Everything you've always wanted to know about confidence intervals

(CI = (estimator) ± (margin of error)
= (estimator) ± (critical value)(standard error))

Parameter	Estimator (Statistic)	Sampling Dist. Std. Dev.	Sampling Dist. Std. Error	Conditions for use	Critical Value	Confidence Interval
μ	$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	$s_{\bar{x}} = \frac{s}{\sqrt{n}}$	Large random sample: *($n \geq 30$)*	$z_{\alpha/2}$	$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$
				Small random sample: *($n < 30$, parent population approximately normal)*	$t_{\alpha/2}$	$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$, $d.f. = n-1$
p	$\hat{p} = \frac{X}{n}$	$\sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Random sample $n\hat{p} \geq 5, n(1-\hat{p}) \geq 5$	$z_{\alpha/2}$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$\mu_1 - \mu_2$ *(independent samples)*	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Large random sample ($n_1 \geq 30$ *and* $n_2 \geq 30$)	$z_{\alpha/2}$	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
			$s_{\bar{x}_1-\bar{x}_2} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$	Small sample ($\sigma_1^2 = \sigma_2^2, n_1 < 30$ *or* $n_2 < 30$, *two random samples drawn from independent, approximately normal populations)*	$t_{\alpha/2}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, $d.f. = n_1 + n_2 - 2$
			$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Small sample ($\sigma_1^2 \neq \sigma_2^2$, *two random samples drawn from independent, approximately normal populations)*	$t_{\alpha/2}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $d.f. = n_1 + n_2 - 2$ $(n_1 = n_2)$, $d.f. =$ (software) $(n_1 \neq n_2)$, or $\min\{n_1 - 1, n_2 - 1\}$.
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	Random samples $n_1\hat{p}_1 \geq 5, n_1(1-\hat{p}_1) \geq 5$ $n_2\hat{p}_2 \geq 5, n_2(1-\hat{p}_2) \geq 5$	$z_{\alpha/2}$	$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
β_1 *(slope of regression line)*	$\hat{\beta}_1$	$\sigma_{\hat{\beta}_1}$	$s_{\hat{\beta}_1}$ (from computer printout – shows as "Stdev" for predictor variable)	For each x, the corresponding values of y are normally distributed and all have the same standard deviation. The mean values of the y's lie on a line.	$t_{\alpha/2}$	$\hat{\beta}_1 \pm t_{\alpha/2}\, s_{\hat{\beta}_1}$, $d.f. = n - 2$

Note: ***Paired samples*** *(dependent samples) are a special case of one-sample statistics*

Everything you've always wanted to know about hypothesis testing

(Test Statistic* = $\frac{(estimator) - hypothesized\ value}{standard\ error}$)

Null Hypothesis	Estimator (Statistic)	Sampling Dist. Std. Dev.	Sampling Dist. Std. Error	Conditions for use	Test Statistic
$H_o: \mu = \mu_o$	$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	$s_{\bar{x}} = \frac{s}{\sqrt{n}}$	Large random sample: *($n \geq 30$)*	$z^* = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}$
Note: ***Paired samples*** *(dependent samples) are a special case of one-sample statistics ($H_o: \mu_d = 0$)*				Small random sample: *($n < 30$, population approximately normal)*	$t^* = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}$ $d.f. = n-1$
$H_o: p = p_o$	$\hat{p} = \frac{X}{n}$	$\sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{p_o(1-p_o)}{n}}$	Random Sample $np_o \geq 5, n(1-p_o) \geq 5$	$z^* = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$
$H_o: \mu_1 - \mu_2 = 0$ *or* $H_o: \mu_1 = \mu_2$ *(independent samples)*	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Large random samples ($n_1 \geq 30$ *and* $n_2 \geq 30$)	$z^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
			$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	Small samples *($\sigma_1^2 = \sigma_2^2$, $n_1 < 30$ or $n_2 < 30$, two random samples from independent approximately normal populations)*	$t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, $d.f. = n_1 + n_2 - 2$
			$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Small sample *($\sigma_1^2 \neq \sigma_2^2$, two random samples from approximately normal populations)*	$t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $d.f. = n_1 + n_2 - 2 \quad (n_1 = n_2)$, $d.f. = (software) \quad (n_1 \neq n_2)$ or $min\{n_1 - 1, n_2 - 1\}$
$H_o: p_1 - p_2 = 0$ *or* $H_o: p_1 = p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, *where* $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$	$n_1\hat{p}_1 \geq 5, n_1(1-\hat{p}_1) \geq 5$ $n_2\hat{p}_2 \geq 5, n_2(1-\hat{p}_2) \geq 5$	$z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
$H_o: \beta_1 = 0$	$\hat{\beta}_1$ (OR: *b*)	$\sigma_{\hat{\beta}_1}$	$s_{\hat{\beta}_1}$ (from computer printout – shows as "Stdev" for predictor variable)	For each *x*, the corresponding values of *y* are normally distributed and all have the same standard deviation. The mean values of the *y's* lie on a line.	$t^* = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$, *d.f.* $= n - 2$

From STEM center

IMPORTANT FORMULAS

Chapter 3: Numerical Summaries of Data

Sample mean:

$$\bar{x} = \frac{\sum x}{n}$$

Population mean:

$$\mu = \frac{\sum x}{N}$$

Range:

Range = largest value − smallest value

Population variance:

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

Sample variance:

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

Coefficient of variation:

$$\text{CV} = \frac{\sigma}{\mu}$$

z-score:

$$z = \frac{x-\mu}{\sigma}$$

Interquartile range:

$\text{IQR} = Q_3 - Q_1$ = third quartile − first quartile

Lower outlier boundary:

$Q_1 - 1.5\,\text{IQR}$

Upper outlier boundary:

$Q_3 + 1.5\,\text{IQR}$

Chapter 4: Summarizing Bivariate Data

Correlation coefficient:

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$$

Slope of least-squares regression line:

$$b_1 = r\frac{s_y}{s_x}$$

y-intercept of least-squares regression line:

$$b_0 = \bar{y} - b_1\bar{x}$$

Equation of least-squares regression line:

$$\hat{y} = b_0 + b_1 x$$

Chapter 5: Probability

General Addition Rule:

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Multiplication Rule for Independent Events:

$P(A \text{ and } B) = P(A)P(B)$

Addition Rule for Mutually Exclusive Events:

$P(A \text{ or } B) = P(A) + P(B)$

Rule of Complements:

$P(A^c) = 1 - P(A)$

General Method for Computing Conditional Probability:

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

General Multiplication Rule:

$P(A \text{ and } B) = P(A)P(B \mid A) = P(B)P(A \mid B)$

Permutation of r items chosen from n:

$${}_nP_r = \frac{n!}{(n-r)!}$$

Combination of r items chosen from n:

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

Chapter 6: Discrete Probability Distributions

Mean of a discrete random variable:

$\mu_X = \sum[x \cdot P(x)]$

Variance of a discrete random variable:

$\sigma_X^2 = \sum[(x - \mu_X)^2 \cdot P(x)] = \sum[x^2 \cdot P(x)] - \mu_X^2$

Standard deviation of a discrete random variable:

$\sigma_X = \sqrt{\sigma_X^2}$

Mean of a binomial random variable:

$\mu_X = np$

Variance of a binomial random variable:

$\sigma_X^2 = np(1 - p)$

Standard deviation of a binomial random variable:

$\sigma_X = \sqrt{np(1 - p)}$

Mean of Poisson random variable:

$\mu_X = \lambda t$

Variance of Poisson random variable:

$\sigma_X^2 = \lambda t$

Standard deviation of Poisson random variable:

$\sigma_X = \sqrt{\lambda t}$

Chapter 7: The Normal Distribution

z-score:

$$z = \frac{x - \mu}{\sigma}$$

Convert z-score to raw score:

$x = \mu + z\sigma$

Standard deviation of the sample mean:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

z-score for a sample mean:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Standard deviation of the sample proportion:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}}$$

z-score for a sample proportion:

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

Chapter 8: Confidence Intervals

Confidence interval for a mean, standard deviation known:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Sample size to construct an interval for μ with margin of error m:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{m}\right)^2$$

Confidence interval for a mean, standard deviation unknown:

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Confidence interval for a proportion:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Sample size to construct an interval for p with margin of error m:

$$n = \hat{p}(1 - \hat{p})\left(\frac{z_{\alpha/2}}{m}\right)^2$$ if a value for $\hat{p}$ is available

$$n = 0.25\left(\frac{z_{\alpha/2}}{m}\right)^2$$ if no value for $\hat{p}$ is available

Confidence interval for the variance of a normal distribution:

$$\frac{(n - 1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n - 1)s^2}{\chi^2_{1-\alpha/2}}$$

Confidence interval for the standard deviation of a normal distribution:

$$\sqrt{\frac{(n - 1)s^2}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{(n - 1)s^2}{\chi^2_{1-\alpha/2}}}$$

Chapter 9: Hypothesis Testing

Test statistic for a mean, standard deviation known:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Test statistic for a mean, standard deviation unknown:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Test statistic for a proportion:

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

Test statistic for a standard deviation:

$$\chi^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2}$$

Chapter 10: Two-Sample Confidence Intervals

Confidence interval for the difference between two means, independent samples:

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Confidence interval for the difference between two means, matched pairs:

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

Confidence interval for the difference between two proportions:

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

$$< p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

Chapter 11: Two-Sample Hypothesis Tests

Test statistic for the difference between two means, independent samples:

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

Test statistic for the difference between two means, matched pairs:

$$t = \frac{\bar{d} - \mu_0}{s_d/\sqrt{n}}$$

Test statistic for the difference between two proportions:

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

where $\hat{p}$ is the pooled proportion $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

Test statistic for two standard deviations:

$$F = \frac{\text{Larger of } s_1^2 \text{ and } s_2^2}{\text{Smaller of } s_1^2 \text{ and } s_2^2}$$

Chapter 12: Tests with Qualitative Data

Chi-square statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Expected frequency for goodness-of-fit:

$$E = np$$

Expected frequency for independence or homogeneity:

$$E = \frac{\text{Row total} \cdot \text{Column total}}{\text{Grand total}}$$

Chapter 13: Inference in Linear Models

Residual standard deviation:

$$s_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}}$$

Confidence interval for the mean response:

$$\hat{y} \pm t_{\alpha/2} \cdot s_e\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

Standard error for b_1:

$$s_b = \frac{s_e}{\sqrt{\sum(x-\bar{x})^2}}$$

Test statistic for slope b_1:

$$t = \frac{b_1}{s_b}$$

Confidence interval for slope:

$$b_1 - t_{\alpha/2} \cdot s_b < \beta_1 < b_1 + t_{\alpha/2} \cdot s_b$$

Prediction interval for an individual response:

$$\hat{y} \pm t_{\alpha/2} \cdot s_e\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

Chapter 14: Analysis of Variance

Treatment sum of squares:

$$SSTr = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \cdots + n_I(\bar{x}_I - \bar{\bar{x}})^2$$

Error mean square:

$$MSE = \frac{SSE}{N-I}$$

Treatment mean square:

$$MSTr = \frac{SSTr}{I-1}$$

***F* statistic for one-way ANOVA:**

$$F = \frac{MSTr}{MSE}$$

Error sum of squares:

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2$$

Test statistic for Tukey–Kramer test:

$$q = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

Chapter 15: Nonparametric Statistics

Test statistic for the sign test:

$$z = \frac{x + 0.5 - n/2}{\sqrt{n}/2} \quad \text{if } n > 25$$

If $n \leq 25$, the test statistic is x, the number of times the less frequent sign occurs.

Mean of *S*, the sum of the ranks for the rank-sum test:

$$\mu_S = \frac{n_1(n_1 + n_2 + 1)}{2}$$

Standard deviation of *S*, the sum of the ranks for the rank-sum test:

$$\sigma_S = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

Test statistic for the rank-sum test:

$$z = \frac{S - \mu_S}{\sigma_S}$$

ISBN - 978-1-259-29255-2
MHID - 1-259-29255-X