Summary of Hypothesis Tests and Confidence Intervals

Tuesday, 8 November 2022 21:10

Or, calculate the p-value, and determine if it is significant or not based off the specified Alpha (0.05).

X bar = sample mean, Mu0 = null hypothesis of population mean, sigma = population std.
S = sample std,

Summary of Hypothesis Tests and Confidence Intervals

	H_0	The statistical value/test	H_1	Rejection region	Confidence interval	In R
1.	$\mu = \mu_0$ Test for a sample mean against a Normal Distribution Scipy.ttest_1samp	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ Sigma/sqt(n) = variance/n. This is the estimate of the std of our population, using the std of the sample of the sample mean. This is a pretty close approximation, but not exact σ is known and $X \sim N$ Or $n > 30$	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_{1-\alpha}$ Table $z > z_{1-\alpha}$ $\lvert z \rvert > z_{1-\alpha/2}$	of: μ $\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$	z.test(Data, mu, Sigma.X, alternative = ...)
2.	$\mu = \mu_0$ Student's T-test, for sample mean against a T distribution	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ DF $\nu = \text{n} - 1$ σ is unknown, $X \sim N$ and $n \leq 30$	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ Table $t > t_\alpha$ $\lvert t \rvert > t_{\frac{\alpha}{2}}$	of: μ $\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$	t.test(Data, mu, alternative = ...)
3.	$\mu_1 - \mu_2 = d_0$ Testing difference of two independent sample means (scipy.ttest_ind())	z test $z = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ Often equal 0 σ_1, σ_2 are known $X_1, X_2 \sim N$ or $n_1, n_2 > 30$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_{1-\alpha}$ $z > z_{1-\alpha}$ $\lvert z \rvert > z_{1-\alpha/2}$	of: $\mu_1 - \mu_2$ $\bar{x}_1 - \bar{x}_2 \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	Z.test(X = X_Data, Y = Y_Data, mu = 0, Sigma.X = 15, Sigma.Y = 15)
4.	$\mu_1 - \mu_2 = d_0$ Welch's T-Test for unequal variances	$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $\sigma_1 = \sigma_2$ are unknown $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ $v = n_1 + n_2 - 2$ $X_1, X_2 \sim N$ and $n_1 \leq 30$ or $n_2 \leq 30$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $\lvert t \rvert > t_{\frac{\alpha}{2}}$	of: $\mu_1 - \mu_2$ $\bar{x}_1 - \bar{x}_2 \pm t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	t.test(X = X_Data, Y = Y_Data, Var.equal = True / False) Welch's
5.	$\mu_1 - \mu_2 = d_0$ T test for Paired Samples	Average of difference $t = \frac{\bar{d} - d_0}{s_d/\sqrt{n}}$ $\text{v} = \text{n} - 1$ Paired samples $D \sim N, n \leq 30$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $\lvert t \rvert > t_{\frac{\alpha}{2}}$	of: μ_d $\bar{d} \pm t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}}$	t.test(X = Xp, Y = Yp, Paired = True)
6.	$p = p_0$ proportions of populations	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ $n\hat{p}, n(1-\hat{p}) \geq 5$	$p < p_0$ $p > p_0$ $p \neq p_0$	$z < -z_{1-\alpha}$ $z > z_{1-\alpha}$ $\lvert z \rvert > z_{1-\alpha/2}$	of: p $\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	

	H_0	The statistical value/test	H_1	Rejection region	Confidence interval	
7.	$p_1 - p_2 = d_0$ Difference of proportions	$z = \frac{\hat{p}_1 - \hat{p}_2 - d_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$ $n_1\hat{p}_1, n_1(1-\hat{p}_1) \geq 5$ $n_2\hat{p}_2, n_2(1-\hat{p}_2) \geq 5$ $d_0 = 0$ then we take the following as the denominator: $\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ for $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$	$p_1 - p_2 < d_0$ $p_1 - p_2 > d_0$ $p_1 - p_2 \neq d_0$	$z < -z_{1-\alpha}$ $z > z_{1-\alpha}$ $\lvert z \rvert > z_{1-\alpha/2}$	of: $p_1 - p_2$ $\hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	
8.	$\sigma^2 = \sigma_0^2$ Chi squared: Testing a variance Variance of a population is = x	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ $\text{v} = \text{n} - 1$ $X \sim N$	$\sigma^2 < \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha}$ $\chi^2 > \chi_\alpha^2$ $\chi^2 < \chi^2_{1-\frac{\alpha}{2}}$ or $\chi^2 > \chi^2_{\frac{\alpha}{2}}$	of: σ^2 between $\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}$ and $\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}$	// Install "EnvStats" VarTest(X = X_data, alternative = "two.sided", Sigma.squared = σ^2
9.	$\sigma_1^2 = \sigma_2^2$ F Test: Testing a variance against another variance	$F = \frac{s_1^2}{s_2^2}$ V = Degrees of freedom $\text{v}_1 = \text{n}_1 - 1$ $\text{v}_2 = \text{n}_2 - 1$ If the test is two-sided, then X_1 is the variable with the greatest variance according to H_1 $X_1, X_2 \sim N$	$\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	$F > F_\alpha(\text{v}_1, \text{v}_2)$ $F < F_{1-\alpha/2}(\text{v}_1, \text{v}_2)$ or $F > F_{\alpha/2}(\text{v}_1, \text{v}_2)$	of: $\frac{\sigma_1^2}{\sigma_2^2}$ between $\frac{s_1^2}{s_2^2} \frac{1}{F_{\frac{\alpha}{2}}(\nu_1, \nu_2)}$ and $\frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}}(\nu_2, \nu_1)$	var.test(X, Y, alternative = "two.sided") Example: Testing the variance of a sample of Men's Wages vs. women's wages.

$$F_\alpha(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$$