STAT 108 Lecture 17 Tu, Nov 3, 2015

9.4 Hypotheses Tests About a Population Proportion: Large Samples

A sample of size *n* ($n \geq 30$) is drawn, and the sample proportion $\hat{p}$ is observed. Suppose we want to test $H_0 : p = p_0$ against $H_1 : p \geq p_0$ or $H_1 : p \leq p_0$ or $H_1 : p \neq p_0$ where *p* is the true population proportion.

Recall that by the CLT, an approximate distribution of $\hat{p}$ is normal with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.

Thus, under $H_0: p = p_0$, the distribution of $\hat{p}$ is approximately normal with mean p_0 and standard deviation $\sqrt{\frac{p_0(1-p_0)}{n}}$.

Therefore, the test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$$

which has an approximately standard normal distribution under H_0.

How to conduct hypothesis test for *p*

- Write H_0 and H_1.
- Compute test statistic z_0.
- Compute *P*-value:
    - $P(z \geq z_0)$ if $H_1 : p \geq p_0$
    - $P(z \leq z_0)$ if $H_1 : p \leq p_0$
    - $2P(z \leq -|z_0|)$ if $H_1 : p \neq p_0$.
- Compare *P*-value to significance level α.
- State decision and draw conclusion.

Example. In a random sample of 100 voters, 63% said that they vote in favor of a proposal. Can we claim at the 1% significance level that majority of votes is in favor of the proposal?

Solution. We test $H_0 : p = 0.5$ against $H_1 : p > 0.5.$ It is given that $n = 100, \hat{p} = 0.63,$ and $\alpha = 0.01.$ The test statistic is

$$z = \frac{0.63 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = 2.6.$$

P-value $= P(z > 2.6) = P(z < -2.6)$

$= 0.0047 < 0.01 = \alpha$, thus, we reject H_0, and conclude that majority of votes is in favor of the proposal.

Example. To check whether a machine is producing fewer than 4% defective items, a random sample of size 200 is drawn, and 6 items are found to be defective. Test the hypothesis.

<u>Solution</u>. It is given that $n = 200$ and $\hat{p} = 6/200 = 0.03$. To test $H_0 : p = 0.04$ against $H_1 : p < 0.04$, we compute the test statistic

$$z = \frac{0.03 - 0.04}{\sqrt{\frac{0.04(1-0.04)}{200}}} = -0.72.$$

The *P*-value = $P(z < -0.72) = 0.2358 > 0.05 = \alpha$.

Hence, we accept H_0, and conclude that the machine is <u>not</u> producing fewer than 4% defectives.

Example. LBTA claims that 80% of bus trips are as-scheduled. A random sample of 60 buses reveals that 68% of the buses were on time at a certain check point. Verify the claim.

Solution. We would like to test $H_0: p = 0.8$ against $H_1: p \neq 0.8$. We are given that $n = 60$, and $\hat{p} = 0.68$. The test statistic is

$$z = \frac{0.68 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{60}}} = -2.32.$$

P-value $= 2P(z < -2.32) = (2)(0.0102)$ $= 0.0204 < 0.05 = \alpha$, thus, we reject H_0, and conclude that proportion of buses that run as scheduled is different from 80%.

10.5 Hypotheses Testing About Two Population Proportions: Large Samples

Suppose we draw two samples of sizes n_1 and n_2 from two populations. Let x_1 and x_2 denote the number of objects of interest in these samples, and let $\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$ be the two sample proportions.

Denote by p_1 and p_2 the true unknown population proportions. Suppose we want to test $H_0 : p_1 = p_2$ against $H_1 : p_1 \underset{>}{\geq} p_2$ or $H_1 : p_1 \underset{<}{\leq} p_2$ or $H_1 : p_1 \neq p_2$.

Under H_0, both population proportions are equal. To estimate this common proportion $p = p_1 = p_2$, we pool the samples to obtain the **pooled estimate**

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

By the CLT, $\hat{p}_1$ is approximately normally distributed with mean p and variance $\frac{p(1-p)}{n_1}$. Likewise, $\hat{p}_2$ is approximately normal with mean p and variance $\frac{p(1-p)}{n_2}$.

Thus, the difference $\hat{p}_1 - \hat{p}_2$ is approximately normal with mean $p - p = 0$ and variance

$$\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2} = p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Therefore, the test statistic is

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

which under H_0 is approximately a standard normal random variable.

Example. To test whether vitamin C is a preventive measure for common cold, 500 people took vitamin C, and 500 people took a sugar pill. In the first sample, 200 people had cold, while in the second sample, 230 had cold. Test the claim at the 1% significance level.

Solution. Given $n_1 = n_2 = 500, x_1 = 200, x_2 = 230,$ and $\alpha = 0.01.$ We want to test $H_0 : p_1 = p_2$ against $H_1 : p_1 < p_2.$

We estimate

$$\hat{p}_1 = \frac{200}{500} = 0.4,\ \hat{p}_2 = \frac{230}{500} = 0.46,\ \text{and}\ \ \hat{p} = \frac{200+230}{500+500} = 0.43.$$

The test statistic is

$$z = \frac{0.4 - 0.46}{\sqrt{(0.43)(1-0.43)\left(\frac{1}{500} + \frac{1}{500}\right)}} = -1.92.$$

The P-value $= P(z < -1.92) = 0.0274 > 0.01 = \alpha$,

hence, we accept H_0, and conclude that vitamin C is <u>not</u> a preventive measure for common cold.

Example. In a sample of 50 freshmen, 12% have a job, while in a sample of 70 seniors, 20% have a job. Test at the 5% level of significance whether the true population proportions are the same or not.

Solution. Given $n_1 = 50, n_2 = 70, \hat{p}_1 = 0.12,$ $\hat{p}_2 = 0.2,$ and $\alpha = 0.05.$ We want to test $H_0 : p_1 = p_2$ against $H_1 : p_1 \neq p_2.$

We compute the pooled estimate of the proportion as follows:

$x_1 = n_1\hat{p}_1 = (50)(0.12) = 6,\; x_2 = n_2\hat{p}_2 = (70)(0.2) = 14,$

and

$$\hat{p} = \frac{6+14}{50+70} = \frac{20}{120} = 0.167.$$

The test statistic is

$$z = \frac{0.12 - 0.2}{\sqrt{(0.167)(1 - 0.167)\left(\frac{1}{50} + \frac{1}{70}\right)}} = -1.16.$$

The P-value = $2P(z < -1.16) = (2)(0.1230)$

$= 0.2460 > 0.05 = \alpha.$

Thus, we accept H_0, and conclude that the true population proportions are the same.

Exercise. Managers of a large electric utility company claim that at least 70% of their customers are very satisfied with the service they receive. To test this claim, the local newspaper surveyed 100 customers, using simple random sampling. Among the sampled customers, 73 percent say they are very satisfied. Based on these findings, can we reject the managers' claim? Use a 0.05 level of significance.

Solution. We test $H_0: p = 0.7$ against $H_1: p \geq 0.7$. It is given that $n = 100, \hat{p} = 0.73$, and $\alpha = 0.05$. The test statistic is

$$z = \frac{0.73 - 0.7}{\sqrt{\dfrac{0.7(1-0.7)}{100}}} = 0.65.$$

P-value = $P(z > 0.65) = P(z < -0.65) = 0.2678$ $> 0.05 = \alpha$, hence, we accept H_0, and conclude that the claim is false.

Exercise. To compare the efficacy of two insect sprays, A and B, an experiment is conducted. Two rooms of equal size are sprayed with the same amount of spray, one room with A, the other with B. Two hundred insects are released into each room, and after one hour the numbers of dead insects are counted. There are 120 dead insects in the room sprayed with A and 90 in the room sprayed with B. Do the data provide enough evidence to indicate that spray A is more efficient than spray B? Use $\alpha = 0.01$.

Solution. Given $n_A = n_B = 200,\ \hat{p}_A = \dfrac{120}{200} = 0.6,$ $\hat{p}_B = \dfrac{90}{200} = 0.45,$ and $\alpha = 0.05.$ We test $H_0\colon\ p_A = p_B$ against $H_1\colon p_A > p_B.$

The pooled estimate of the proportion is

$$\hat{p} = \frac{120+90}{200+200} = \frac{210}{400} = 0.525.$$

The test statistic is

$$z = \frac{0.6 - 0.45}{\sqrt{(0.525)(1-0.525)\left(\frac{1}{200}+\frac{1}{200}\right)}} = 3.00.$$

The P-value = $P(z > 3.00) = P(z < -3.00)$

$$= 0.0014 < 0.01 = \alpha.$$

Thus, we reject H_0, and conclude that spray A is more efficient than spray B.