Hypothesis testing

1. Null & Alternative
2. Significance Level
   Rejection Region
3. test statistic
   confidence Interval
   p-value
4. Make a conclusion about your hypothesis.

Single Sample mean

$X_1 \cdots X_n \sim N(\mu, \sigma^2)$

unknown

Does the data inform you about a specified value μ_0

μ_0 is what we beleive to be true.

$H_0: \mu = \mu_0$

$H_a: \mu \neq \mu_0$

$\mu > \mu_0$

$\mu < \mu_0$

choose 1 alternative hypothesis.

test statistic.

$n < 30$ $\quad \sigma$ is known

$$ts = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

$n < 30$ $\quad \sigma$ is unknown

$$x_1 \dots x_n \sim N(\mu, \sigma^2)$$

$$ts = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{DF}$$

$$DF = n - 1$$

$n \geq 30$ $\quad \sigma$ is unknown

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

$\sim N(0,1)$
via CLT
$s^2 \xrightarrow[n \to \infty]{} \sigma^2$

$\sim t_{DF}$
$DF = n-1$

$t_{DF} \to N$
$DF \to \infty$
$DF = 30$

Single Sample proportion

$X_1, \dots X_n \sim \text{Bernoulli}(p)$

$\hat{p} = \frac{1}{n} \Sigma X_i$ = sample proportion

$H_0 \; p = p_0$

$H_a : p \neq p_0$

$p > p_0$

$p < p_0$

Large Sample Theory

$n p_0 > 10 \qquad n(1-p_0) > 10$

$$ts = z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$$

Two Sample Means. $X \perp Y$

$X_1, \dots X_m \sim N(\mu_1, \sigma_1^2)$

$Y_1 \dots Y_n \sim N(\mu_2, \sigma_2^2)$

$\mu_1 - \mu_2$

Δ_0 = difference

$H_0: \mu_1 - \mu_2 = \Delta_0$ 0 is most common.

$H_a: \mu_1 - \mu_2 \neq \Delta_0$

$\mu_1 - \mu_2 > \Delta_0$

$\mu_1 - \mu_2 < \Delta_0$

Test statistic

σ_1 & σ_2 known

$$ts = z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$$

$n, m > 30$

σ_1^2 & σ_2^2 unknown

$$TS = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} \sim t_{DF}$$

$$DF = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{\left(s_1^2/m\right)^2}{m-1} + \frac{\left(s_2^2/n\right)^2}{n-1}}$$

round down to whole number.

If you believe

$$\sigma_1^2 = \sigma_2^2$$

$$ts = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{S_p^2\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim t_{DF}$$

S_p^2 = pooled variance

$$S_p^2 = \frac{m-1}{m+n-2} S_1^2 + \frac{n-1}{m+n-2} S_2^2$$

$$Df = m+n-2$$

2-sample proportions.

$X_1 \ldots X_m \sim \text{Bernoulli}(p_1)$

$Y_1 \ldots Y_n \sim \text{Bernoulli}(p_2)$

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \neq 0$

$p_1 - p_2 > 0$

$p_1 - p_2 < 0$

Large Samples Theory

$$t.s. = z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m}+\frac{1}{n}\right)}} \sim N(0,1)$$

$\hat{p}_1 = \frac{1}{m}\sum X_i \qquad \hat{p}_2 = \frac{1}{n}\sum Y_i$

weighted average

$$\hat{p} = \frac{m}{m+n}\hat{p}_1 + \frac{n}{m+n}\hat{p}_2$$

$$\hat{q} = 1-\hat{p}$$

Conditions

$m\hat{p}_1 > 10 \qquad m(1-\hat{p}_1) > 10$

$n\hat{p}_2 > 10 \qquad n(1-\hat{p}_2) > 10$

2 - Sample Variances

This must be true:

$X_1 \cdots X_m \sim N(\mu_1, \sigma_1^2)$

$Y_1 \cdots Y_n \sim N(\mu_2, \sigma_2^2)$

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \leftarrow$ Pooled variance

$H_a: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \leftarrow$ not pooled

$ts = F = \frac{S_1^2}{S_2^2} \sim F_{v_1, v_2}$

$v_1 = m-1 \qquad v_2 = n-1$

put the larger sample variance as S_1^2

If Data is not normal

Levene's Test

A median - based test

Z - table.

$Z \sim N(0,1)$

Standard Normal.

$Z = 0.25$

$P(Z \leq 0.25) = 0.5987$

$P(Z \geq 0.25) = 1 - 0.5987$

$P(Z \geq 0.25) = 0.4013$

$z = 2.848$

$z = 2.85$

$P(z \le 2.85) = 0.9978$

$z = 6.80$

$P(z \le 6.80) \approx 1$

$z = -6.80$

$P(z \le -6.80) \approx 0$

Critical Value

$\alpha = 0.05$

$\alpha = 0.025$

$\mu \ne 0$

$z_{\alpha/2} = \pm 1.96$

$\alpha/2 = 0.05$

$z_{\alpha/2} = \pm 1.645$

$\alpha/2 = 0.1$

$z_{\alpha/2} = \pm 1.28$

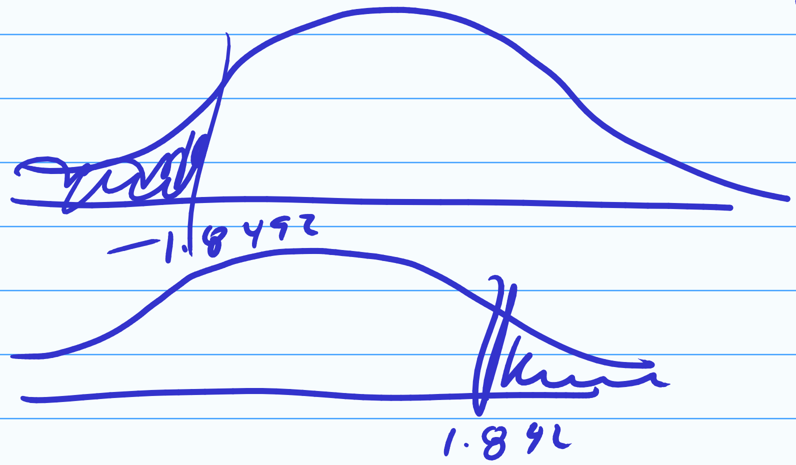
$DF = 20$

$0.005 > P(t > 2.844) > 0.001$

$DF = 25$ $\alpha = 0.05$

$0.05 > P(t < -1.842) > 0.025$

still Reject Ho



$DF = 48$

$0.025 > P(t > 2.381) > 0.01$

CV $DF = 28$ $\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{cv} = \pm 2.048$