Mean : $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Variance : $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

Standard Deviation : $s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

1-Sample z Test : $z = \dfrac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$

1-Sample t Test : $t = \dfrac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$

2-Sample z Test : $z = \dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

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

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

2-Prop z-Test $z = \dfrac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\hat{p}_1 = \frac{x}{n_1}$, $\hat{p}_2 = \frac{y}{n_2}$, $\hat{p} = \frac{x+y}{n_1+n_2}$

2-Sample F-Test : $F = \dfrac{S_1^2}{S_2^2}$

Paired t-test : $\dfrac{\bar{D}-D_0}{\frac{S_D}{\sqrt{n}}}$, Normally $D_0 = 0$

Wilson's Adjustment

- Adjusted standard error: $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2}{4n^2}}$
- Confidence Interval : $\dfrac{\hat{p}+\frac{z^2}{2n} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2}{4n^2}}}{1+\frac{z^2}{n}}$

Student's t-Distribution : $t = \dfrac{x-\mu}{\frac{s}{\sqrt{n}}}$

Confidence Interval for population proportion p:

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

Unequal variance two-sample t-test for Independent samples: Welch-Satterthwaite Approximate t-test:

test statistic: $t = \dfrac{\bar{x}-\bar{Y}-D_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$

df for unequal variance two-sample t-test for independent samples:

$$v = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)}+\frac{S_2^4}{n_2^2(n_2-1)}}$$

Pooled standard deviation for Paired t-test:

$$S_D = \sqrt{\frac{1}{n-1}\left(\sum D_i^2 - \frac{(\sum D_i)^2}{n}\right)}$$

Confidence Intervals:

1-Prop z interval:

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

2-Prop z Interval:

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

2-Sample z Interval:

$$(\bar{x}_1-\bar{x}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

2-Sample t Interval:

$$(\bar{x}_1-\bar{x}_2) \pm t_{\frac{\alpha}{2},df}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

Two-sample test of hypothesis about population variance σ^2.

- Both populations are Normally distributed.
- Two samples are independent.

Case 1: $H_0: \sigma_1^2 = \sigma_2^2$, $H_a: \sigma_1^2 > \sigma_2^2$

$$F = \frac{S_1^2}{S_2^2}$$

Reject H_0 if $F > F_\alpha$ is based on (n_1-1, n_2-1) df.

Case 2: $H_0: \sigma_1^2 = \sigma_2^2$, $H_a: \sigma_1^2 < \sigma_2^2$

$$F = \frac{S_2^2}{S_1^2}$$

Reject H_0 if $F > F_\alpha$ where F_α is based on (n_2-1, n_1-1) df.

Decision rule simple: (Test for the equality of two proportions)

Case 1: $H_0: p_1-p_2=0$ vs. $H_a: p_1-p_2>0$

Reject H_0 if $z > z_\alpha$

Case 2: $H_0: p_1-p_2=0$ vs. $H_a: p_1-p_2<0$

Reject H_0 if $z < -z_\alpha$

Case 3: $H_0: p_1-p_2=0$ vs. $H_a: p_1-p_2 \neq 0$

Reject H_0 if $z > z_{\frac{\alpha}{2}}$ OR $z < -z_{\frac{\alpha}{2}}$

Two-sample test of hypothesis about population mean difference $(\mu_1-\mu_2)$

(Small sample case)

- Both populations are Normally distributed.
- Two samples are independent.
- Two population variances are equal.

$t = \dfrac{\bar{x}-\bar{Y}-D_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $S_p = \sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

Small sample confidence interval for $(\mu_1-\mu_2)$

$$(\bar{x}-\bar{Y}) \pm t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Case 3: $H_0: \sigma_1^2 = \sigma_2^2$, $H_a: \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)}$$

Reject H_0 if $F > F_{\frac{\alpha}{2}}$ where the df is given by the case 1 or case 2 depending on the form of the F-statistic.

Power of a Test: The probability of correctly rejecting the H_0, i.e., the probability that we accept H_a when in fact H_a is true.

$$\text{Power} = 1 - P(\text{Type II}) = 1 - \beta$$

Sample size Estimation:

Mean Estimation: $n = \left(\frac{Z \cdot \sigma}{E}\right)^2$

Proportion Estimation: $n = \frac{Z^2 \cdot \hat{p} \cdot (1-\hat{p})}{E^2}$

$\hat{p}$ is the estimated proportion, and if no preliminary estimate is available, we can use $\hat{p} = 0.5$ to maximize the sample size.

If the population standard deviation is unknown, we can divide the range by 4 to estimate the standard deviation. $\sigma = \frac{R}{4}$

The value of alpha (α) is the significance level of the test, which is the probability of rejecting the null hypothesis when it is true.

The rejection region is the set of values of the test statistic that would lead us to reject the null hypothesis.

Smaller the p-value, more supportive it is of the alternative hypothesis.

Reject H_0 if p-value < alpha (α)

p-value: Observed Significance level of a test is the probability of observing a value of the test statistic that at least as supportive of the alternative hypothesis as the one observed from the sample data.

Confidence Interval for population proportion p

$n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$

if not met: Modified sample proportion

$$\tilde{p} = \frac{x+2}{n+4}$$

A higher confidence level provides a greater degree of certainty about the estimate, which requires a wider interval.

Reduce the width of confidence interval

- Increase the sample size
- Decrease the confidence level.

F-Distribution

- A continuous distribution that sits on the positive side of the real line, and skewed to the right.
- Determined by two parameters. (V_1, V_2).
  - V_1 is called the numerator degrees of freedom
  - V_2 is called the denominator degrees of freedom.