University of Guelph Numeracy Project

About Proportions: Examples

UNIVERSITY of GUELPH

TABLE OF CONTENTS

About Proportions: Examples

Introduction to Proportions

Introduction to Proportions

► A recent poll conducted on 500 child educators in Toronto indicated that 25% of these individuals believe that children are evil. How was this estimate derived?

$$\hat{p} = \frac{X}{N}$$

$$\hat{p} = \frac{125}{500}$$

$$\hat{p} = 0.25 \text{ or } 25\%$$

Therefore, when a sample of 500 child educators was polled, 125 (or 25%) answered that they felt children are evil.

Hypothesis Tests for p

Hypothesis Tests for p

However, researchers in teaching education want to test whether the true proportion (p) is DIFFERENT from 0.05, at the 10% significance level. The hypotheses will be:

$$Ho : p = 0.05$$
$$Ha : p \neq 0.05$$

If we plug our values into our equation:

$$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o\ (1-p_o)}{n}}}$$

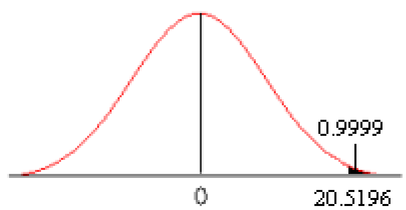
$$Z = \frac{0.25 - 0.05}{\sqrt{\frac{0.05\ (1-0.05)}{500}}}$$

Z = 20.5196

Since we are looking to see whether our true proportion is DIFFERENT from 0.05, we are going to look at a 2-sided z-test.

In order to calculate the p-value from our Z-score, we will be looking at the following Z-distribution:

The Z-score is 20.5196, we can take this and use it to find a p-value in a probability table. We have already done that for you.



Value From Probability Table = 0.9999

1 – 0.9999 = 0.0001
p-value = 2 x 0.0001 = 0.0002 (since we are doing a 2-sided test)

Therefore, we do have significant evidence that the true proportion is different than 0.05, as it is significant at the 10% level.

Confidence Intervals and Wilson Estimate

Confidence Intervals for p

► Here is a table that shows the most commonly used confidence levels and Z* values:

Z*	1.645	1.960	2.576
C	90%	95%	99%

For the purposes of this example, let us say we wish to determine our confidence interval with 90% confidence. Using the values that we know for $\hat{p}$, n and Z* from previously:

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \quad = \quad 0.25 \pm 1.645\sqrt{\frac{0.25(1-0.25)}{500}}$$

Therefore, we are 90% confident that the true proportion of ALL Torontonian child educators who feel children are evil lies within the interval 0.25 ± 0.031855.

Recall that when we tested the hypothesis that our true parameter p is NOT different than 0.05, we had significant evidence that the true proportion was different than 0.05. This was due to the fact that it was significant at the 10% level.

This significant finding is not surprising, as 0.05 lies outside our interval!

The Wilson Estimate

► Recall that we were 90% confident that the true proportion of ALL Torontonian child educators who feel children are evil lies within the interval 0.25 ± 0.031855.

By re-adjusting $\hat{p}$ using the Wilson estimate, we find:

$$\tilde{p} = \frac{X+2}{n+4} \quad = \quad \frac{127}{504} \quad = \quad 0.2520$$

$$\tilde{p} \pm Z^* \sqrt{\frac{\tilde{p}\,(1-\tilde{p})}{n+4}} \quad = \quad 0.2520 \pm 1.645\sqrt{\frac{0.2520\ (1-0.2520)}{504}}$$

= 0.2520 ± 0.031813, very similar to our previous confidence interval. This is most likely due to our large sample size.

Comparing Two Proportions

Introduction: Comparing Two Proportions

► A study of 3926 births by Caesarean section revealed the following breakdown:

Gender of Child	***n***	**>10 lbs**
Male	2100	437
Female	1826	734

Do you think there is a significant difference between the proportions for males and females?

1) Draw 2 independent samples to determine p for each:

Let $\hat{p}_{Male} = \frac{X_1}{n_1}$ $\hat{p}_{Female} = \frac{X_2}{n_2}$

$\hat{p}_M = \frac{437}{2100}$ $\hat{p}_F = \frac{734}{1826}$

Pooled Proportion, $\hat{p}$:

$$\hat{p} = \frac{437 + 734}{2100 + 1826} = \frac{1171}{3926}$$

2) State hypotheses:

H_o: $p_F = p_M$

H_a: $p_F \neq p_M$

3) Calculate Z:

$$Z = \frac{\hat{p}_M - \hat{p}_F}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_M} + \frac{1}{n_F}\right)}}$$

$$Z = \frac{\frac{437}{2100} - \frac{734}{1826}}{\sqrt{\frac{1171}{3926}\left(1 - \frac{1171}{3926}\right)\left(\frac{1}{2100} + \frac{1}{1826}\right)}}$$

Z = -452.36

4) Find the p-value for the 2-sided alternative:

$p < 0.0001$, as indicated by z-tables.

Remember, since we are looking at a 2-sided alternative, we will have to multiply our p-value by 2:

So, $p < 0.0002$

Therefore, we are fairly certain that there is significant evidence that females born by Caesarean weigh more than their male counterparts, as results are significant at both the 5% and 1% level.

5) Calculate confidence interval:

$$C = \hat{p}_M - \hat{p}_F \pm \underbrace{z^* \sqrt{\frac{\hat{p}_M(1-\hat{p}_M)}{n_M} + \frac{\hat{p}_F(1-\hat{p}_F)}{n_F}}}_{\text{Standard Error}}$$

Substituting our values into the equation for a 90% confidence interval, we get:

$$C = \frac{437}{2100} - \frac{734}{1826} \pm 1.645\sqrt{\frac{\frac{437}{2100}\left(1 - \frac{437}{2100}\right)}{2100} + \frac{\frac{734}{1826}\left(1 - \frac{734}{1826}\right)}{1826}}$$

C = -0.19388 ± 0.023845

Therefore, we are 90% confident that the true values of our 2 proportions lie within this interval.

It is also important to note that Wilson's estimate can be applied to 2 proportions in a similar fashion to that of a single proportion.