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 other cases, we may have a single binary variable, such as a political poll about whether one candidate (or side of an issue) would receive more votes than an alternative candidate or whether there are more survey responses favoring one choice over another when there are two possible options, such as "yes or no, do you support affirmative action in college admissions?". For these kinds of research questions, the statistical test investigates whether there is a significantly higher percentage choosing one option than the other.

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 from a YouGov survey about voter participation in the 2020 election.[2] The results presented here (n = 1,092) concern whether those who were surveyed were more likely to support the mandate in the Affordable Care Act to require health insurance for everyone. The percentage who said they supported the mandate was 45.2% (n = 494) and those who opposed was 54.8% (n = 598). To determine whether this is a significant difference, we need only choose one proportion—the proportion for either support or oppose. It does not matter. The null hypothesis is that the proportion in the population who support and oppose the measure are perfectly split 50/50 (i.e., the proportion is .50), so π = .5. If the proportion of the sample who support vs. oppose differs from what we expect due to sampling variability (chance), then one option is favored significantly more than another.

If we plug in our obtained values (but rounding in my example just to simplify), we get the following result:

$$\begin{aligned} z &= \frac{p-\pi}{\sqrt{\pi(1-\pi)/n}} \\ &= \frac{.55-.50}{\sqrt{.50(1-.50)/1092}} \\ &= \frac{.05}{.0151} \\ &= 3.31 \end{aligned}$$

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

---

[1] This formula has a parallel to our single-group t test standard error formula, $s_{\bar{Y}} = s/\sqrt{n}$, 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 $p(1\text{-}p)$ instead and is called the *Wald* test.

[2] These results use a random sample taken from a YouGov/Harvard/MIT survey called the Cooperation Election Study (CES) conducted by Stephen Ansolabehere & Brian Schaffner, 2022, "CES Common Content, 2021", https://doi.org/10.7910/DVN/OPQOCU

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 3.31 exceeds this cutoff value ($p < .05$), there is a significant difference between the proportion that oppose the mandate and those that support the mandate (significantly more oppose).

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 (.55 who oppose) 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):

$$LCL = .55 - (1.96)(.0151) = .55 - .03 = .52$$
$$UCL = .55 + (1.96)(.0151) = .55 + .03 = .58$$

Thus, the 95% confidence interval is .52-.58. 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 .03 subtracted to find the confidence interval multiplied by 100 to find a percent (i.e., $.03 \times 100 = 3.0\%$) or we can compute the margin of error by subtracting the LCL from the UCL and dividing by two [$(.58 - .52)/2 \times 100 = .06/2 \times 100 = 3.0\%$]. 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 voter survey into frequencies, we would obtain the following result displayed in a two-cell table:

Oppose	Support	*Total*
598	494	*1092*

Using the chi-square formula, we would get the following result (note: expected frequencies often have decimals), where $E = 1092/2 = 546$:

$$\begin{aligned}\chi^2 &= \sum \frac{(O-E)^2}{E} \\ &= \frac{(598-546)^2}{546} + \frac{(494-546)^2}{546} \\ &= \frac{2704}{546} + \frac{2704}{546} \\ &= 4.95 + 4.95 \\ &= 9.90\end{aligned}$$

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 oppose than support the mandate.

The z test and the chi-square test are equivalent, and, in fact, $z^2 = \chi^2$. We should expect them to always lead to the same conclusion. *Note that because of my rounding of the response proportions initially, these values are only approximately equal here,* $(3.31)^2 \approx 9.90$.

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, $w = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{9.90}{1092}} = .10$, 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. The Analyze -> Nonparametric tests -> Legacy dialogs -> binomial which gives simpler output and does not require defining the variable type, but it uses the Fisher's exact test that I do not usually recommend.[3]

```
nptests   /onesample test (mandate) 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

*The (1) refers to the group with code = 1. The variable must be nominal for the successcategorical command to
be recognized.  To change the variable to nominal use:  variable level mandate(nominal). This only matters if
testvalue is not .5.

*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.
```

One-Sample Binomial Test Summary

Total N	1092
Test Statistic	598.000
Standard Error	16.523
Standardized Test Statistic	3.117
Asymptotic Sig.(2-sided test)	.002

Confidence Interval Summary

Confidence Interval Type	Parameter	Estimate	95.0% Confidence Interval	
			Lower	Upper
One-Sample Binomial Success Rate (Likelihood)	Probability(mandate Health Care -- Restore the Affordable Care Act's mandate that all individuals be required to purchase health insurance.=0 Oppose).	.548	.518	.577

[3] When using Analyze -> Nonparametric tests -> One sample -> settings tab, and then check either (Binomial test) or (Chi-square test). SPSS may require that you assign the variable as binary "nominal".

You can also conduct the chi-square test with the nptests procedure—***there is no need to do chi-square and the z-test***, and for this problem, the z-test is more common because you can get confidence intervals.

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

One-Sample Chi-Square Test Summary

Total N	1092
Test Statistic	9.905 [a]
Degree Of Freedom	1
Asymptotic Sig.(2-sided test)	.002

a. There are 0 cells (0%) with expected values less than 5. The minimum expected value is 546.

R

The lessR function Prop_test gives frequencies but also the chi-square test. You could also use the `summarytools` package `freq` function for the frequencies and proportions of each variable. The R base function prop.test requires manual entry of frequency of one cell and the total n, so you must obtain the frequencies first.

```
> Prop_test(mandate)

<<< Chi-squared test for given probabilities

variable: mandate

--- Description
                 0         1
---------  --------  --------
observed       598       494
expected   546.000   546.000
residual     2.225    -2.225
stdn res     3.147    -3.147

--- Inference

Chi-square statistic: 9.905
Degrees of freedom: 1
Hypothesis test of equal population proportions: p-value = 0.002
```

The above analysis provides a significance test of the hypothesis, but if you want confidence intervals you need to do another analysis.

```
> #then enter in the number of cases into prop.test(x,n,p,continuity correction option)
> #where x is the number of successes (oppose mandate)
> prop.test(598, 1092, p=0.5, correct=FALSE)

         1-sample proportions test without continuity correction

data:  598 out of 1092, null probability 0.5
X-squared = 9.9048, df = 1, p-value = 0.001649
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5179826 0.5769216
sample estimates:
       p
0.547619
```

Sample write-up

A z-proportions test was used to test whether significantly more respondents support or oppose the health insurance mandate. Of the 1092 voters surveyed, 598 (54.8%) opposed the mandate and 463 (45.2%) supported the mandate. The difference was statistically significant, z =3.12, p = .002, indicating that the opposition to the mandate was greater than what would be expected due to chance.[4] The margin of error for this survey was 3.0%.[5]

Note that, the z value can be obtained from the chi-square in the R output by taking the square root (square root of 9.9048 = 3.12).

[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, $\chi^2(1)$ = 9.90, p = .002, Cohen's w = .10.

[5] Confidence limits could be reported instead of the margin of error, e.g., z =3.12, p = .002, 95% CI[.52,.58].