Confidence Intervals for the Mean

Parameter	Assumptions	Formula
Mean μ	Data normally distributed or n is large ($n > 30$); σ^2 known	$\mu \in \left(\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$ with $(100-\alpha)\%$ confidence
	n is large ($n > 30$); σ^2 unknown	$\mu \in \left(\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$
	Data normally distributed and n is small; σ^2 unknown	$\bar{x} \pm t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$
Difference in means $\mu_X - \mu_Y$ Case of 2 independent distributions	Data normally distributed or large samples; σ_X^2, σ_Y^2 known	$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$
	Large samples; σ_X^2, σ_Y^2 unknown	$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$
	Data normally distributed and small samples; σ_X^2, σ_Y^2 unknown but $\sigma_X^2 = \sigma_Y^2$	$(\bar{x} - \bar{y}) \pm t_{\alpha/2}(n_X + n_Y - 2) \cdot \sqrt{s_p^2 \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}$ where estimate of the pooled variance is $s_p^2 = \frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X+n_Y-2}$
	Data normally distributed and small samples; σ_X^2, σ_Y^2 unknown and $\sigma_X^2 \neq \sigma_Y^2$	$(\bar{x} - \bar{y}) \pm t_{\alpha/2}(n_X + n_Y - 2) \cdot \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$

Notation:

$z_{\alpha/2}$ is the $\alpha/2$ critical value of the standard normal distribution

$t_{\alpha/2}(n-1)$ is the $\alpha/2$ critical value of the $t(n-1)$ distribution

μ: population mean; σ^2: population variance

n: sample size; $\bar{x}$: average of observed sample values; s^2: variance of observed sample values ($s^2 = 1/(n-1) \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2$)

Confidence Intervals for a Proportion		
Parameter	**Assumptions**	**Formula**
Proportion p	Data binomially distributed; n large, i.e., $np \geq 10$ and $n(1-p) \geq 10$	$p \in \left(\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ with $(100-\alpha)\%$ confidence
Difference in proportions $p_X - p_Y$ Case of 2 independent distributions	Data binomially distributed; large samples	$\hat{p}_X - \hat{p}_Y \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_X \cdot (1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y \cdot (1-\hat{p}_Y)}{n_Y}}$
Notation: p: proportion of successes in the population n: sample size; x: number of successes in the sample; $\hat{p} = \frac{x}{n}$: proportion of successes in the sample		

Hypothesis Testing for the Mean

1^{st} Formulate the hypotheses

2^{nd} Compute test statistic (under the null hypothesis)

3^{rd} Compute p-value (or compare test statistic with critical value) and draw a conclusion

	For μ with $X \sim N(\mu, \sigma)$ (*)	**For $\mu_X - \mu_Y$ with $X \sim N(\mu_X, \sigma_X)$ (*) and $Y \sim N(\mu_Y, \sigma_Y)$ (*)**
		(For two independent samples)
1^{st}	$H_0 : \mu = \mu_0$ $H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$ or $H_a : \mu \neq \mu_0$	$H_0 : \mu_X = \mu_Y$ $H_a : \mu_X > \mu_Y$ or $H_a : \mu_X < \mu_Y$ or $H_a : \mu_X \neq \mu_Y$
2^{nd}	Use $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ if σ known or n large (**) or $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ otherwise	Use $z = \frac{(\bar{x}_X-\bar{x}_Y)-(0)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}$ if σ known or n large (**) or $t = \frac{(\bar{x}_X-\bar{x}_Y)-(0)}{\sqrt{\frac{s_X^2}{n_X}+\frac{s_Y^2}{n_Y}}}$ otherwise (***)
3^{rd}	$P(Z \geq z)$ or $P(Z \leq z)$ or $2 \cdot P(Z \geq \lvert z \rvert)$ if σ known or $n > 40$ $P(T \geq t)$ or $P(T \leq t)$ or $2 \cdot P(T \geq \lvert t \rvert)$ otherwise where $Z \sim N(0,1)$ and $T \sim t_{(n-1)}$	$P(Z \geq z)$ or $P(Z \leq z)$ or $2 \cdot P(Z \geq \lvert z \rvert)$ if σ known or $n > 40$ $P(T \geq t)$ or $P(T \leq t)$ or $2 \cdot P(T \geq \lvert t \rvert)$ otherwise where $Z \sim N(0,1)$ and $T \sim t_{(n_X+n_Y-2)}$

Notes:

μ: population mean; σ^2: population variance

n: sample size; $\bar{x}$: average of observed sample values; s^2: variance of observed sample values ($s^2 = 1/(n-1) \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2$)

(*) or apx. Normal if X is not normal but n is large (by the CLT)

(**) If σ unknown and n is large, use s in place of σ

(***) If have reasons to believe $\sigma_X = \sigma_Y = \sigma$ then use $T = \frac{(\bar{X}_X-\bar{X}_Y)-(\mu_X-\mu_Y)}{s_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}} \sim t_{(n_X+n_Y-2)}$ where $s_p = \sqrt{\frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}}$

Hypothesis Testing for a Proportion

1^{st} Formulate the hypotheses

2^{nd} Compute test statistic (under the null hypothesis)

3^{rd} Compute p-value (or compare test statistic with critical value) and draw a conclusion

	For p with $X \sim bin(n, p)$	**For $p_X - p_X$ with $X \sim bin(n_X, p_X)$ and $Y \sim bin(n_Y, p_Y)$**
1^{st}	$H_0 : p = p_0$ $H_a : p > p_0$ or $H_a : p < p_0$ or $H_a : p \neq p_0$	$H_0 : p_X = p_Y$ $H_a : p_X > p_Y$ or $H_a : p_X < p_Y$ or $H_a : p_X \neq p_Y$
2^{nd}	$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	$z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n_X + \hat{p}_Y(1-\hat{p}_Y)/n_Y}}$
3^{rd}	$P(Z \geq z)$ or $P(Z \leq z)$ or $2 \cdot P(Z \geq \lvert z \rvert)$ where $Z \sim N(0, 1)$	$P(Z \geq z)$ or $P(Z \leq z)$ or $2 \cdot P(Z \geq \lvert z \rvert)$ where $Z \sim N(0, 1)$

Notes:

All the hypothesis tests for p on this table assume that n is large (i.e. $np \geq 10$ and $n(1-p) \geq 10$).

p: proportion of successes in the population

n: sample size; x: number of successes in the sample; $\hat{p} = \frac{x}{n}$: proportion of successes in the sample