

Single-Group Statistical Tests with a Binary Dependent Variable

***z* test for proportions**

Many surveys use a simple statistical test that is analogous to the single sample *t* test we used to investigate whether a company paid a higher than (state) average wage. In this case, the researcher is interested in whether one candidate (or side of an issue) would receive more votes than an alternative candidate. Survey participants are asked a single question which has two possible options, such as “yes” or “no.” The statistical test investigates whether there are significantly more “yes” than “no” responses.

There are two tests designed for this circumstance. One of these tests is a *z* test that is very similar to the single-group *t* test, called the *z* test for the difference between two proportions. The formula looks like this:

$$z = \frac{p - \pi}{\sqrt{\pi(1 - \pi) / n}}$$

In the formula, *p* is the proportion of the sample choosing one of the options in the survey (e.g., “yes”), *π* is the null hypothesis value (i.e., the proportion expected if there is no difference between “yes” and “no”), and *n* is the sample size. If you look carefully, you will see that this formula parallels the single-group *t* test, because the denominator (bottom portion) is a standard error, which we could call *s_π*,

$$z = \frac{p - \pi}{s_{\pi}}$$

where $s_{\pi} = \sqrt{\pi(1 - \pi) / n}$ for the standard error.¹ The top part of the equation is parallel as well, because it concerns the difference between the sample and population means ($\bar{X} - \mu$).

As an example, I use data extrapolated from reports of a 2020 presidential election poll in Georgia.² The results in a head-to-head question with just Biden and Trump, in which I excluded the undecideds (*n* = 988), had Biden up among registered voters by 51.0% to 44.0%. To determine whether this is a significant difference, we need only choose one proportion—the proportion for either Biden or Trump, it does not matter. The null hypothesis is that voters in the population are perfectly split 50/50 (i.e., the proportion is .50), so *π* = .5. If the proportion of the sample for one candidate differs from this value relative to what we expect due to sampling variability (chance), then one candidate has a significant lead over the other. If we plug in our obtained values, we get the following result:

$$\begin{aligned} z &= \frac{p - \pi}{\sqrt{\pi(1 - \pi) / n}} \\ &= \frac{.54 - .50}{\sqrt{.50(1 - .50) / 988}} \\ &= \frac{.04}{.0159} \\ &= 2.52 \end{aligned}$$

This obtained value is compared to the critical value obtained in the *z*-table (Table C.2 in the text) that corresponds to the outer 2.5% of the sampling distribution, which is our conventional significance cutoff. With the *z* test, the critical value is always 1.96 for two-tailed significance regardless of sample size (i.e., there is only one normal curve). Because our computed value of 2.52 exceeds this cutoff value (*p* < .05, because the values only exceeds the cutoff for *α* = .05 two-tailed), there is a significant difference between the proportion that preferred Biden and the proportion that preferred Trump.

¹ This formula has a parallel to our single-group *t* test standard error formula, $s_{\bar{x}} = s / \sqrt{n}$, because $\pi(1 - \pi)$ is a convenient formula for the calculation of the variance of a proportion (i.e., the test is really parallel to the single-group *t* test where the variance is known, because we use the population variance $\pi(1 - \pi)$). This form is called the *score* test. Another form uses $p(1 - p)$ instead and is called the *Wald* test.

² These results are taken from a Quinnipiac University poll from Oct 14, 2020 in Georgia among likely voters, <https://poll.qu.edu/georgia/release-detail?ReleaseID=3679>. Methodological details are here https://poll.qu.edu/images/polling/ga/ga10142020_demos_bgwc96.pdf. These results are “extrapolated” here because the survey is weighted for demographics, because I excluded other categories (“other” “wouldn’t vote” and “don’t know/refused”), and because some rounding is necessary to construct the counts to match the percents given in the report.

With a z -proportions test, one can also construct “confidence limits” or a “confidence interval.” The confidence limits describe the amount of sampling variability that might be expected from random chance. In other words, if we were to draw a large number of random samples from the same population, we would not get the same proportion estimate (.54 for Biden) each time. We would expect some variability in this estimate resulting from random sampling chance. The 95% confidence interval is an estimate of the range of these possible values (more precisely, 95% of this range). In the case of the z test, we use the normal distribution and our estimate of standard error to construct the interval using the following formula.

$$p \pm (z_{critical})(s_{\pi}),$$

where the $z_{critical}$ is the critical value, which is 1.96 whenever the normal distribution is used. For our example above, we get the following values for the lower confidence limit (LCL) and the upper confidence limit (UCL):

$$LCL = .54 - (1.96)(.0159) = .54 - .031 = .51$$

$$UCL = .54 + (1.96)(.0159) = .54 + .031 = .57$$

Thus, the 95% confidence interval is .51-.57. This interval does not include the null hypothesis value of .50, suggesting that the difference from an equal proportion is unlikely to be due to random sampling chance. Whenever the confidence limits include the null value, you will find that the significance test will have a non-significant result. Half of this confidence interval is what is commonly called the *margin of error*, and is typically expressed in terms of a percentage. We can just use the .031 subtracted to find the confidence interval multiplied by 100 to find a percent (i.e., $.031 \times 100 = 3.1\%$) or we can compute the margin of error by subtracting the LCL from the UCL and dividing by two [$(.57 - .51)/2 \times 100 = .06/2 \times 100 = 3.0\%$]. The two methods are equivalent but may differ slightly depending on whether or when rounding is used.

Chi-square test

A second, equivalent test for this problem is a chi-square test. The chi-square compares frequencies obtained in the sample to those expected according to the null hypothesis (i.e., no difference in the population). The chi-square formula looks like this:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where \sum is the summation sign, indicating addition across all the “cells,” O is the observed frequency (obtained from the survey), and E is the frequency expected if the two “cells” were equal. If we translate our presidential survey into frequencies, we would obtain the following result displayed in a two-cell table:

Biden	Trump	Total
534	454	988

Using the chi-square formula, we would get the following result (note: expected frequencies often have decimals):

$$\begin{aligned} \chi^2 &= \sum \frac{(O - E)^2}{E} \\ &= \frac{(534 - 494)^2}{494} + \frac{(454 - 494)^2}{494} \\ &= \frac{1600}{494} + \frac{1600}{494} \\ &= 3.24 + 3.24 \\ &= 6.48 \end{aligned}$$

This computed value is compared to a critical value obtained from the chi-square table (Table C.4 in the text). It is a 1-*degree of freedom* (df) test, and chi-square for a two-tailed 1- df test is always 3.84. Our computed value does exceed this value, so voters were significantly more likely to prefer Biden over Trump.

The z test and the chi-square test will always give identical results, in fact, $z^2 = \chi^2$ (allowing for rounding error).

More than Two Cells

The chi-square formula is quite general, and as long as we can compute the expected frequencies based on what is expected due to chance (or another null hypothesis), we can simply employ the same equation to test whether any of the cell frequencies are different from one another whether there are three, four, or more cells. For example, we could compare Republicans, Democrats, and independents or yes, no, and undecided or multiple candidates, in which case expected frequencies would be computed by multiplying n by .333.

Effect Size

It seems that researcher's rarely report effect size for simple chi-square tests like these (probably because software packages typically do not print it out), but it is useful to go beyond just determining significance. Cohen's w (Cohen, 1988) is based on the magnitude of the differences between the observed and expected values, and so it is easily computed once the chi-square has been obtained. Cohen suggested that .1 is small, .3 is medium, and .5 is large.

$$w = \sqrt{\sum \frac{(O - E)^2}{E}} = \sqrt{\frac{\chi^2}{N}}$$

In our example, $w = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{6.48}{988}} = .08$, a small effect

SPSS

Please note that when using the menus, there are a couple of ways to obtain the single group z -proportions test. I recommend Analyze -> Nonparametric tests -> Legacy dialogs -> binomial which gives simpler output and does not require defining the variable type.³

```
nptests /onesample test (response) binomial (testvalue=.5 successcategorical=list(1) likelihood ).
*For binomial (z-proportion) test, successcategorical=list(1) chooses the value of 1 (Biden) as the comparison proportion
*testvalue=.5 gives the null proportion (default and can be omitted)
*likelihood gives CIs based on the sample SE estimate (Wald) rather than the null value SE estimate.
*The z-value printed uses a continuity correction (and will not match other programs unless the continuity correction is requested.
```

One-Sample Binomial Test Summary

Total N	988
Test Statistic	534.000
Standard Error	15.716
Standardized Test Statistic	2.513
Asymptotic Sig.(2-sided test)	.012

Confidence Interval Summary

Confidence Interval Type	Parameter	Estimate	95.0% Confidence Interval	
			Lower	Upper
One-Sample Binomial Success Rate (Likelihood)	Probability(intended vote=Biden).	.540	.509	.571

```
nptests /onesample test (response) chisquare.
```

One-Sample Binomial Test Summary

Total N	988
Test Statistic	534.000
Standard Error	15.716
Standardized Test Statistic	2.513
Asymptotic Sig.(2-sided test)	.012

³ When using Analyze -> Nonparametric tests -> One sample, SPSS requires that you assign the variable as binary "nominal". Then, by default SPSS will give results in the "model viewer" form, which is a graphical depiction of the result (just so you know, I hate this!). I turn this off by going to Edit->Options->output tab->select "Pivot tables and charts."

R

```
> #get the counts of the number of 1s, can also use the summarytools freq function  
> #this also gives the chi-square version of the test  
> library(lessR)  
> SummaryStats(response)
```

```
--- response ---
```

```
Missing values of response: 0
```

	0	1	Total
Frequencies:	454	534	988
Proportions:	0.460	0.540	1.000

```
Chi-squared test of null hypothesis of equal probabilities  
Chisq = 6.478, df = 1, p-value = 0.011
```

```
> #then enter in the number of cases into prop.test(x,n,p,continuity correction option)  
> #where x is the number of successes (Biden voters)  
> prop.test(534, 988, p=0.5, correct=FALSE)
```

```
1-sample proportions test without continuity correction
```

```
data: 534 out of 988, null probability 0.5  
X-squared = 6.4777, df = 1, p-value = 0.01092  
alternative hypothesis: true p is not equal to 0.5  
95 percent confidence interval:  
 0.5093139 0.5713442  
sample estimates:  
      p  
0.5404858
```

Sample write-up

A z -proportions test was used to test whether significantly more likely voters preferred Joseph Biden over Donald Trump for president. Of the 988 voters surveyed, 534 (54.0%) preferred Biden and 454 (46.0%) preferred Trump. The difference was statistically significant, $z = 2.51$, $p = .012$, indicating that the greater preference for Biden was greater than what would be expected due to chance.⁴ The margin of error for this survey was 3.1%.⁵

⁴ In practice, because the chi-square and the z -proportion tests are equivalent, there would be no need to do both. Either one might be used by a researcher, although survey results are more often reported in the media in terms of percentages and margin of error. For the chi-square test, I would suggest also reporting Cohen's w . For example, $\chi^2(1) = 6.48$, $p = .011$, Cohen's $w = .08$.

⁵ Confidence limits could be reported instead of the margin of error, e.g., $z = 2.51$, $p = .012$, 95% CI [.51, .57].