Confidence Intervals and Tests of Significance

1 Confidence Intervals

The confidence interval is an extension of the idea of a point estimation of the parameter to an interval that is likely to contain the true parameter value.

A level C confidence interval for a population parameter θ is an interval computed from the sample data having probability C of producing an interval containing θ.

For an estimate of a population mean or proportion, a level C confidence interval often has the form

$$\text{estimate} \pm t^* \times \text{standard error}$$

where t^* is the upper $\frac{1-C}{2}$ critical value for the t distribution with the appropriate number of degrees of freedom. If the number of degrees of freedom is infinite, we use the normal distribution to detemine the critical value, usually denoted by z^*.

The margin of error $\quad m = t^* \times \text{standard error}$ decreases if

- C decreases
- the standard deviation decreases
- n increases

The procedures for finding the confidence interval are summarized in the table below.

procedure	parameter	estimate	standard error	degrees of freedom
one sample	μ	$\bar{x}$	$\frac{s}{\sqrt{n}}$	$n-1$
two sample	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\min(n_1, n_2) - 1$
pooled two sample	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$n_1 + n_2 - 2$
one proportion	p	$\hat{p}$	$\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}},\ \tilde{p} = \frac{x+2}{n+4}$	∞
two proportion	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	∞

The first confidence interval for $\mu_1 - \mu_2$ is the two-sample t procedure. If we can assume that the two samples have a common standard deviation, then we pool the data to compute s_p, the pooled standard deviation. Matched pair procedures use a one sample procedure on the difference in the observed values.

For these tests, we need a sample size large enough so that the central limit theorem is a sufficiently good approximation. For one population tests for means, $n > 30$ and data not strongly skewed is a good rule of thumb. For two population tests, $n > 40$ may be necessary. For population proportions, we ask that the mean number of successes and the mean number of failures each be at least 10.

2 Tests of Significance

2.1 Test for Population Means

- Hypotheses are stated in terms of *population parameter*.
- The null hypothesis H_0 is a statement that no effect is present.
- The alternative hypothesis H_1 is a statement that a parameter differs from its null value in a specific direction (one-sided alternative) or in either direction (two-sided alternative).
- A test statistic is designed to assess the strength of evidence against H_0.
- If a decision must be made, specify the significance level α.
- Assuming H_0 is true, the p-value is the probability that the test statistic would take a value as extreme or more extreme than the value observed.
- If the p-value is smaller than the significance level α, then the data are statistically significant at level α.

t-procedure	null hypothesis
single sample	$H_0 : \mu = \mu_0$
two samples	$H_0 : \mu_1 = \mu_2$

The test statistic

$$t = \frac{\text{estimate} - \text{parameter}}{\text{standard error}}.$$

The P-value is determined by the distribution of a random variable having a t distribution with the appropriate number of degrees of freedom. The appropriate estimates, standard errors, and degrees of freedom are stated on the previous page.

2.2 Test for Population Proportions

The design is based on Bernoulli trials. This is considered valid is the sample is small ($< 10\%$) compared to the total population and np_0 is at least 10.

sample proportions	null hypothesis	test statistic
single proportion	$H_0 : p = p_0$	$z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$
two proportions	$H_0 : p_1 = p_2$	$z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$

The pooled sample proportion $\hat{p} = (X_1 + X_2)/(n_1 + n_2)$.