point estimate = $\bar{x}$ or $\hat{p}$, critical value : z-score, standard error : $\frac{\theta}{\sqrt{n}}$, $\frac{\alpha}{2} = \frac{1-\text{C-level}}{2}$, Sample size = $n = \left(\frac{z\cdot\theta}{m}\right)^2$, margin of error = (critical value)·(standard error)

C interval : $\bar{x} - m < \mu < \bar{x} + m$

$\hat{p} - m < p < \hat{p} + m$

Construct Confidence interval

known θ

z-method : (Z interval)

$z_{\alpha/2} = \text{invNorm}(\text{C-level}, 0, 1, \text{center})$

$m = z_{\alpha/2} \cdot \frac{\theta}{\sqrt{n}}$

unknown θ

T-method : (T interval)

$t_{\alpha/2} = \text{invT}\left(1-\frac{\alpha}{2}, n-1\right)$

$m = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

C-interval for proportion : (1 prop Z Int)

$\hat{p} = \frac{x}{n}$

$z_{\alpha/2} = \text{invNorm}(\text{C-level}, 0, 1, \text{center})$

$m = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{m}\right)^2$ or $0.25\left(\frac{z_{\alpha/2}}{m}\right)^2$

critical values using chi-square :

$\chi^2_{1-\frac{\alpha}{2}} = \left(1-\frac{\alpha}{2}, df\right)$

$\chi^2_{\frac{\alpha}{2}} = \left(\frac{\alpha}{2}, df\right)$

C-interval for population θ is

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

null hypothesis : $H_0 : \mu = \mu_0$

alternate hypothesis : $H_1 : \mu < \mu_0$, $\mu > \mu_0$, $\mu \neq \mu_0$

level of significance : $\alpha = 0.05$ (if not mentioned)

Type I error : reject H_0 when H_0 is true ($p > \alpha$)

Type II error : do not reject H_0 when H_0 is false ($p < \alpha$)

If $p < \alpha$, reject H_0. Not enough evidence

If $p > \alpha$, do not reject H_0. Enough evidence

Smaller p is, stronger against H_0

Hypothesis test

Known θ

z-test :

Test statistic : $z = \frac{\bar{x} - \mu_0}{\left(\frac{\theta}{\sqrt{n}}\right)}$

p-value :

$p = \text{normalcdf}(z, \infty, 0, 1)$ left-tailed ($-\infty$, z) / right-tailed

$p = 2\cdot\text{normalcdf}(z, \infty, 0, 1)$ two-tailed

unknown θ

T Test :

test statistic : $t = \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$

p-value :

$p = \text{tcdf}(t, \infty, n-1)$ left-tailed ($-\infty$, t) / right-tailed

$2\cdot\text{tcdf}(t, \infty, n-1)$ two-tailed

Hypothesis test for proportion

$H_0 : p = p_0$

$H_1 : p < p_0$, $p > p_0$, $p \neq p_0$

1 prop Z Test

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

p-value $= \text{normalcdf}(z, \infty, 0, 1)$ left-tailed ($-\infty$, z) / right-tailed

$2\cdot\text{normalcdf}(z, \infty, 0, 1)$ two-tailed

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$, $\mu_1 > \mu_2$, $\mu_1 \neq \mu_2$

standard error of $\bar{x}_1 - \bar{x}_2 = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

degree of freedom : smaller of $n_1 - 1$ and $n_2 - 1$

Two means : Independent samples

2-SampTTest :

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

p-value : $p = \text{tcdf}(z, \infty, df)$ left-tailed ($-\infty$, z) / right-tailed

$2\cdot\text{tcdf}(z, \infty, df)$ two-tailed

$H_0 : p_1 = p_2$

$H_1 : p_1 < p_2$, $p_1 > p_2$, $p_1 \neq p_2$

mean $= p_1 - p_2$, standard deviation $= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$, standard error $= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Two proportions

2-PropZTest :

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

p-value : $p = \text{normalcdf}(z, \infty, 0, 1)$ left-tailed ($-\infty$, z) / right-tailed

matched pairs : dependent samples

$d = \bar{x}_1 - \bar{x}_2$

$\bar{d}$ = mean of d

$H_0 : \mu_d = 0$

$H_1 : \mu_d < 0$, $\mu_d > 0$, $\mu_d \neq 0$

Two means : paired samples

T Test :

Test statistic : $t = \frac{\bar{d} - \mu_0}{\left(\frac{s_d}{\sqrt{n_d}}\right)}$ always 0

p-value : $p = \text{tcdf}(z, \infty, n_d - 1)$

Assumptions : SRS and $n > 30$ or normally distributed

Assumptions for proportion : SRS, population $\geq 20\cdot n$, categories = 2

and each categories > 10