Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

INTRODUCTION

Suppose that you are interested in investigating whether there are differences between men and women who text while driving. After studying these populations of men and women, you begin to develop the opinion that men send more texts while driving than women.

To investigate this possibility, you randomly sample the populations of men and women who text while driving. The data you gather include 85 randomly selected men who text while driving, and among these are 31 who text frequently. Also, there are 97 women who text while driving, including 22 who text frequently.

The following questions will guide you through the process of testing hypotheses regarding the difference between two population proportions.

1. Begin by verifying that the criteria for approximate normality are met. Verify that each sample contains at least 10 successes and 10 failures. In this context a success is a person who texts frequently while driving.

   For the men, there are _____ successes and _____ failures.

   For the women, there are _____ successes and _____ failures.

   Are these large enough to assume approximate normality of the underlying sampling distributions?

2. While we assume that the population proportions are equal (with zero difference – this will be our null hypothesis), we would like to test the claim that the population proportion of men who text frequently while driving (7 or more texts per week) is larger than the population proportion of women (this is the alternative hypothesis).
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

A Write a sentence that describes the null hypothesis in the context of the difference of the population proportions.

B Write a sentence that describes the alternative hypothesis in the context of the difference of the population proportions.

3 The test statistic for this hypothesis test is the \( Z \)-score of the observed difference, \( p_m - p_w \).

A Compute the sample proportions for the men, \( p_m \), and for the women, \( p_w \).

\[
p_m = \text{______}
\]

\[
p_w = \text{______}
\]

B Compute the difference between these sample proportions, rounding to two places after the decimal.

\[
p_m - p_w =
\]

C The null hypothesis for a two proportions test generally claims that the population proportions are equal. Under this assumption, this hypothetical common proportion is estimated by combining the samples to create a single proportion, \( \bar{p} \). Compute this below.

\[
\bar{p} = \frac{x_m + x_w}{n_m + n_w} = \text{______}
\]

We now estimate the standard error of differences as we have done before, but here we use the pooled proportion as the estimate for the proportion common to both populations. This computation is performed below.

\[
s_{p_m-p_w} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_m} + \frac{\bar{p}(1-\bar{p})}{n_w}} = \sqrt{\frac{0.291(1-0.291)}{85} + \frac{0.291(1-0.291)}{97}} \approx 0.07
\]

Under the assumption of equal population proportions, a collection of differences, \( p_m - p_w \), is simulated in the figure below.
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

Each data point on the dotplot represents the difference of two proportions. The first is a sample proportion of men who text frequently while driving from a sample of size 85. The second is a sample proportion of women who text frequently while driving from a sample of size 97.

Note that the assumption of equal population proportions yields a distribution of differences whose mean appears to be zero.

To learn more precisely the number of standard errors that the difference, \( p_m - p_w \), lies from the mean of differences, \( \pi_m - \pi_w = 0 \), we compute a z-score. As always, the Z-score is computed by subtracting the mean, and dividing by the estimated standard error. The resulting Z-score is the test statistic for the two proportions test:

\[
Z = \frac{(p_m - p_w) - (\pi_m - \pi_w)}{S_{p_m-p_w}}
\]

D Add the difference computed in Question 3.b to the dotplot above. Is the difference observed in your sample data in or near the range of values that you consider unusual?
E Using the assumed mean difference, \( \pi_m - \pi_w = 0 \), and the estimated standard error,

\[
s_{p_m - p_w} \approx 0.07,
\]

compute the Z-score of the difference computed in A above.

\[
Z = \frac{(p_m - p_w) - (\pi_m - \pi_w)}{s_{p_m - p_w}} = 
\]

F Does this Z-score appear to be unusually high or low?

4 To determine whether this Z-score indicates a significant deviation from the assumed mean difference of zero, we use a P-value.

As we have seen previously, the P-value is the probability of observing a Z-score which is at least as extreme as the test statistic (under the assumption of a null hypothesis). The figure below graphically depicts the regions defined by the test statistic, \( Z = 2.00 \), under the standard normal (Z) distribution.

![Z-score distribution](image)

\( P\)-value = _______

B Suppose that we require the P-value be less than or equal to 5% before we consider it to be significant (for \( \alpha = 0.05 \), a 5% level of significance). Is the P-value small enough to consider the difference, \( p_m - p_w \), (from 3.b) significantly different from zero?
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

C You may recall from the one proportion hypothesis test that the null hypothesis is rejected whenever the \( P \)-value is less than the level of significance. Do you reject the null hypothesis here?

D Does this deviation above zero of the given difference, \( p_m - p_w \), lead you to believe that the difference between population proportions might be greater than zero as well?

E Suppose that the difference between the population proportions of men and women who text frequently while driving is greater than zero. Is the proportion higher for the men or the women?

5 Write a conclusion regarding the differences between the proportions of men and women who text frequently while driving. Remember that your decision was based on sample data only. It is, therefore, subject to the possibility of error. Your wording should carefully balance the likelihood of a correct decision with the uncertainty which accompanies decisions made using data from samples.
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

TAKE IT HOME

Worldwide there appear to be differences between the proportions of girls and boys who are permitted by their circumstances to attend elementary school. The following values are simulated based on parameters estimated in 2008 by Unicef.

Suppose that, from a sample of 825 girls (of school age or older), randomly sampled from all over the world, 693 were permitted to attend elementary school. In a sample of 861 boys (of school age or older), randomly sampled worldwide, 750 were permitted to attend elementary school.

We would like to investigate the claim that the population proportion of girls who are permitted to go to school is lower than the population proportion for boys.

For this test, we will use a level of significance of 5%.

1. We begin by verifying that the assumptions of approximate normality for the test are satisfied. Are the criteria for approximate normality met?

2. Write the null and alternative hypotheses as sentences below.

   A. The null hypothesis assumes that the population proportions are equal – that their difference is zero. Write a sentence which describes the null hypothesis regarding the population proportions of girls and boys who are permitted to attend school.

   B. Write a sentence which describes the alternative hypothesis regarding the population proportions of girls and boys who are permitted to attend school.
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

3 The test statistic for this hypothesis test is the Z-score of the observed difference, \( p_g - p_b \). Computing it is our next goal.

A Compute the sample proportions for the girls, \( p_g \), and for the boys, \( p_b \).

\[ p_g = \]
\[ p_b = \]

B Compute the difference between the sample proportions.

\[ p_g - p_b = \]

C Next, compute the pooled proportion, \( \bar{p} \).

\[ \bar{p} = \frac{x_g + x_b}{n_g + n_b} = \]

With the pooled proportion, we estimate the standard error of differences

\[ s_{p_g - p_b} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_g} + \frac{\bar{p}(1 - \bar{p})}{n_b}} = \sqrt{\frac{0.856(1 - 0.856)}{825} + \frac{0.856(1 - 0.856)}{861}} \]
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

Under the assumption of equal proportions, a collection of differences, $p_g - p_b$, is simulated in the figure below. Note that the assumption of equal proportions yields a distribution of differences whose mean is zero.

D  Does the distribution of differences appear to be approximately normal?

E  Using the assumed mean difference of $\pi_g - \pi_b = 0$, and the estimated standard error,

$$s_{p_g - p_b} \approx 0.0171,$$

compute the Z-score of the difference computed in $b$ above.

$$Z = \frac{(p_m - p_w) - (\pi_m - \pi_w)}{s_{p_g - p_b}} =$$

F  Does the test statistic seem to support the claim that the difference in the proportions of girls and boys who are permitted to attend school is less than zero?

G  Does the test statistic seem to support the claim that the proportion of girls permitted to attend school is less than the proportion for boys?
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

H Does this agree with your choice for the alternative hypothesis in Question 2.b?

4 To determine whether this Z-score indicates a significant deviation from the assumed mean difference of zero, we use a $P$-value. For a left-tailed test, the $P$-value is the probability of observing a Z-score that is less than the test statistic.

The figure below graphically depicts the region under the normal distribution associated with this $P$-value. Your test statistic should be around $Z = -1.81$, depending on how you rounded your numbers.

A What is the $P$-value associated with the test statistic, $Z = -1.81$, on this left-tailed test?

B Is the $P$-value less than the level of significance ($\alpha = 0.05$)?

C Does the $P$-value indicate that the difference between the observed difference and our assumed difference of zero is statistically significant?
Lesson 9.3.1
A Statistical Test for the Difference in Two Population Proportions

5 Use plain language and complete sentences to write a conclusion regarding the difference in proportions of girls and boys who are permitted to go to school.

+++++

This lesson is part of STATWAY™, A Pathway Through College Statistics, which is a product of a Carnegie Networked Improvement Community that seeks to advance student success. Version 1.0, A Pathway Through Statistics, Statway™ was created by the Charles A. Dana Center at the University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This version 1.5 and all subsequent versions, result from the continuous improvement efforts of the Carnegie Networked Improvement Community. The network brings together community college faculty and staff, designers, researchers and developers. It is an open-resource research and development community that seeks to harvest the wisdom of its diverse participants in systematic and disciplined inquiries to improve developmental mathematics instruction. For more information on the Statway Networked Improvement Community, please visit carnegiefoundation.org. For the most recent version of instructional materials, visit Statway.org/kernel.

+++++

STATWAY™ and the Carnegie Foundation logo are trademarks of the Carnegie Foundation for the Advancement of Teaching. A Pathway Through College Statistics may be used as provided in the CC BY license, but neither the Statway trademark nor the Carnegie Foundation logo may be used without the prior written consent of the Carnegie Foundation.