Comparing Two Populations

Comparing two proportions

Confidence interval

- Confidence Interval for the difference between two population proportions p_1 and p_2 at a specified confidence level is:

$$CI = (\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$\hat{p}_1, \hat{p}_2$ - Proportions of samples 1 & 2; z^* - Critical value for the CI;
n_1, n_2 - Size of samples 1 & 2

Z^* (Z_{crit}) Values

Confidence	Z^*
90%	1.645
95%	1.96
98%	2.33
99%	2.58

Significance Test

As with one-sample tests, calculate a P-value from the z statistic.

Null and alternative hypotheses

- H_0: $p_1 - p_2 = 0$
- H_a: $p_1 > p_2$
- H_a: $p_1 < p_2$
- H_a: $p_1 \neq p_2$

Validity requirements

These equations are valid if:
- Random data
- 10% rule
  $n_{1,2} \leq 0.1 N_{1,2}$
- Large counts
  $n \cdot \hat{p} \geq 10$
  $n(1-\hat{p}) \geq 10$

Z-Statistic

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

$\hat{p}_1, \hat{p}_2$ - Proportions of samples 1 & 2; p_c - Pooled proportion
n_1, n_2 - Size of samples 1 & 2

Obtaining P-value from z score

Upper-tailed test	H_a: $p_1 - p_2 > 0$	P-value = $1 - P(z)$
Lower-tailed test	H_a: $p_1 - p_2 < 0$	P-value = $P(z)$
Two-tailed test	H_a: $p_1 - p_2 \neq 0$	if $z > 0$, P-value = $2(1 - P(z))$ if $z < 0$, P-value = $2P(z)$

p_1, p_2 - Proportions of populations 1 & 2;
z - Z-score/test statistic; $P(z)$ - P-value from z table

Calculator Note

Population proportions
- *2-PropZInt* Interval
- *2-PropZTest* Significance

Population means
- *2-SampTInt* Interval
- *2-SampTTest* Significance

Comparing Two Means, μ_1 and μ_2

Confidence Interval

- Standard Deviation for the difference between two populations whose stddev's are known:

$$\sigma = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

For t^*, use the smaller degrees of freedom of n_1 or n_2

- Confidence Interval for the difference between two population means μ_1 and μ_2 at a given confidence level is:

$$CI = (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

For t^*, use the smaller degrees of freedom of n_1 or n_2

Validity requirements

These equations are valid if:

- Random data
- 10% rule
  $n_{1,2} \leq 0.1\, N_{1,2}$
- Large counts
  $n_1 \geq 30$
  $n_2 \geq 30$

Significance Test

T-Statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\mu_1 - \mu_2$ will usually be zero

x_1, x_2 - Means of samples; **s_1, s_2** - Std. dev. of samples
$\mu_1 - \mu_2$ - Hypothesized difference in population means
n_1, n_2 - Size of samples

Calculator Note

Population proportions

- *2-PropZInt* Interval
- *2-PropZTest* Significance

Population means

- *2-SampTInt* Interval
- *2-SampTTest* Significance