Test of two means in two univariate normal population

$X_{11}, X_{12}, \cdots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ # n_1

$X_{21}, X_{22}, \cdots, X_{2n_2} \sim N(\mu_2, \sigma_2^2)$ # n_2

A set of random sample n_1 from 1st population

A set of random " n_2 " 2nd "

suppose $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ are not known.

We want to test whether the means of two populations are same.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

$H_2: \mu_1 < \mu_2$

$H_3: \mu_1 \neq \mu_3$

see if σ_1^2, σ_2^2 are known, the test statistic will be based on $\bar{x}_1$ and $\bar{x}_2$, i.e.,

$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$\therefore Z = \dfrac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \sigma_2^2/n_2}} \sim N(0,1)$

When σ_1^2, σ_2^2 are unknown, we've to make an assumption $\boxed{\sigma_1^2 = \sigma_2^2}$.

Under this assumption, we know,

independent

$\dfrac{(n_1 - 1)s_1^2}{\sigma_1^2} \sim \chi^2_{n_1 - 1}$

$\dfrac{(n_2 - 1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_2 - 1}$

where s_1^2 = sample variance $= \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2$

$s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2$

$\Rightarrow \dfrac{(n_1 - 1)s_1^2}{\sigma_1^2} + \dfrac{(n_2 - 1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_1 + n_2 - 2}$

$\Rightarrow$ Under the assumption, $\sigma_1 = \sigma_2 = \sigma$ (common variance)

$$E\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right] = \sigma^2$$

So, $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ = pooled sample variance will be estimate of σ^2.

Again $\dfrac{(n_1 + n_2 - 2)s^2}{\sigma^2} \sim \chi^2_{n_1 + n_2 - 2}$

Thus test statistic for $H_0: \mu_1 = \mu_2$ is

$$\underbrace{\frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}_{N(0,1)} \Bigg/ \underbrace{\sqrt{\frac{(n_1+n_2-2)s^2}{\sigma^2}\Big/(n_1+n_2-2)}}_{\sqrt{\chi^2_{n_1+n_2-2}/n_1+n_2-2}} \sim t_{n_1+n_2-2}$$

↑ Fisher's t. statistic

$$\Rightarrow \quad \frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Under $H_0: \mu_1=\mu_2$, $t = t_{cal} = \dfrac{\bar{x}_1-\bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

We reject H_0, if $t_{cal} > t_{\alpha, n_1+n_2-2}$

$H_1: \mu_1 < \mu_2$, $\quad t_{cal} < -t_{\alpha, n_1+n_2-2}$

$H_1: \mu_1 \neq \mu_2$ $\quad |t_{cal}| > t_{\alpha/2, n_1+n_2-2}$.

$\boxed{\text{Remember}}$ If the test is $H_0: \mu_1-\mu_2 = \xi$ (a specified value) say $H_0: \mu_1 = 2+\mu_2$, then $\xi = 2$.

test statistic will be.

$$\frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Under H_0, $\dfrac{\bar{x}_1-\bar{x}_2-\xi}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

Test between two variances

when μ_1, μ_2 unknown

$H_0: \sigma_1^2 = \sigma_2^2$

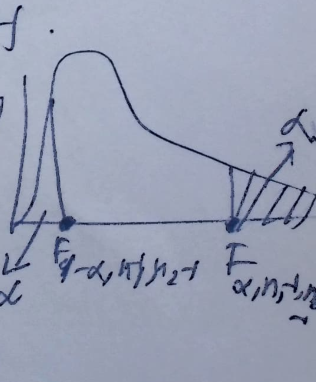
$$F = \frac{(n_1-1)s_1^2/\sigma_1^2}{(n_2-1)s_2^2/\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2}\,\frac{(n_1-1)s_1^2}{(n_2-1)s_2^2} \sim F_{n_1-1, n_2-1}.$$

Under H_0, $F_{cal} = \dfrac{(n_1-1)s_1^2}{(n_2-1)s_2^2} \sim F_{n_1-1, n_2-1}$

We reject $H_0: \sigma_1^2 = \sigma_2^2$ ag. $H_1: \sigma_1^2 > \sigma_2^2$ if.

$F_{cal} > F_{\alpha, n_1-1, n_2-1}$

F also positively skewed ranging from $0 < F < \infty$



for $H_1: \sigma_1^2 < \sigma_2^2$

$F_{cal} < F_{1-\alpha, n_1-1, n_2-1}$.

Paired t test

$(X, Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$H_0: \mu_1 = \mu_2$, $\sigma_1^2, \sigma_2^2, \rho$ are unknown.

Concept

A paired t test is used when we are interested in the difference between two variables for the same subject. For this type of t test we have to consider dependent paired observation.

Example: suppose we have 12 patients whose cholesterol level is montinored every week. So cholesterol level at the beginning of time and cholesterol level after 4 weeks — these two variables are interrelated but subjects (patients) remain the same.

$Z = X - Y$.

$E(Z) = \mu_1 - \mu_2 = \mu_Z$

$V(Z) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 = \sigma_Z^2$. $E(\bar{Z}) = \mu_Z$, $V(\bar{Z}) = \frac{\sigma_Z^2}{n}$.

Under null hyp. $\mu_Z = 0$.

Test statistic: $\frac{\sqrt{n}(\bar{Z} - \mu_Z)}{\sigma_Z}$ where n being the no of obs.

$\frac{\sqrt{n}(\bar{Z} - \mu_Z)}{\sigma_Z} \sim N(0,1)$.

when σ_Z is not known, we replace σ_Z by s_Z.

Now, the test statistic is

$$\frac{\sqrt{n}(\bar{Z} - \mu_Z)}{s_Z} \sim t_{n-1}$$

Under H_0, $\frac{\sqrt{n}\,\bar{Z}}{s_Z} \sim t_{n-1}$.

Remember

For practical calculation data for paired t will be as (x_i, y_i), $i = 1(1)n$, find $z_i = x_i - y_i$, then find $\bar{z}$ (sample mean) and $s_Z^2 = \frac{1}{n-1}\sum(z_i - \bar{z})^2$.

test statistic $\frac{\sqrt{n}(\bar{Z} - \mu_Z)}{s_Z} \sim t_{n-1}$