One Sample: Test on a Single Proportion

We shall consider the problem of testing the hypothesis that the proportion of successes in a binomial experiment equals some specified value. That is, we are testing the null hypothesis H0 that p = p0, where p is the parameter of the binomial distribution.

The alternative hypothesis may be one of the usual one-sided or two-sided alternatives:

$$p < p_0, \quad p > p_0, \quad p \neq p_0$$

To test the hypothesis

H_0: $p=p_0$,

H_1: $p<p_0$,

we use the binomial distribution to compute the P -value

$$P = P(X \leq x \text{ when } p = p_0).$$

The value x is the number of successes in our sample of size n. If this P-value is less than or equal to α, our test is significant at the α level and we reject H_0 in favor of H_1.

Similarly, to test the hypothesis

$$H_0: p=p_0,$$

$$H_1: p>p_0,$$

at the α-level of significance, we compute

$$P = P(X \geq x \text{ when } p = p_0)$$

and reject H_0 in favor of H_1 if this P-value is less than or equal to α.

Finally, to test the hypothesis

$$H_0: p=p_0,$$

$$H_1: p \neq p_0,$$

at the α-level of significance, we compute

$$P = 2P(X \leq x \text{ when } p = p_0)$$

or

$$P = 2P(X \geq x \text{ when } p = p_0)$$

and reject H_0 in favor of H_1 if the computed P-value is less than or equal to α.

Testing a Proportion (Small Samples)

1. $H_0: p = p_0$.
2. One of the alternatives H_1: $p < p_0$, $p > p_0$, or $p \neq p_0$.
3. Choose a level of significance equal to α.
4. Test statistic: Binomial variable X with $p = p_0$.
5. Computations: Find x, the number of successes, and compute the appropriate P-value.
6. Decision: Draw appropriate conclusions based on the P -value.

Problem 1:

A marketing expert for a pasta-making company believes that 40% of pasta lovers prefer lasagna. If 9 out of 20 pasta lovers choose lasagna over other pastas, what can be concluded about the expert's claim? Use a 0.05 level of significance.

- The number of people prefers lasagne over other pasta is $x = 9$
- The level of significance is $\alpha = 0.05$
- The sample size is, $n = 20$

⋆ The null hypothesis is

$$H_0 : p \leq 0.40$$

⋆ The alternative hypothesis is

$$H_1 : p > 0.40$$

X be binomial variable with $p = 0.04$ and $n = 20$

- Let x be the number of success (i.e. the number of pasta lovers who prefer lasagne) in our sample of size 20

The number of success is $x = 9$

The mean is

$$\begin{aligned} E(X) &= n \cdot p \\ &= 20 \cdot 0.4 \\ &= 8 \end{aligned}$$

The P value is

$$\begin{aligned} P - \text{value} &= P(X \geq 9 \text{ when } p = 0.4) \\ &= \sum_{x=9}^{20} b(x; 20, 0.4) \\ &= \left\{ 1 - \sum_{x=0}^{8} \binom{20}{x} (0.4)^x (1-0.4)^{20-x} \right\} \\ &= [1 - 0.5956] \\ &= 0.4044 \end{aligned}$$

The P-value is greater than the level of significance, so we fail to reject the null hypothesis and conclude that the proportion of pasta lovers prefer is not over 40%.

The normal curve approximation, with parameters $\mu = np_0$ and $\sigma^2 = np_0q_0$, is usually preferred for large n and is very accurate as long as p_0 is not extremely close to 0 or to 1. If we use the normal approximation, the z-value for testing $p = p_0$ is given by

$$z = \frac{x - np_0}{\sqrt{np_0q_0}} = \frac{\hat{p} - p_0}{\sqrt{p_0q_0/n}},$$

which is a value of the standard normal variable Z. Hence, for a two-tailed test at the α-level of significance, the critical region is $z < -z\alpha/2$ or $z > z\alpha/2$. For the one-sided alternative $p < p_0$, the critical region is $z < -z\alpha$, and for the alternative $p > p_0$, the critical region is $z > z\alpha$.

Problem A new radar device is being considered for a certain missile defense system. The system is checked by experimenting with aircraft in which a kill or a no kill is simulated. If, in 300 trials, 250 kills occur, accept or reject, at the 0.04 level of significance, the claim that the probability of a kill with the new system does not exceed the 0.8 probability of the existing device.

Solution:

In $n = 300$ trials, $x = 250$ kills occurred.

We need to accept or reject, at the 0.04 level of significance, the claim that the probability of a kill with the new system does not exceed the 0.8 probability of the existing device.

So,

- we test the null hypothesis $H_0 : p = 0.8$,

- against the alternative hypothesis $H_1 : p < 0.8$,

- at the 0.04 level of significance.

• The z-value for testing $p = p_0$ is given by

$$z = \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}},$$

which is a value of the standard normal variable Z.

• If x represents the number of kills that occur in n trials, then the proportions of successful killing is

$$\widehat{p} = \frac{x}{n} = \frac{250}{300} = 0.83$$

- So, the z-value for testing $p = 0.8$ is given by

$$z = \frac{0.83 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{300}}} \approx 1.3$$

For the one-sided alternative $p < 0.8$, at the 0.04 level of significance, the critical region is $z < -z_{0.04}$.

we obtain that $z_{0.04} = 1.75$, i.e. the critical region is $z < -1.75$.

Since $z = 1.3 > -1.75 = -z_{0.04}$, we do not reject H_0 and conclude that the probability of a kill with the new system does exceed the 0.8 probability of the existing device.

Two Samples: Tests on Two Proportions

The critical regions for the appropriate alternative hypotheses are set up as before, using critical points of the standard normal curve. Hence, for the alternative $p1 /= p2$ at the α-level of significance, the critical region is $z < -z\alpha/2$ or $z > z\alpha/2$.

For a test where the alternative is $p1 < p2$, the critical region is $z < -z\alpha$, and when the alternative is $p1 > p2$, the critical region is $z > z\alpha$.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}.$$

A vote is to be taken among the residents of a town and the surrounding county to determine whether a proposed chemical plant should be constructed. The construction site is within the town limits, and for this reason many voters in the county believe that the proposal will pass because of the large proportion of town voters who favor the construction. To determine if there is a significant difference in the proportions of town voters and county voters favoring the proposal, a poll is taken. If 120 of 200 town voters favor the proposal and 240 of 500 county residents favor it, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use an $\alpha = 0.05$ level of significance.

Let p1 and p2 be the true proportions of voters in the town and county, respectively, favoring the proposal.

1. H_0: $p_1 = p_2$.
2. H_1: $p_1 > p_2$.
3. $\alpha = 0.05$.
4. Critical region: $z > 1.645$.

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0.60, \quad \hat{p}_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0.48, \quad \text{and}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 240}{200 + 500} = 0.51.$$

$$z = \frac{0.60 - 0.48}{\sqrt{(0.51)(0.49)(1/200 + 1/500)}} = 2.9,$$

$$P = P(Z > 2.9) = 0.0019.$$

Decision: Reject H0 and agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters.

In a study on the fertility of married women conducted by Martin O'Connell and Carolyn C. Rogers for the Census Bureau in 1979, two groups of childless wives aged 25 to 29 were selected at random, and each was asked if she eventually planned to have a child. One group was selected from among wives married less than two years and the other from among wives married five years. Suppose that 240 of the 300 wives married less than two years planned to have children some day compared to 288 of the 400 wives married five years. Can we conclude that the proportion of wives married less than two years who planned to have children is significantly higher than the proportion of wives married five years? Make use of a P -value.

⋆ Let n_1 and n_2 be the sample sizes for the two groups respectively.
It is given $n_1 = 300$; $n_2 = 400$

⋆ Let x_1 be the number of wives married less than two years who planned to have children in the sample. It is given $x_1 = 240$.

⋆ Let x_2 be the number of wives married five years who planned to have children in the sample. It is given $x_2 = 288$

- The null hypothesis is: $H_0 : p_1 = p_2$

- The alternative hypothesis is: $H_1 : p_1 > p_2$

The level of significance is $\alpha = 0.05$

$\star$ Let p_1 be the true population proportion of wives married less than two years who planned to have children.

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{240}{300} = 0.8$$

$\star$ Let p_2 be the true population proportion of wives married five years respectively.

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{288}{400} = 0.72$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{240 + 288}{300 + 400} = 0.75$$

The test statistic is

$$
\begin{aligned}
z &= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\
z &= \frac{0.8 - 0.72}{\sqrt{0.75 \times (1 - 0.75)\left(\frac{1}{300} + \frac{1}{400}\right)}} \\
&= \frac{0.08}{0.0329} \\
&= 2.433
\end{aligned}
$$

The P - value is

$$\begin{aligned} P - \text{value} &= P(Z > 2.43) \\ &= 1 - P(Z < 2.43) \\ &= 1 - 0.9925 \\ &= 0.0075 \end{aligned}$$

The P value is smaller than the significance level 0.05. Therefore, we reject H_0 and conclude that the proportion of wives married less than two years who planned to have children is significantly higher than the proportion of wives married five years.

An urban community would like to show that the incidence of breast cancer is higher in their area than in a nearby rural area. (PCB levels were found to be higher in the soil of the urban community.) If it is found that 20 of 200 adult women in the urban community have breast cancer and 10 of 150 adult women in the rural community have breast cancer, can we conclude at the 0.05 level of significance that breast cancer is more prevalent in the urban community?

⋆ Let n_1 denotes the sample size for the urban community. It is given $n_1 = 200$.
⋆ Let n_2 denotes the sample size for the rural community. It is given $n_2 = 150$.

• Let p_1 denotes the true population proportion of women who have breast cancer in the urban community.
• Let p_2 denotes the true population proportion of women who have breast cancer in the rural community.

$\star$ The null hypothesis is

$$H_0 : p_1 = p_2$$

The alternative hypothesis is

$$H_1 : p_1 > p_2$$

The level of significance is $\alpha = 0.05$

$\star$ Let x_1 denotes the number of women in the sample, from urban communities, who have breast cancer. It is given $x_1 = 20$.
$\star$ Let x_2 denotes the number of women in the sample, from rural communities, who have breast cancer. It is given $x_2 = 10$.

The test statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $\hat{p}_1 = \dfrac{x_1}{n_1}, \hat{p}_2 = \dfrac{x_2}{n_2}$, and $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$.

⋆ Therefore, we have:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{20}{200} = 0.1$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{10}{150} = 0.067$$

$$\begin{aligned} \hat{p} &= \frac{x_1 + x_2}{n_1 + n_2} \\ &= \frac{20 + 10}{200 + 150} \\ &= 0.085 \end{aligned}$$

$\star$ Hence, we have:

$$
\begin{aligned}
z &= \frac{0.1 - 0.067}{\sqrt{0.085 \cdot (1 - 0.085)\left(\frac{1}{200} + \frac{1}{150}\right)}} \\
&= 1.1
\end{aligned}
$$

Lets find the P-value:

$$
\begin{aligned}
\text{P-value} &= P(z > 1.1) \\
&= 1 - P(z < 1.1) \\
&= 1 - 0.8643 \\
&= 0.1357
\end{aligned}
$$

Because the P-value is greater than the level of significance (which is $\alpha = 0.05$), we fail to reject H_0 hypothesis and conclude that breast cancer is more frequent in the urban community.