<u>Two Populations Hypothesis Testing</u>

<u>Two Proportions (Large Independent Samples)</u>

Two samples are said to be <u>independent</u> if the data from the first sample is not connected to the data from the second sample. If the two data sets are connected, then the samples are said to be <u>dependent.</u> Dependent samples are also referred to as <u>paired samples</u> or <u>matched samples.</u>

$$\widehat{p_1} = \frac{x_1}{n_1}$$

$$\widehat{p_2} = \frac{x_2}{n_2}$$

To compare the two populations, we use the difference between the two sample proportions (**<u>point estimate)</u>**:

$\widehat{p_1} - \widehat{p_2}$

1. The two samples must be independent
2. The samples must be large enough to use the normal distribution. The products

$n_1\widehat{p_1} \geq 5$

$n_1(1 - \widehat{p_1}) \geq 5$

$n_2\widehat{p_2} \geq 5$

$n_2(1 - \widehat{p_2}) \geq 5$

p_1=population 1 proportion p_2=population 2 proportion n_1=sample size of population 1 n_2=sample size of population 2 x_1=number of successes in population 1 x_2=number of successes in population 2 $\widehat{p_1}$=sample proportion of population 1 $\widehat{p_2}$=sample proportion of population 2

If the above conditions are met, the sampling distribution for $\widehat{p_1} - \widehat{p_2}$ (the difference between the samples proportions) has a normal distribution with mean:

$\mu_{\widehat{p_1} - \widehat{p_2}} = p_1 - p_2$

<u>Standard Error</u>

$$\sigma_{\widehat{p_1} - \widehat{p_2}} = \sqrt{\frac{\widehat{p_1}(1 - \widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1 - \widehat{p_2})}{n_2}}$$

Created by: Megan Krou
Spring 2017
STUDENT LEARNING ASSISTANCE CENTER (SLAC)
Texas State University

Confidence Interval

$(\widehat{p_1} - \widehat{p_2}) \pm \left(z_{\frac{\alpha}{2}}\right)\left(\sigma_{\widehat{p_1}-\widehat{p_2}}\right)$

*Note: $\widehat{p_1}$ and $\widehat{p_2}$ estimates are used when p_1 and p_2 are unknown.

Standardized test statistic

$$Z = \frac{(\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

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

$$\bar{p} = \frac{n_1\widehat{p_1} + n_2\widehat{p_2}}{n_1 + n_2}$$

Determine the decision rule: Reject Ho if z <_____

If left tail =normsinv(α)

Determine the decision rule: Reject Ho if z >_____

If right tail =normsinv(1-α)

Determine the decision rule: Reject Ho if $|z|$>______

If two tail test =normsinv$\left(1 - \frac{\alpha}{2}\right)$

Two Means (Sigma Known)

Bounds

$\bar{x} \pm \left(z_{\frac{\alpha}{2}}\right)\left(\frac{\sigma}{\sqrt{n}}\right)$

$\overline{x_1} - \overline{x_2} \pm \left(z_{\frac{\alpha}{2}}\right)\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}$

$\overline{x_1} - \overline{x_2}$ is the "point estimator"

$\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}$ is the standard deviation

$\overline{x_1}$=sample mean of sample 1
$\overline{x_2}$=sample mean of sample 2
σ_1=population standard deviation of population 1
σ_2=population standard deviation of population 2
$(\sigma_1)^2$=population variance of population 1
$(\sigma_2)^2$=population variance of population 2

Have to be large enough: $n_1 \geq 30$ *and* $n_2 \geq 30$

$$Z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}}$$

Two Means (Sigma Unknown)

Use t-distribution when population standard deviation is not known

$$\overline{x_1} - \overline{x_2} \pm \left(t_{\frac{\alpha}{2}}\right)\left(\sigma_{\overline{x_1}-\overline{x_2}}\right)$$

$\overline{x_1} - \overline{x_2}$ is the "point estimator"

Standard error when the population variances are equal

$$\sigma_{\overline{x_1}-\overline{x_2}} = S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Degrees of freedom $n_1 + n_2 - 2$

s_1=sample standard deviation of sample 1
s_2=sample standard deviation of sample 2
$(s_1)^2$=sample variance of sample 1
$(s_2)^2$=sample variance of sample 2

Standard error when the population variances are not equal

$$\sigma_{\overline{x_1}-\overline{x_2}} = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

Degrees of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$

Pooled sample standard deviation

$$S_P = \sqrt{\frac{(n_1 - 1)(s_1)^2 + (n_2 - 1)(s_2)^2}{n_1 + n_2 - 2}}$$

Common population variance

$$S_P{}^2 = \frac{(n_1 - 1)(s_1)^2 + (n_2 - 1)(s_2)^2}{n_1 + n_2 - 2}$$

Find $t_{\frac{\alpha}{2}}$ by looking up in table depending on the degrees of freedom and $\frac{\alpha}{2}$

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sigma_{\overline{x_1} - \overline{x_2}}}$$

$*\mu_1 - \mu_2 = 0$

Decision Rule: Reject Ho if t > ____
If right tail: look up α and degrees of freedom

Reject Ho if $|t|$ > ____
If two tail: look up degrees of freedom and $\frac{\alpha}{2}$

Two Means (Dependent Samples)

When dependent samples are involved, the data is thought of as paired data.
Difference between the pairs of data values is referred to as a paired difference.

Paired difference
$d = x_1 - x_2$

The difference between two population means, when dependent samples are used, is equivalent to the mean of the paired differences.

If the following two conditions are met, the sampling distribution for $\bar{d}$, the mean of the differences of the paired data entries in the dependent samples, has a t-distribution with n-1 degrees of freedom, where n is the number of data pairs.

1. the two samples must be dependent (paired)
2. each population has a normal distribution

Mean of the population
$\mu_d = \mu_1 - \mu_2$

Point estimate/Mean of the paired differences
$$\bar{d} = \frac{\sum d}{n}$$

n = number of data pairs

Bounds

$$\bar{d} \pm \left(t_{\frac{\alpha}{2}}\right)\left(\frac{s_d}{\sqrt{n}}\right)$$

Sample standard deviation

$$s_d = \sqrt{\frac{n(\sum d^2) - (\sum d)^2}{n(n-1)}}$$

Test statistic

$$t = \frac{\bar{d} - \mu_d}{\left(\frac{s_d}{\sqrt{n}}\right)}$$

n – 1= degrees of freedom

Decision Rule: Reject Ho if t > ____
If right tail: look up n-1 degrees of freedom and α

Decision Rule: Reject Ho if t < ____
If left tail: look up n-1 degrees of freedom and α and make it negative

Decision Rule: Reject Ho if $|t|$ > ____
If two tail: look up n-1 degrees of freedom and $\frac{\alpha}{2}$

Two Population Variances

In order to compare two population variances, must ensure that the following two conditions are met:
1. the two populations must be independent, not matched or paired in any way, and
2. the two populations must be normally distributed

F-distribution

- Skewed to the right
- Values of F are always greater than 0
- Shape is completely determined by its two parameters, the degrees of freedom of the numerator and the degrees of freedom of the dominator of the ratio

Degrees of freedom of numerator = $n_1 - 1$
Degrees of freedom of denominator = $n_2 - 1$

Ho/Ha: ${\sigma_1}^2$

$$F = \frac{{s_1}^2}{{s_2}^2}$$

If left tail: Decision Rule: Reject Ho if F ≤ $F_{1-\alpha}$

If right tail: Decision Rule: Reject Ho if F ≥ F_{α}

If two tail:
Decision Rule: Reject Ho if F ≤ $F_{1-\frac{\alpha}{2}}$

OR

Reject Ho if F ≥ $F_{\frac{\alpha}{2}}$

n_1=sample size of sample 1
n_2=sample size of sample 2
σ_1=population standard deviation of population 1
σ_2=population standard deviation of population 2
$(\sigma_1)^2$=population variance of population 1
$(\sigma_2)^2$=population variance of population 2
s_1=sample standard deviation of sample 1
s_2=sample standard deviation of sample 2
${s_1}^2$=sample variance of sample 1
${s_2}^2$=sample variance of sample 2

Created by: Megan Krou
Spring 2017
STUDENT LEARNING ASSISTANCE CENTER (SLAC)
Texas State University