Chapter 10

• Two samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$ (from normal populations or n and m are large enough); $EX_i = \mu_X$, $\text{Var}(X_i) = \sigma_X^2$, $EY_i = \mu_Y$, $\text{Var}(Y_i) = \sigma_Y^2$. Testing $H_0 : \mu_X - \mu_Y = d_0$ (d_0 is a given number, usually $d_0 = 0$). Alternatives a) $H_1 : \mu_X - \mu_Y > d_0$, b) $H_1 : \mu_X - \mu_Y < d_0$, c) $H_1 : \mu_X - \mu_Y \neq d_0$. The level of significance α.

1) σ_X^2 and σ_Y^2 are known. The test statistic

$$Z = \frac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}.$$

Z has the standard normal distribution under H_0 (exactly or approximately). The critical region: a) $Z \geq z_\alpha$, b) $Z \leq -z_\alpha$, c) $|Z| \geq z_{\alpha/2}$.

2) σ_X^2 and σ_Y^2 are unknown but equal: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. The test statistic

$$T = \frac{(\bar{X} - \bar{Y}) - d_0}{S_p\sqrt{1/n + 1/m}}$$

where

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n + m - 2}.$$

T has the t-distribution with $n + m - 2$ degrees of freedom under H_0 (exactly or approximately). The critical region: a) $T \geq t_{\alpha,n+m-2}$, b) $T \leq -t_{\alpha,n+m-2}$, c) $|T| \geq t_{\alpha/2,n+m-2}$.

3) σ_X^2 and σ_Y^2 are unknown and unequal. The test statistic

$$T = \frac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{S_X^2/n + S_Y^2/m}}.$$

T has the t-distribution with v degrees of freedom under H_0 (approximately), where v is the nearest integer to

$$\frac{(S_X^2/n + S_Y^2/m)^2}{(S_X^2/n)^2/(n-1) + (S_Y^2/m)^2/(m-1)}.$$

The critical region: a) $T \geq t_{\alpha,v}$, b) $T \leq -t_{\alpha,v}$, c) $|T| \geq t_{\alpha/2,v}$.

• Paired data.

$$X_1, X_2, ..., X_n,\ EX_i = \mu_X\ (X_i = \mu_X + \epsilon_{1i}),$$

$$Y_1, Y_2, ..., Y_n,\ EY_i = \mu_Y\ (Y_i = \mu_Y + \epsilon_{2i}).$$

X-s are independent, Y-s are independent but X_i og Y_i are dependent for each i. Testing $H_0 : \mu_X - \mu_Y = d_0$ (d_0 is a given number, usually $d_0 = 0$). Alternatives a) $H_1 : \mu_X - \mu_Y > d_0$, b) $H_1 : \mu_X - \mu_Y < d_0$, c) $H_1 : \mu_X - \mu_Y \neq d_0$. The level of significance α. Test statistic

$$T = \sqrt{n}\frac{\bar{D} - d_0}{S_D}$$

where

$$D_i = X_i - Y_i,\ \bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i,\ S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2.$$

T has the t-distribution with $n-1$ degrees of freedom under H_0 (exactly or approximately). The critical region: a) $T \geq t_{\alpha,n-1}$, b) $T \leq -t_{\alpha,n-1}$, c) $|T| \geq t_{\alpha/2,n-1}$.

• Binomial data. One variable (sample). $X \sim$binomial(n,p). Testing $H_0 : p = p_0$. Alternatives a) $H_1 : p > p_0$, b) $H_1 : p < p_0$, c) $H_1 : p \neq p_0$. The level of significance α.

The test statistic

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}.$$

Z has the standard normal distribution under H_0 (approximately). Critical region: a) $Z \geq z_\alpha$, b) $Z \leq -z_\alpha$, c) $|Z| \geq z_{\alpha/2}$.

• Binomial data. Two variables (samples). $X \sim$binomial(n,p_X), $Y \sim$binomial(m,p_Y). Testing $H_0 : p_X = p_Y$. Alternatives a) $H_1 : p_X > p_Y$, b) $H_1 : p_X < p_Y$, c) $H_1 : p_X \neq p_Y$. The level of significance α.

The test statistic

$$Z = \frac{\frac{X}{n} - \frac{Y}{m}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

where

$$\hat{p} = \frac{X+Y}{n+m}.$$

Z has the standard normal distribution under H_0 (approximately). Critical region: a) $Z \geq z_\alpha$, b) $Z \leq -z_\alpha$, c) $|Z| \geq z_{\alpha/2}$.

• Test concerning variance (one sample).

Testing $H_0 : \sigma^2 = \sigma_0^2$. Alternatives a) $H_1 : \sigma^2 > \sigma_0^2$, b) $H_1 : \sigma^2 < \sigma_0^2$, c) $H_1 : \sigma^2 \neq \sigma_0^2$. Signifikansnivå α.

The test statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}.$$

χ^2 has chi-squared distribution with $n-1$ degrees of freedom under H_0. Critical region: a) $\chi^2 \geq \chi^2_{\alpha,n-1}$, b) $\chi^2 \leq \chi^2_{1-\alpha,n-1}$, c) $\chi^2 \leq \chi^2_{1-\alpha/2,n-1}$ or $\chi^2 \geq \chi^2_{\alpha/2,n-1}$.

• Two samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$ (from normal populations or n and m are large enough); $\text{Var}(X_i) = \sigma_X^2$, $\text{Var}(Y_i) = \sigma_Y^2$. Testing $H_0 : \sigma_X^2 = \sigma_Y^2$. Alternatives a) $H_1 : \sigma_X^2 > \sigma_Y^2$, b) $H_1 : \sigma_X^2 < \sigma_Y^2$, c) $H_1 : \sigma_X^2 \neq \sigma_Y^2$. The level of significance α.

The test statistic

$$F = \frac{S_X^2}{S_Y^2}.$$

Critical region: a) $F \geq f_{\alpha,n-1,m-1}$, b) $F \leq f_{1-\alpha,n-1,m-1}$, a) $F \geq f_{\alpha/2,n-1,m-1}$ or $F \leq f_{1-\alpha/2,n-1,m-1}$