STP 226
ELEMENTARY STATISTICS

NOTES

PART 1V INFERENTIAL STATISTICS

CHAPTER 11

INFERENCES FOR POPULATION PROPORTIONS

Population Proportion and Sample Proportion

Consider a population in which each member either has or does not have a specified attribute. Then we use the following notation and terminology.

Population proportion, p : The proportion (percentage) of the entire population that has the specified attribute. (p – unknown but fixed)

Sample proportion , $\hat{p}$ (p hat) : The proportion (percentage) of a sample from the population that has the specified attribute. ($\hat{p}$ - variable and varies from sample to sample)

Where $\hat{p} = \frac{x}{n}$, x is the number of members in the sample that have the specified attribute and n is the sample size

eg. One sample may have 5 females out of 20 people where another sample may have 7 out of 20 which makes $\hat{p}$ different for each sample

$\hat{p} = \frac{x}{n} = \frac{5}{20} = 0.25\,(25\%)$ for the first sample and $\hat{p} = \frac{x}{n} = \frac{7}{20} = 0.35\,(35\%)$ for the next sample

x - number of successes (number in the sample that have the specified attribute)

n – x – number of failures (number in sample that do not have the specified attribute)

The Sampling Distribution of the Proportion

For samples of size n,

1. The mean of $\hat{p}$ equals the population proportion: $\mu_{\hat{p}} = p$

2. The standard deviation of $\hat{p}$ equals the square root of the product of the population proportion and one minus the population proportion divided by the sample size:

3. $\hat{p}$ is approximately normally distributed for large n.

Large-Sample Confidence Interval for a Population Proportion

The one-sample z-interval procedure for a population proportion

ASSUMPTIONS

The number of successes, x, and the number of failures, n – x, are both 5 or greater.

STEPS

1. For a confidence level of 1 - α, use Table II to find $z_{\alpha/2}$

2. The confidence interval for p is from

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n} \quad \text{to} \quad \hat{p} + z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$$

where $z_{\alpha/2}$ is found in step 1, n is the sample size, and $\hat{p} = \frac{x}{n}$ is the sample proportion.

Margin of Error for the Estimate of p

The margin of error for the estimate p is $E = z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$

Solving the margin of error formula $E = z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$ for n gives sample size formula

$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$.

Sample size for estimating p

A (1 - α)-level confidence interval for a population proportion having a margin of error of at most E can be obtained by choosing

$n = 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2$ rounded up to the nearest whole number. If we can make an educated guess, $\hat{p}_g$ (g for guess), for the observed value of $\hat{p}$, then we should choose

$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$, rounded up to the nearest whole number.

11.2 Hypotheses tests for One Population Proportion

For a large n, the standardized version of $\hat{p}$, $z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$ has approximately the standard normal distribution

The One-Sample z-test for a Population Proportion

ASSUMPTION

Both np_0 and $n(1-p_0)$ are 5 or greater.

STEPS

1. The null hypothesis is H_0: $p = p_0$ and the alternative hypothesis is one of the following:

$H_a : p \neq p_0$	Two-tailed
$H_a : p < p_0$	Left-tailed
$H_a : p > p_0$	Right-tailed

2. Decide on the significance level, α.

3. The critical value(s) are

$\pm z_{\alpha/2}$	Two-tailed
$-z_{\alpha}$	Left-tailed
z_{α}	Right-tailed

   Use Table II to find the critical value(s)

4. Compute the value of the test statistic, $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

5. If the value of the test statistic falls in the rejection region, reject H_0 otherwise, do not reject H_0.

6. State the conclusion in words.

11.3 Inferences for Two Population Proportions Using Independent Samples

The Sampling Distribution of the Difference Between Two Population Proportions for Independent Samples

For independent samples of sizes n_1 and n_2 from two populations,

1. $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$

2. $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$, and

3. $\hat{p}_1 - \hat{p}_2$ is approximately normally distributed for large n_1 and n_2.

In particular, for large samples, the possible differences between the two sample proportions have approximately a normal distribution with mean $p_1 - p_2$ and standard deviation $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$.

From $p = \frac{x}{n}$ we can get our estimates as follows

$\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$ for the two samples respectively

Large-Sample Hypothesis Tests for Two Population Proportions Using Independent Samples

For large samples, $z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}}$ is approximately the standard normal distribution.

To test H_0: $p_1 = p_2$ the standardized test statistic becomes $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)/n_1 + p(1-p)/n_2}}$

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

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

the estimate of p is $\hat{p}_p = \dfrac{x_1 + x_2}{n_1 + n_2}$ (the pooled sample proportion)

The Two-Sample z-test for Two Population Proportions

ASSUMPTIONS

1. Independent samples

2. x_1, $n_1 - x_1$, x_2, and $n_2 - x_2$ are all 5 or greater.

STEPS

1. The null hypothesis is H_0: $p_1 = p_2$ and the alternative hypothesis is one of the following:

   $H_a : p_1 \neq p_2$ Two-tailed
   $H_a : p_1 < p_2$ Left-tailed
   $H_a : p_1 > p_2$ Right-tailed

2. Decide on the significance level, α.

3. The critical value(s) are

   $\pm z_{\alpha/2}$ Two-tailed
   $-z_{\alpha}$ Left-tailed
   z_{α} Right-tailed

   Use Table II to find the critical value(s)

4. Compute the value of the test statistic, $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)}\sqrt{(1/n_1 + 1/n_2)}}$

5. If the value of the test statistic falls in the rejection region, reject H_0 otherwise, do not reject H_0.

6. State the conclusion in words.

The p-value approach compares the p-value for the test statistic with the α level (the p-value of the hypothesis test)

The Two-Sample z-interval Procedure for Two Population Proportions

ASSUMPTIONS

1. Independent samples

2. x_1, $n_1 - x_1$, x_2, and $n_2 - x_2$ are all 5 or greater.

STEPS

1. For a confidence level of 1 - α, use Table II to find $z_{\alpha/2}$.

2. The endpoints of the confidence interval for $p_1 - p_2$ are

$$(\hat{p}_1 - \hat{p}_2) \pm \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$$

Margin of Error and Sample Size for Estimating $p_1 - p_2$

The margin of error for the estimate of $p_1 - p_2$ is

$$E = z_{\alpha/2} \cdot \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \ .$$

It equals half the length of the confidence interval and represents the precision with which the difference the sample proportions, $\hat{p}_1 - \hat{p}_2$, estimates the difference between the population proportions, $p_1 - p_2$, at the specified confidence level.

A (1 - α)-level confidence interval for the difference between two population proportions having a margin of error of at most E can be obtained by choosing

$n_1 = n_2 = 0.5\left(\frac{z_{\alpha/2}}{E}\right)^2$ rounded up to the nearest whole number. If we can make educated guesses, $\hat{p}_{1g}$ and $\hat{p}_{2g}$, for the observed values of $\hat{p}_1$ and $\hat{p}_2$, then we should instead choose

$$n_1 = n_2 = (\hat{p}_{1g}(1-\hat{p}_{1g}) + \hat{p}_{2g}(1-\hat{p}_{2g}))\left(\frac{z_{\alpha/2}}{E}\right)^2$$

rounded up to the nearest whole number.