

More Proportions Practice Problems

MAT241 Class

Here are some practice problems from our Textbook and MyOpenMath.

Introduction to Inference (MoM Week 5 HW)

Problem 1: For each of the following situations, state whether the parameter of interest is a mean or a proportion. It may be helpful to examine whether individual responses are numerical or categorical.

- In a survey, one hundred college students are asked how many hours per week they spend on the Internet.
 - Participants are being asked: “How many hours per week do you spend on the internet?”
 - Sample Answer: 5 hours, 6 and a half hours...numerical values.
 - So this study is investigating a *population mean* (the average number of hours per week on the internet).
- In a survey, one hundred college students are asked: “What percentage of the time you spend on the Internet is part of your course work?”
 - Participants are asked: “What percentage of the time you spend on the Internet is part of your course work?”
 - Sample Answer: 75%, 43%
 - Investigating *average* percentage of time spent.
- In a survey, one hundred college students are asked whether or not they cited information from Wikipedia in their papers.
 - Participants asked: “Have you cited Wikipedia on a paper?”
 - Sample answer: Yes, No
 - This is a study on *proportions*

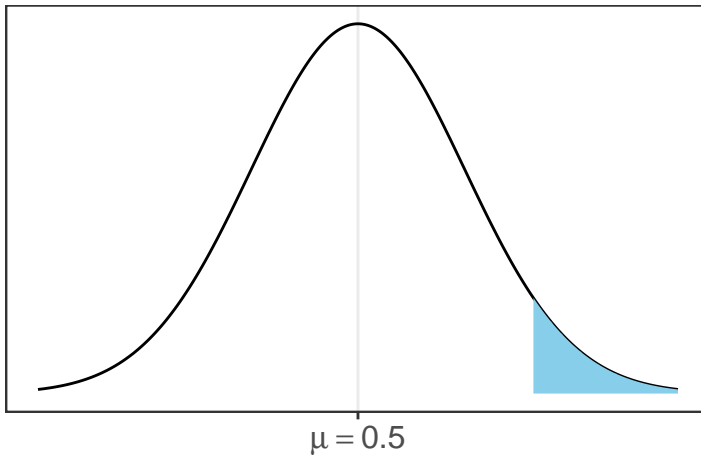
- In a survey, one hundred college students are asked what percentage of their total weekly spending is on alcoholic beverages.
 - Participants asked: “What percentage of your weekly spending is on alcohol?”
 - Sample Answer: 12%, 5%, numeric percentages
 - This is a study on *means* (the average percentage of total weekly spending)
- In a sample of one hundred recent college graduates, it is found that 85 percent expect to get a job within one year of their graduation date.
 - Participants asked: “Do you expect to find a job within one year of graduating?”
 - Sample Answer: yes/no
 - This is a study on *proportions* (the proportion of recent college graduates who expect to find a job within one year of graduating).

Problem 2: A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter, and the standard error for this estimate was 2.4%. Conduct a hypothesis test at the $\alpha = 0.05$ level of significance to determine whether a majority of US adult Twitter users get at least some news on Twitter.

Solution.

- **Hypotheses:** $H_0 : \mu = 0.5$
 $H_a : \mu > 0.5$
- **Picture of Alternative Hypothesis:**

Shaded Region is Samples Favorable to H_a



- **Set α Level:** Notice that $\alpha = 0.05$ in the problem statement.
- **Compute the Test Statistic:**

```
#Here is a code cell
null_value <- 0.5
point_estimate <- 0.52
st_error <- 0.024

test_stat <- (point_estimate - null_value)/st_error

test_stat
```

```
[1] 0.8333333
```

- **Compute the p -value:** Notice that our p -value is the area to the right of our test statistic. We use the picture of the alternative hypothesis to tell this.

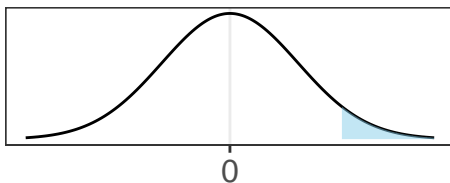
```
p_val <- 1 - pnorm(test_stat, 0, 1)

p_val
```

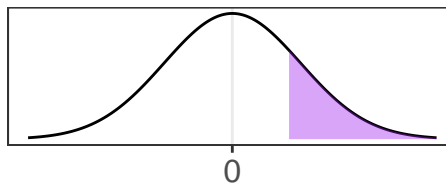
```
[1] 0.2023284
```

Comparison of α and p -value: Our p -value Exceeds α

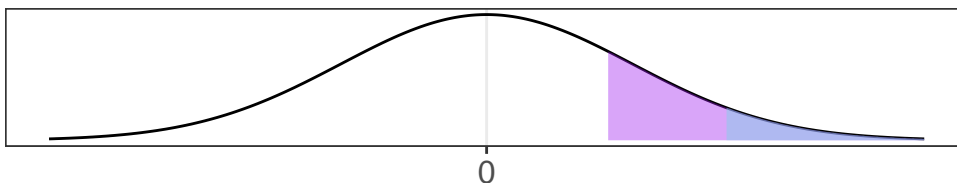
$\alpha = 0.05$



$p\text{-value} \approx 0.2023$



p -value (purple), α (blue)



- **Compare the p -value to the Level of Significance:** Since our p -value exceeds the level of significance, our sample is not one of those “unlikely” samples which are favorable to the alternative hypothesis. That is, our sample is *compatible* with a world in which the null hypothesis is true – we do **not**

have significant evidence to suggest that more than 50% of US Adult Twitter users get some news from Twitter.

Problem 5: A store randomly samples 603 shoppers over the course of a year and finds that 142 of them made their visit because of a coupon they'd received in the mail. Construct a 95% confidence interval for the fraction of all shoppers during the year whose visit was because of a coupon they'd received in the mail.

Problem 6: A tutoring company would like to understand if most students tend to improve their grades (or not) after they use the company's services. They sample 200 of the students who used their service in the past year and ask them if their grades have improved or declined from the previous year. Of the 200 sampled, 185 said that their grades had improved. Determine whether the data provides evidence to suggest that the companies tutoring services suggest that over 90% of customers report improved grades after using the tutoring services. Use the $\alpha = 0.1$ level of significance.

Problem 11: A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter (Pew, 2013). The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion. Construct a 99% confidence interval for the fraction of U.S. adult Twitter users who get some news on Twitter, and interpret the confidence interval in context.

Inference on One and Two Proportions (MoM Week 6 HW)

Problem 3: Among a simple random sample of 331 American adults who do not have a four-year college degree and are not currently enrolled in school, 48% said they decided not to go to college because they could not afford school. Calculate a 90% confidence interval for the proportion of Americans who decide to not go to college because they cannot afford it, and interpret the interval in context.

Solution. In this problem, we are asked to construct a confidence interval, with 90% confidence. We'll follow the steps to do so below.

We are trying to capture a single population proportion here. The formula for a confidence interval is (point estimate) \pm (critical value) (S_E).

- **Point Estimate:** The point estimate is the sample proportion. Here, that is 48%.
- **Standard Error (S_E):** Using the standard error decision tree, we find that $S_E = \sqrt{\frac{p(1-p)}{n}}$. Notice also that there is no information about *degrees of freedom (df)* in this box, so we are free to use the normal distribution with this problem.

- **Critical Value:** Since we are working with a normal distribution, we can use the small table at the top of the standard error decision tree to determine that the critical value is $z_{\alpha/2} = 1.65$.

Since we have all of the components of the confidence interval formula, we are now ready to construct our interval.

```
point_estimate <- 0.48
st_error <- sqrt(0.48*(1 - 0.48)/331)
critical_value <- 1.65

lower <- point_estimate - (critical_value*st_error)
upper <- point_estimate + (critical_value*st_error)

c(lower, upper)
```

```
[1] 0.4346902 0.5253098
```

Interpretation: Given the results above, we are 90% confident that, of American Adults who chose not to go to college, the proportion making that decision because they could not afford to is between 43.47% and 52.53%. Note that we **cannot** claim that a majority or minority of American Adults made this decision for this reason since our confidence interval contains proportions both below and above 50%.

Problem 4: A 2010 Pew Research foundation poll indicates that among 1,099 college graduates, 33% watch The Daily Show. Meanwhile, 22% of the 1,110 people with a high school degree but no college degree in the poll watch The Daily Show. Construct a 95% confidence interval for $(p_{\text{college grad}} - p_{\text{HS or less}})$, where p is the proportion of those who watch The Daily Show

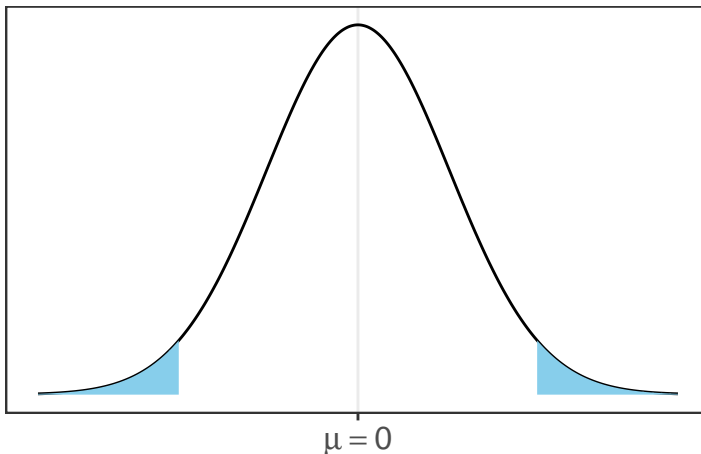
Problem 5: Researchers studying the link between prenatal vitamin use and autism surveyed the mothers of a random sample of children aged 24 - 60 months with autism and conducted another separate random sample for children with typical development. The table below shows the number of mothers in each group who did and did not use prenatal vitamins during the three months before pregnancy (periconceptional period). Conduct a hypotheses to test for independence of use of prenatal vitamins during the three months before pregnancy and autism. (Schmidt, 2011)

	Autism	Typical Development	Total
No vitamin	111	70	181
Vitamin	143	159	302
Total	254	229	483

Solution. Note that we are conducting a hypothesis test here for a difference of two population *proportions*. The first population is *mothers and children where the mother took a prenatal vitamin during pregnancy*, and the second population is *mothers and children who did not take a prenatal vitamin during pregnancy*.

- **Hypotheses:** The hypotheses are $H_0 : p_{\text{vitamin}} = p_{\text{no vitamin}}$. While this framing of the hypotheses is fine, it hides (i) the null value and (ii) the population parameter from us. It is generally better to rewrite these hypotheses so that we have “zero on one side”. That is, the hypotheses are
 $H_0 : p_{\text{vitamin}} - p_{\text{no vitamin}} = 0$
 $H_a : p_{\text{vitamin}} - p_{\text{no vitamin}} \neq 0$
- **Picture of Alternative Hypothesis:**

Shaded Region is Samples Favorable to H_a



- **Set α Level:** Notice that we'll assume $\alpha = 0.05$ since we aren't told otherwise. Note that this α is split up across the two tails, so each shaded area in the image above represents 2.5% of samples.
- **Compute the Test Statistic:**

```

null_value <- 0
p_vitamin <- 143/302
p_no_vitamin <- 111/181
point_estimate <- p_vitamin - p_no_vitamin

#Compute Standard Error
term1 <- p_vitamin*(1 - p_vitamin)/302
term2 <- p_no_vitamin*(1 - p_no_vitamin)/181
st_error <- sqrt(term1 + term2)

```

```
test_stat <- (point_estimate - null_value)/st_error
```

```
test_stat
```

```
[1] -3.023898
```

- **Compute the p -value:** Notice that our p -value needs to account for samples whose difference in proportions is less than 0 and also samples whose difference in proportions is greater than 0. Our p -value will need to include two areas, according to our picture of the alternative hypotheses! Because of this, we'll calculate the area in one tail of the distribution and then double it.

```
tail_area <- pnorm(test_stat, 0, 1)
```

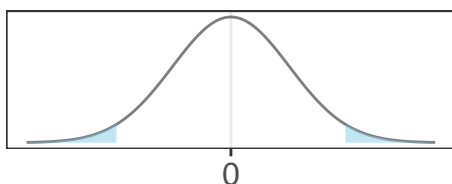
```
p_val <- 2*tail_area
```

```
p_val
```

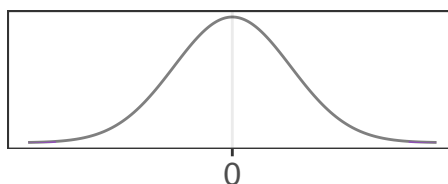
```
[1] 0.002495405
```

Comparison of α and p -value: Our p -value Is Less Than α

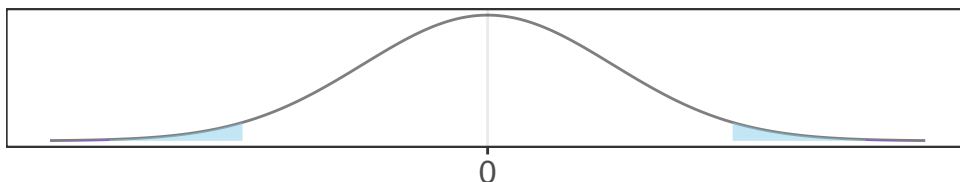
$\alpha = 0.05$



$p\text{-value} \approx 0.0025$



p -value (purple), α (blue)



- **Compare the p -value to the Level of Significance:** Since our p -value is less than the level of significance, our sample is one of those “unlikely” samples which are favorable to the alternative hypothesis. That is, our sample is *incompatible* with a world in which the null hypothesis is true – we **do** have significant evidence to suggest that there is an association between prenatal vitamin use and the development of autism in children.

Note. We have not tested whether the use of the particular prenatal vitamin increases or decreases the risk of development with autism here.