Single-Group Statistical Tests with a Binary Dependent Variable

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 survey example, 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:

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

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 $s_\pi = \sqrt{\pi\left(1-\pi\right)/n}$.[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, let's assume we conducted a survey of 500 likely voters to forecast the 2012 Presidential election to see whether voters preferred President Obama over former Massachusetts governor Mitt Romney. Let's also assume our results indicated that 52% (or .52 expressed as a proportion) preferred Obama (i.e., 48% preferred Romney). If the voters were perfectly split, 50% would be expected to vote for each. Thus, our null hypothesis was that, in the population, there was a proportion (π) of .5 who preferred Obama. If we plug in our obtained values, we get the following result:

$$\begin{aligned} z &= \frac{p-\pi}{\sqrt{\pi\left(1-\pi\right)/n}} \\ &= \frac{.52-.50}{\sqrt{.50\left(1-.50\right)/500}} \\ &= \frac{.02}{.02236} \\ &= .89 \end{aligned}$$

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 .89 does not exceed this cutoff value, there is no significant difference between the proportion that preferred Obama and the proportion that preferred Romney.

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 (.52 for Obama) 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.

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

[1] This formula has a parallel to our single-group t-test standard error formula, $s_{\bar{Y}} = s/\sqrt{n}$, because $\pi\left(1-\pi\right)$ is a convenient formula for calculation the variance of a proportion (note: the test is really parallel to the single-group t-test where the variance is known, because we use the population variance).

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):

$$LCL = .52 - (1.96)(.02236) = .52 - .04 = .48$$
$$UCL = .52 + (1.96)(.02236) = .52 + .04 = .56$$

Thus, the 95% confidence interval is .48-.56. This interval includes the null hypothesis value of .50, suggesting that the difference from an equal proportion may 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. So, the confidence limits are just a different way of viewing the significance test, with some added information about how variable your sample estimate might be expected to be. Half of this confidence interval is what is commonly called the *margin of error*, and is typically expressed in terms of a percentage. The margin of error for our hypothetical survey then is (.52 - .48) X 100 = .04 X 100 = 4%.

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:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

where $\sum$ is the summation sign, indicating addition across all the "cells," f_o is the observed frequency (obtained from the survey), and f_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:

Romney	**Obama**	*Total*
240	**260**	*500*

Using the chi-square formula, we would get the following result:

$$\begin{aligned}\chi^2 &= \sum \frac{(f_o - f_e)^2}{f_e} \\ &= \frac{(240-250)^2}{250} + \frac{(260-250)^2}{250} \\ &= \frac{100}{250} + \frac{100}{250} \\ &= .4 + .4 \\ &= .8\end{aligned}$$

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 not exceed this, so voters were not significantly more likely to prefer Obama over Romney.

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

Sample write-up

A z-proportions test was used to test whether significantly more likely voters preferred President Obama over Mitt Romney for president. Of the 500 voters surveyed, 260 (52%) preferred Obama and 240 (48%) preferred Romney. The difference was not statistically significant (z = .89, ns), indicating that the greater preference for Obama was not greater than what would be expected due to chance.[2] The margin of error for this survey was 4%.

---

[2] 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.