Equivalence of Chi-Square Tests and *Z*-Tests for Proportions

In this section we will show that *Z*-tests for proportions are algebraically equivalent to chi-square tests. We will begin with the test for a single proportion where the hypotheses are

$$H_0\colon p = p_0 \text{ and } H_a\colon p \neq p_0 .$$

To perform a chi-square test, we would have observed and expected values as shown in the table below:

	Yes	No	Total
Observed	y	$n-y$	n
Expected	np_0	$n(1-p_0)$	n

Since $\hat{p} = y/n$, $y = n\hat{p}$, this table is equivalent to

	Yes	No	Total
Observed	$n\hat{p}$	$n-n\hat{p}$	n
Expected	np_0	$n(1-p_0)$	n

The chi-square statistic with one degree of freedom is

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^{2} \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \\
&= \frac{(y-np_0)^2}{np_0} + \frac{[(n-y)-n(1-p_0)]^2}{n(1-p_0)} \\
&= \frac{(n\hat{p}-np_0)^2}{np_0} + \frac{[(n-n\hat{p})-n(1-p_0)]^2}{n(1-p_0)} \\
&= \frac{n^2(\hat{p}-p_0)^2}{np_0} + \frac{n^2[(1-\hat{p})-(1-p_0)]^2}{n(1-p_0)} \\
&= \frac{n(\hat{p}-p_0)^2}{p_0} + \frac{n(-\hat{p}+p_0)^2}{(1-p_0)} \\
&= n(\hat{p}-p_0)^2\left\{\frac{1}{p_0} + \frac{1}{1-p_0}\right\} \\
&= n\left(\hat{p}-p_0\right)^2\left\{\frac{1}{p_0(1-p_0)}\right\} \\
&= \frac{(\hat{p}-p_0)^2}{p_0(1-p_0)/n}.
\end{aligned}$$

This last expression we recognize as Z^2 with $Z^2 = \left[\dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0\left(1-p_0\right)}{n}}}\right]^2$

Note that large values of n are required for $\sum_{i=1}^{2}\frac{(\text{observed}-\text{expected})^2}{\text{expected}}$ to be approximately χ^2 distributed and for $\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ to be approximately Z. We already know that Z^2 is χ^2 with 1 degree of freedom. It should be no surprise that the critical value for χ^2 with 1 degree of freedom with $\alpha = 0.05$ is $1.96^2 = 3.84$.

Two Proportions Tests

Now consider the test for two proportions where the hypotheses are

$$H_0: p_1 = p_2 \text{ and } H_a: p_1 \neq p_2$$

For this test, $\hat{p}_1 = \frac{y_1}{n_1}$, $\hat{p}_2 = \frac{y_2}{n_2}$, and pooled $\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{y}{n}$.

To perform a chi-square test, we have values as shown in the tables below:

Observed Values

	Group 1	Group 2	Total
Success	y_1	y_2	$y_1 + y_2 = y$
Failure	$n_1 - y_1$	$n_2 - y_2$	$n_1 + n_2 - y_1 - y_2 = n - y$
Total	n_1	n_2	$n_1 + n_2 = n$

The expected number of successes is $\frac{n_1 y}{n}$ for Group 1 and $\frac{n_2 y}{n}$ for Group 2. The expected number of failures is $\frac{n_1(n-y)}{n}$ for Group 1 and $\frac{n_2(n-y)}{n}$ for Group 2. These values are summarized in the table below.

Expected Values

	Group 1	Group 2	Total
Success	$\frac{n_1 y}{n} = n_1\hat{p}$	$\frac{n_2 y}{n} = n_2\hat{p}$	y
Failure	$\frac{n_1(n-y)}{n} = n_1(1-\hat{p})$	$\frac{n_2(n-y)}{n} = n_2(1-\hat{p})$	$n - y$
Total	n_1	n_2	n

The chi-square statistic with one degree of freedom is

$$\chi^2 = \sum_{i=1}^{4}\frac{(\text{observed}-\text{expected})^2}{\text{expected}}$$

$$
\begin{aligned}
&= \frac{\left(y_1 - n_1\hat{p}\right)^2}{n_1\hat{p}} + \frac{\left(y_2 - n_2\hat{p}\right)^2}{n_2\hat{p}} + \frac{\left[(n_1 - y_1) - n_1(1-\hat{p})\right]^2}{n_1(1-\hat{p})} + \frac{\left[(n_2 - y_2) - n_2(1-\hat{p})\right]^2}{n_2(1-\hat{p})} \\
&= \frac{\left(n_1\hat{p}_1 - n_1\hat{p}\right)^2}{n_1\hat{p}} + \frac{\left(n_2\hat{p}_2 - n_2\hat{p}\right)^2}{n_2\hat{p}} + \frac{\left[n_1(1-\hat{p}_1) - n_1(1-\hat{p})\right]^2}{n_1(1-\hat{p})} + \frac{\left[n_2(1-\hat{p}_2) - n_2(1-\hat{p})\right]^2}{n_2(1-\hat{p})} \\
&= \frac{n_1(\hat{p}_1 - \hat{p})^2}{\hat{p}} + \frac{n_2(\hat{p}_2 - \hat{p})^2}{\hat{p}} + \frac{n_1(\hat{p} - \hat{p}_1)^2}{(1-\hat{p})} + \frac{n_2(\hat{p} - \hat{p}_2)^2}{(1-\hat{p})} \\
&= n_1\left(\hat{p}_1 - \hat{p}\right)^2\left\{\frac{1}{\hat{p}} + \frac{1}{1-\hat{p}}\right\} + n_2\left(\hat{p}_2 - \hat{p}\right)^2\left\{\frac{1}{\hat{p}} + \frac{1}{1-\hat{p}}\right\} \\
&= \left\{\frac{1}{\hat{p}} + \frac{1}{1-\hat{p}}\right\}\left[n_1\left(\hat{p}_1 - \hat{p}\right)^2 + n_2\left(\hat{p}_2 - \hat{p}\right)^2\right]
\end{aligned}
$$

$$\text{note that } \hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

$$
\begin{aligned}
&= \left\{\frac{1}{\hat{p}} + \frac{1}{1-\hat{p}}\right\}\left[n_1\left(\hat{p}_1 - \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}\right)^2 + n_2\left(\hat{p}_2 - \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}\right)^2\right] \\
&= \frac{1}{\hat{p}(1-\hat{p})}\left[n_1\left(\frac{n_2\hat{p}_1 - n_2\hat{p}_2}{n_1 + n_2}\right)^2 + n_2\left(\frac{n_1\hat{p}_2 - n_1\hat{p}_1}{n_1 + n_2}\right)^2\right] \\
&= \frac{1}{\hat{p}(1-\hat{p})}\left[\frac{n_1 n_2^2\left(\hat{p}_1 - \hat{p}_2\right)^2}{(n_1 + n_2)^2} + \frac{n_2 n_1^2\left(\hat{p}_1 - \hat{p}_2\right)^2}{(n_1 + n_2)^2}\right] \\
&= \frac{1}{\hat{p}(1-\hat{p})}\frac{\left(\hat{p}_1 - \hat{p}_2\right)^2}{\left(n_1 + n_2\right)^2}[n_1 n_2(n_2 + n_1)] \\
&= \frac{1}{\hat{p}(1-\hat{p})}\frac{\left(\hat{p}_1 - \hat{p}_2\right)^2}{\left(n_1 + n_2\right)}(n_1 n_2) \\
&= \frac{\left(\hat{p}_1 - \hat{p}_2\right)^2}{\hat{p}(1-\hat{p})\left(\dfrac{n_1 + n_2}{n_1 n_2}\right)} = \frac{\left(\hat{p}_1 - \hat{p}_2\right)^2}{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)} \\
&= \left[\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{1/n_1 + 1/n_2}}\right]^2 = Z^2
\end{aligned}
$$

Note that this Z is approximate since the variance is estimated from the data. As before, we note that χ^2 on a 2 x 2 table has 1 degree of freedom, and Z^2 is a χ^2 with 1 degree of freedom.