# ONE PROPORTION Z-INTERVAL

During a flu vaccine shortage in the United States, it was believed that 45 percent of vaccine-eligible people received flu vaccine. The results of a survey given to a random sample of 2,350 vaccine-eligible people indicated that 978 of the 2,350 people had received flu vaccine.

Construct and interpret a 99 percent confidence interval for the proportion of vaccine-eligible people who had received flu vaccine.

**Know the conditions for a 1-proportion z interval, specifically how to check normality, (11b, 5b; 10b, 4a; 03, 6b; 03b, 6a; M7, 21)**

We are going to do a 1 proportion z-interval. In order to proceed we must verify that data was taken at random, and that a normal approximation is appropriate.

Random → "A random sample of 2350 vaccine-eligible people"

Normal → $n\hat{p} > 10$ $n(1-\hat{p}) > 10$

$2350(0.4162) > 10$ $2350(0.5838) > 10$

Independent → 2350 is less than 10% of vaccine eligible people

**Know how to calculate a one proportion z interval, (11b, 5a; 03, 6b; 03b, 6a)**

CI = Statistic ± (Critical Value)(Standard Deviation)

CI = $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.4162 \pm 2.576 \sqrt{\frac{(0.4162)(0.5838)}{2350}}$

CI = (0.38998, 0.44236)

**Know how to interpret a one proportion z interval, (11b, 5b; 03, 6b; 03b, 6a; M7, 34)**

We are 99% confident that the proportion of vaccine-eligible people who actually received the flu vaccine is between 0.38998 and 0.44236.

THE THREE STEPS ABOVE MUST BE DONE FOR EVERY CONFIDENCE INTERVAL

# More on Confidence Intervals Part 1

**Know how to interpret a confidence level (i.e. What does it mean to be 95% confident)? (10, 3a; 03b, 6a; M2, 37)**

Using the restaurant situation we would say, "If many samples were taken from this community and many confidence intervals were created, 99% of those intervals would contain the true proportion of vaccine-elgible people who have been vaccinated with the flu vaccine." **(Note: This interpretation applies to all confidence levels)**

**Know how changing the sample size affects the margin of error, (M7, 30)**

The more samples that you collect, the smaller your margin of error is. This makes sense, because if you take more observations, you should be getting a better estimate of the true average. **(Note: This applies to all confidence intervals)**

**Know what happens to a confidence interval and the margin of error if you increase the confidence level, (M2, 13)**

The more confident you are, the wider your confidence interval will become. In order to be extremely confident, you have to have a large margin of error. I am 100% confident that you will get a between a 1 and 5 on the AP exam. I am only 60% confident that you will get between a 3 and a 4. **(Note: This applies to all confidence intervals)**

**What exactly is a confidence interval? How do I use it to argue for or against a claim (e.g. Based on your confidence interval is there evidence that boys are smarter than girls?)? (11b, 5a; 10, 3b;10b, 6d; 03, 6b)**

Use your confidence interval above to comment on the belief that 45 percent of the vaccine-eligible people had received flu vaccine.

We believe that the proportion is between 0.38998 and 0.44236. Since 0.45 is not in our confidence interval, we don't believe that 45% of vaccine-eligible people had received the flu vaccine.

**If I need to have a certain margin of error, how do I establish how many observations to make? (11b, 5b; 10, 3c; 08b, 3a; 05, 5b; 03b, 6b; M2, 26)—635, 671**

Suppose a similar survey will be given to vaccine-eligible people in Canada by Canadian health officials. A 99 percent confidence interval for the proportion of people who will have received flu vaccine is to be constructed. What is the smallest sample size that can be used to guarantee that the margin of error will be less than or equal to 0.02?

$$m = z^* \sqrt{\frac{p(1-p)}{n}}$$

Since there is no way that we can know what p will be (that's what we are trying to find), we will use 0.5, because it is the most conservative guess. It will give us a number that GUARANTEES that we won't go over the margin of error that we want.

$$0.02 = 2.576\sqrt{\frac{(0.5)(0.5)}{n}} \quad \rightarrow \quad n = 4147.36$$

We will need to sample 4148. Always round up, because 4147 won't be enough to guarantee.

**If we are taking a sample of 50 from a population, why don't we consider the fact that the probability is changing because we are not replacing? (10b, 4b)**

Let's say that we have 10,000 songs on a computer, and we are going to choose 500 unique songs at random to create a playlist. Some of the songs on the computer were downloaded by Reid, and some were downloaded by Derrel. If 400 of our 500 songs (0.8) were downloaded by Reid, create a 99% confidence interval for the proportion of songs on the computer that were downloaded by Reid. We understand that once you remove a song from the list the probability of choosing a song downloaded by Reid will not be the same as it was when you started. Why is it appropriate to do a confidence interval even though the probability isn't constant?

This is the 10% rule (Rule of thumb). Since we are only taking a small sample from the population (500 out of 10000 → 5%) the probability won't really change that much from song to song. So, it is appropriate to proceed because, even though the probability is changing, it's not changing drastically. It would not be appropriate to proceed if our sample was larger than 10% of the population. We couldn't create a confidence interval if we wanted to choose more than 1000 (10%) songs to add to our playlist.

# 2 PROPORTION Z-INTERVAL

Researchers want to determine whether drivers are significantly more distracted while driving when using a cell phone than when talking to a passenger in the car. In a study involving 288 people, 144 people were randomly assigned to drive in a driving simulator while using a cell phone. The remaining 144 were assigned to drive in the driving simulator while talking to a passenger in the simulator. Part of the driving simulation for both groups involved asking drivers to exit the freeway at a particular exit. In the study, 42 of the 144 cell phone users missed the exit, while 12 of the 144 talking to a passenger missed the exit.

Create and interpret a 90% confidence interval for the difference in the proportion of those who missed the exit while using cell phones and those who were talking to a passenger.

**Know how to check the conditions for a 2 proportion z test/interval, (13, 5b; 12, 4; 09b, 3a; 07, 5c; 07b, 6a; 06b, 2a; 04b, 6a; M7, 39; M2, 22; M2, 40)**

We are going to do a 2 proportion z interval. The conditions that need to be met in order to proceed are that the two samples are independent, random samples and that both distributions are approximately normal.

Random/Independent → Since this is an experiment where people were randomly assigned to the two treatments, it is safe to say that the two groups should be independent.

| | | |
|---|---|---|
| Normal → | np > 10 | n(1 – p) > 10 |
| Cell: | 144(0.2917) > 10 | 144(0.7083) > 10 |
| Pass: | 144(0.0833) > 10 | 144(0.9167) > 10 |

**Know how to calculate a confidence interval for a z confidence interval for a difference of two proportions (2 prop z interval), (09b, 6b; 06b, 2a; M7, 4)**

$$CI = (\hat{p}_{cell} - \hat{p}_{passenger}) \pm z^* \sqrt{\frac{\hat{p}_{cell}(1-\hat{p}_{cell})}{n} + \frac{\hat{p}_{pass}(1-\hat{p}_{pass})}{n}}$$

$$CI = (0.2917 - .0833) \pm 1.645 \sqrt{\frac{(0.2917)(0.7083)}{144} + \frac{(0.0833)(0.9167)}{144}}$$

CI = (0.13542, 0.28125)

Note: We did NOT pool the data, because in a confidence interval we are assuming that the two proportions are different. That means that they would have two different standard deviations. We are just trying to find what that difference is.

**Know how to write a conclusion in context for a two sample/proportion t/z interval (09, 4a; 09, 5b; 06, 4; 06b, 2a; 05, 6a; 04b, 4a)**

We are 90% confident that the difference in the proportion of the drivers who will miss the exit while on their cell phone and those who will miss the exit if they had a passenger is between 0.13542 and 0.28125.

Since both numbers are positive, we have strong evidence that cell phone users are more likely to be distracted than drivers who have a passenger.

THE THREE STEPS ABOVE MUST BE DONE FOR EVERY CONFIDENCE INTERVAL

# ONE PROPORTION Z-TEST

During a flu vaccine shortage in the United States, it was believed that 45 percent of vaccine-eligible people received flu vaccine. The results of a survey given to a random sample of 2,350 vaccine-eligible people indicated that 978 of the 2,350 people had received flu vaccine.

Is there evidence at the 1% significance level that less than 45 percent of the vaccine-eligible people had received flu vaccine?

**Know how to check the conditions for a one proportion z-test, (11, 6a; 12, 5c; 06b, 6b; 05, 4)**

We are going to do a one proportion z-test. The conditions are that the sample was chosen at random, and that there is an approximately normal distribution.

Random → "a random sample of 2350 vaccine-eligible people"

Normal → np > 10 n(1-p) >10

2350(0.45) > 10 2350(0.55) > 10

Note that we use the null hypothesis p to check for normality, because we are assuming that it is true. That is what a null hypothesis is, an assumption of truth.

10% Check → It is safe to say that there are more than 23500 vaccine eligible people.

**Know how to write a pair of hypotheses for a one propotion z test, including how to define parameters, (11, 6a; 06b, 6a; 05, 4; 03, 2a; M2, 2)**

$H_o: p = 0.45$ let p = the proportion of ALL vaccine eligible people who received the flu vaccine

$H_a: p < 0.45$ Note: we use p, not $\hat{p}$

**Know how to find a z-statistic and a P-value for one proportion z-test, (11, 6a; 06b, 6f; 05, 4)**

$$z = \frac{statistic - parameter}{s.d.} = \frac{\frac{978}{2350} - 0.45}{\sqrt{\frac{(0.4162)(0.5838)}{2350}}} = -3.2964$$

P = 0.0004896

**Know how to draw a conclusion based on a P-value for a one proportion z-test in context, (11, 6a; 12, 5b; 06b, 6f; 05, 4)**

We reject the null hypothesis at the 1% significance level because 0.0004896 < 0.01. There is significant evidence that that less than 45 percent of the vaccine-eligible people had received flu vaccine.

THE FOUR STEPS ABOVE MUST BE DONE FOR EVERY HYPOTHESIS TEST

# More on Hypothesis Tests

**Know how to interpret a P-value in context**

For the above problem, we want to interpret 0.0004896.

General: If we assumed that the null hypothesis was true, there is a P-value percent chance of getting the data that we got or more extreme/lower/higher.

Problem Specific: If we assume that 45% of the vaccine-eligible people received a flu vaccine, there is a 0.05% chance of getting a statistics of 41.62% or lower in a random sample of 2350 people.

**Note: interpreting P-values is the same for any hypothesis test**

**Know what a Type 1 and Type 2 error, and be able to distinguish which one has worse consequences in the context of a problem if given a null and alternative hypothesis, (12, 5a; 11b, 4d; 09, 5c; 08b, 4b, 03, 2b)**

In the vaccine problem above, this is how they would be described.

Type 1: Thinking that there there were fewer than 45% of vaccine eligible people getting a flu vaccine, when really 45% were.

Type 2: Thinking that 45% of vaccine eligible people were getting a flu vaccine, when in reality less than 45% were.

**Note: These interpretations apply to all hypothesis tests**

# 2 PROPORTION Z-TEST (The only time that we pool data)

A French study was conducted in the 1990s to compare the effectiveness of using an instrument called a cardiopump with the effectiveness of using traditional cardiopulmonary resuscitation (CPR) in saving lives of heart attack victims. Heart attack patients in participating cities were treated with either a cardiopump or CPR, depending on whether the individual's heart attack occurred on an even-numbered or an odd-numbered day of the month. Before the start of the study, a coin was tossed to determine which treatment, a cardiopump or CPR, was given on the even-numbered days. The other treatment was given on the odd-numbered days. In total, 754 patients were treated with a cardiopump, and 37 survived at least one year; while 746 patients were treated with CPR, and 15 survived at least one year.

Is there evidence that the survival rate for patients treated with a cardiopump is significantly higher than the survival rate for patients treated with CPR?

**Know how to write a pair of hypotheses for a two prop z test, (12, 4; 09b, 3b; 07, 5b; 07b, 6a; 04b, 6a; 03b, 3b)**

$H_0: p_{cardiopump} - p_{cpr} = 0$ Note: We use p, not $\hat{p}$

$H_0: p_{cardiopump} - p_{cpr} > 0$

**Know how to check the conditions for a 2 proportion z test/interval, (13, 5b; 12, 4; 09b, 3a; 07, 5c; 07b, 6a; 06b, 2a; 04b, 6a; M7, 39; M2, 22; M2, 40)**

We are going to do a 2 proportion z-test. In order to proceed, the conditions of random, normal, and independent must be met.

Random/Independent → The coin flip should be fine assuming that there is no reason to believe that even or odd numbered days should affect our results. This randomization should form two independent groups.

Normal → Since, we are doing a two proportion z-test, we must pool our data to check this condition. This is the only time we pool our data.

$$\hat{p}_{pool} = \frac{37+15}{754+746} = 0.0347$$

| | Np > 10 | n(1 – p) > 10 |
|---|---|---|
| Pump: | 754(0.0347) > 10 | 743(0.9653) > 10 |
| CPR: | 756(0.0347) > 10 | 756(0.9653) > 10 |

Note: You don't have to use the pooled statistic for this check

**Know how to find a test statistic for a 2 prop z test, and how to look up its corresponding p-value, (12, 4; 09b, 3b; 07b, 6a; 04b, 6a)**

$$z = \frac{statistic - parameter}{s.d.} = \frac{(\hat{p}_{pump} - \hat{p}_{CPR}) - (p_{pump} - p_{CPR})}{\sqrt{\frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_{pump}} + \frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_{CPR}}}} = \frac{(0.0491 - 0.0201) - 0}{\sqrt{\frac{(0.0347)(0.9653)}{754} + \frac{(0.0347)(0.9653)}{746}}} = 3.06605$$

Note that this is the ONLY time that we pool

P = 0.0010846

**Know how to draw a conclusion in context for a 2 prop z test, (13, 5c; 12, 4; 09b, 3b; 07, 5c;07b, 6a; 04b, 6a)**

We can reject the null hypothesis at the 1% significance level, because 0.0010846 < 0.01. There is evidence that the survival rate for patients treated with a cardiopump is significantly higher than the survival rate for patients treated with CPR.

THE FOUR STEPS ABOVE MUST BE DONE FOR ALL HYPOTHESIS TESTS