table>

(a) Is this study an experiment or an observational study? Explain your answer and discuss the implications this has for establishing a causal relationship between receiving the flu vaccine and contracting the flu.

(b) The conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the difference between the proportion of adults not receiving the vaccine who contract the flu and the proportion of adults receiving the vaccine who contract the flu this particular flu season.

In many of these types of studies, researchers are interested in the ratio of the odds of contracting the flu for those not receiving the vaccine and those receiving the vaccine. This ratio is usually referred to as an odds ratio (OR), and is given by

\[ OR = \frac{p_{NV}/(1 - p_{NV})}{p_{V}/(1 - p_{V})} \]

where \( p_{NV} \) represents the proportion of adults not receiving the vaccine who contract the flu and \( p_{V} \) represents the proportion of adults receiving the vaccine who contract the flu.

For example, an odds ratio of 1 indicates that the odds of contracting the flu are the same for adults who do not and who do receive the flu vaccine. Whereas, an odds ratio of 1.5 indicates that the odds of contracting the flu for adults not receiving the flu vaccine are 1.5 times the odds for adults receiving the vaccine. An estimator of the odds ratio is the sample odds ratio

\[ \hat{OR} = \frac{\hat{p}_{NV}/(1 - \hat{p}_{NV})}{\hat{p}_{V}/(1 - \hat{p}_{V})}. \]

(c) Using the data from the U.S. Influenza Vaccine Effectiveness Study presented above, compute the estimate of the odds ratio.
The sampling distribution of $\hat{OR}$ is skewed. However, when both sample sizes $n_{NV}$ and $n_V$ are relatively large, the distribution of $\ln(\hat{OR})$, the natural logarithm of the sample odds ratio, is approximately normal with a mean of $\ln(OR)$ and an estimated standard error of

$$\sqrt{\frac{1}{n_{NV,F}} + \frac{1}{n_{NV,NF}} + \frac{1}{n_{V,F}} + \frac{1}{n_{V,NF}}}$$

where $n_{NV,F}$ represents the number of adults in the sample that were not vaccinated and did contract the flu, $n_{NV,NF}$ represents the number of adults in the sample that were not vaccinated and did not contract the flu, $n_{V,F}$ represents the number of adults in the sample there were vaccinated and did contract the flu, and $n_{V,NF}$ represents the number of adults in the sample that were vaccinated and did not contract the flu.

When a 95 percent confidence interval for $\ln(OR)$ is known, an approximate 95 percent confidence interval for $OR$ can be constructed by exponentiating (applying the inverse of the natural logarithm to) the endpoints of the confidence interval for $\ln(OR)$.

(d) The conditions for inference are met, and a 95 percent confidence interval for $\ln(OR)$ based on the data from the study presented above is $(0.69495, 1.16421)$. Construct and interpret a 95 percent confidence interval for the odds ratio of contracting the flu for those not receiving the vaccine to those receiving the vaccine.

(e) What is an advantage of using the interval in part (d) over using the interval in part (b)?
**Intent of Question**

The primary goals of this investigative task were to assess students’ ability to (1) evaluate whether a study is an observational study or an experiment and explain how this affects the generalizability of the results, (2) construct and interpret a confidence interval for the difference between the two proportions, (3) estimate an odds ratio and construct and interpret a confidence interval for an odds ratio, and (4) compare the confidence interval for the difference between two proportions and the confidence interval for the odds ratio.

**Solution**

**Part (a):**

This was an observational study because the researchers did not impose a treatment on the participants and the participants were not randomly assigned to two groups. Rather, the participants chose whether or not they would receive a flu vaccine and the researchers retrospectively observed whether or not the participants subsequently contracted the flu. As such, this study cannot establish a causal relationship between receiving a flu vaccine and contracting the flu.

**Part (b):**

Step 1: Identify the appropriate confidence interval by name or formula and check appropriate conditions. (The question states that the conditions for inference have been met.)

Two-sample $z$ confidence interval for the difference of two proportions

OR

The formula for the confidence interval given in step 2 is provided

Step 2: Correct mechanics.

\[
\hat{p}_{NV} = \frac{239}{1085} \approx 0.22 \quad \text{and} \quad \hat{p}_V = \frac{124}{1236} \approx 0.10
\]

Then, the 95 percent confidence interval for \((p_{NV} - p_V)\) is

\[
(\hat{p}_{NV} - \hat{p}_V) \pm z^* \sqrt{\frac{\hat{p}_{NV}(1 - \hat{p}_{NV})}{n_{NV}} + \frac{\hat{p}_V(1 - \hat{p}_V)}{n_V}}
\]

\[
= \left(\frac{239}{1085} - \frac{124}{1236}\right) \pm 1.96 \sqrt{\frac{239}{1085} \left(1 - \frac{239}{1085}\right) + \frac{124}{1236} \left(1 - \frac{124}{1236}\right)}
\]

\[
= 0.11995 \pm 1.96(0.01521)
\]

\[
= 0.11995 \pm 0.02981
\]

\[
= (0.0901, 0.1498)
\]
Step 3: Interpretation.

We can be 95 percent confident that the true difference in the proportion of unvaccinated and vaccinated persons contracting the flu is between 0.0901 and 0.1498. Because this interval is entirely above zero, this suggests that the proportion contracting the flu is higher among unvaccinated persons.

Part (c):

The estimate of the odds ratio is

$$\hat{OR} = \frac{\hat{p}_{NV}(1 - \hat{p}_{NV})}{\hat{p}_{V}(1 - \hat{p}_{V})} = \frac{\frac{239}{1085}/\left(1 - \frac{239}{1085}\right)}{\frac{124}{1236}/\left(1 - \frac{124}{1236}\right)} \approx 2.5334$$

Part (d):

A 95 percent confidence interval for the odds ratio is found by evaluating

$$e^{0.69495}$$ to $$e^{1.16421}$$, which is 2.004 to 3.203.

We can be 95 percent confident that the odds ratio is between 2.004 and 3.203. For this flu season, the odds of contracting the flu for unvaccinated persons are between 2.004 and 3.203 times the odds of contracting the flu for vaccinated persons.

Part (e):

When the proportions of people contracting the disease are low, as is the case with 0.22 and 0.10, it may be more meaningful or impactful to know that the odds of a person contracting the disease without the vaccine are 2 to 3.2 times the odds of contracting the disease with the vaccine, rather than to know that the difference in the proportions of people who contract the disease is 0.09 to 0.15, which does not sound like very much.
**2015 Clemson AP Statistics Practice Exam – Scoring Guidelines**

**Question 6**

**Scoring**

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

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

Essentially correct (E) if the response includes the following three components:
1. The study has been identified as an observational study.
2. The justification is tied to the fact that the researchers did not *impose* a treatment OR states that there was no random assignment of subjects to treatments because the participants chose whether or not to receive the flu vaccine.
3. The discussion of the causal relationship clearly demonstrates an understanding that an observational study cannot establish a cause-and-effect relationship.

Partially correct (P) if the response correctly provides only two of the three components.

Incorrect (I) if the response correctly provides only one or none of the components.

**Section 2** [part (b)] is scored as follows:

Essentially correct (E) if the correct confidence interval is identified and constructed *AND* interpreted in context.

Partially correct (P) if the correct confidence interval is identified and constructed, but the interpretation is not in context or no interpretation is given OR there are calculation errors with the appropriate confidence interval, but the interpretation follows correctly from the interval and is in context.

Incorrect (I) if the confidence interval and interpretation are not reasonable.

**Section 3** [parts (c) and (d)] is scored as follows:

Essentially correct (E) if the response includes the following three components:
1. The estimated odds ratio has been correctly calculated.
2. The confidence interval for the odds ratio has been correctly calculated.
3. The confidence interval has been interpreted in context.

Partially correct (P) if the response correctly provides only two of the three components.

Incorrect (I) otherwise.
2015 Clemson AP Statistics Practice Exam – Scoring Guidelines

Question 6

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

   Essentially correct (E) if the response includes the following two components:
      1. An advantage of using the confidence interval for the odds ratio has been provided.
      2. Both intervals have been addressed.

   Partially correct (P) if the first component is correct.

   Incorrect (I) otherwise.

Each essentially correct (E) section counts as 1 point, and a partially correct (P) section 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), score down.