II Testing problems on two independent ~~variables~~ univariate normal distribution

17th March 2023

Eg 1 A pharmaceutical company manufactures a new drug and wants to test it ag an existing drug on the recovery time of patients suffering from a particular disease.

H_0 : the existing and new drug are equally effective

H_1 : the new drug is more effective than the old drug.

which is equivalent

H_0 : average recovery time for new drug = average recovery time for old drug.

ag H_1: <

If we define X : recovery time of the new drug

and Y : recovery time of old drug.

Assume that $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$

X and Y are independent

Hence we are to test $H_0 : \mu_X = \mu_Y$ ag $H_1 : \mu_X < \mu_Y$

In general let us define two random variables X and Y such that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ and we assume X, Y are independently distributed

i) Test of equality of means

We are to test $H_0 : \mu_X = \mu_Y$

(a) We assume that variances σ_X^2 and σ_Y^2 are known say $\sigma_X^2 = \sigma_{X_0}^2$ and $\sigma_Y^2 = \sigma_{Y_0}^2$, $\sigma_{X_0} > 0$, $\sigma_{Y_0} > 0$

The null hypothesis can be reframed as

$H_0 : \mu_X - \mu_Y = 0$

let us consider a random sample of size n_1 from the distribution of X.

let $X_1, X_2, \ldots, X_{n_1}$ be the samples.

and random sample of size n_2 from the distn of Y. let them be $Y_1, \ldots, Y_{n_2}$

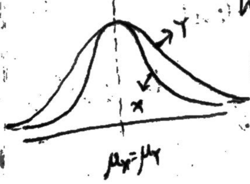
[Sample drawn are randomly selected patients]

Define $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ and $\bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$

An approximate estimator for $(\mu_X - \mu_Y)$ is $(\bar{X} - \bar{Y}) = T$

Consider the quantity $(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)$

If H_0 is true i.e $\mu_X - \mu_Y = 0$ then $\bar{X} - \bar{Y}$ will be close to zero & hence the quantity



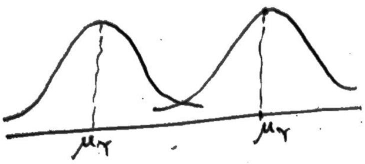
$(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)$ should be close to zero.

Case I : Suppose the alternative is $H_1 : \mu_X - \mu_Y > 0$

Then $\bar{X}$ and $\bar{Y}$ being unbiased estimators of μ_X & μ_Y respectively, will be close to μ_X and μ_Y.

and hence $\bar{X} - \bar{Y} > 0$

Under H_0 (assuming that H_0 is true) the quantity $(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y) = \bar{X} - \bar{Y} > 0$, Thus

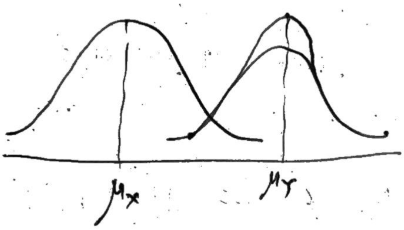


high +ve values of $(\bar{X}-\bar{Y})$ indicate departure of H_0 towards $H_1 : \mu_X - \mu_Y > 0$.

Case II : Suppose the alternative is $H_1 : \mu_X - \mu_Y < 0$

following a similar logic as in case I, under the assumption that H_0 is true, the quantity

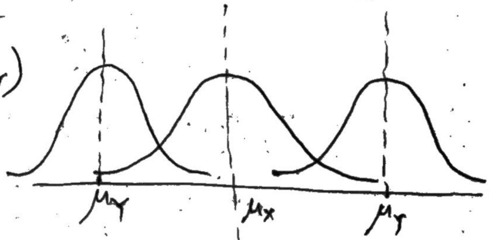
$$(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y) = (\bar{X}-\bar{Y}) < 0$$



Thus high -ve values of $\bar{X}-\bar{Y}$ indicate departure of H_0 towards $H_1 : \mu_X - \mu_Y < 0$

Case III Suppose the alternative is $H_1 : \mu_X - \mu_Y \neq 0$

Combining the logic used in case I and II under assumption H_0 is true, the quality $(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y) = (\bar{X}-\bar{Y})$



is widely different from 0.

Thus high ±ve values indicate departure of H_0 towards $H_1 : (\mu_X - \mu_Y) \neq 0$

Thus the test statistic for testing H_0 should be based on the quantity $(\bar{X}-\bar{Y})$

We have $V(\bar{X}-\bar{Y}) = V(\bar{X}) + V(\bar{Y}) - 2\,cov(\bar{X},\bar{Y})$ and hence

$\because X, Y$ are independent so $\bar{X}, \bar{Y}$

$$cov(\bar{X},\bar{Y}) = 0$$

An apt test statistic for testing H_0 is given by

$$Z = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_{X_0}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_2}}}$$ which is under H_0

$$\Rightarrow Z \sim N(0,1)$$

<u>Critical Region</u>

<u>case I</u> : $H_1 : \mu_X - \mu_Y > 0$. $Z > K$ where K is

such that $P(\text{Type I error}) = \alpha$

$$P(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$$

$$P(Z > K \mid H_0 \text{ is true}) = \alpha$$

$\Rightarrow K = \tau_\alpha =$ upper α-point of a $N(0,1)$ distn

<u>Test Rule</u> : Reject H_0 at α los iff Z_{obs}

$$= \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_{X_0}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_2}}} > \tau_\alpha$$

<u>Case II</u>. $H_1 : \mu_X - \mu_Y < 0$

$Z < K^*$ where K^* is such that

$$P(\text{Type I error}) = \alpha$$

$K^* = \tau_{1-\alpha} = -\tau_\alpha =$ upper $(1-\alpha)$ pt or lower α-point of $N(0,1)$ distn

<u>Test Rule</u> : Reject H_0 at α los iff

$$Z_{obs} = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_{X_0}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_2}}} < -\tau_\alpha$$

Case III $\quad H_1 : (\mu_X - \mu_Y) \neq 0$

$Z < K^{**}$ and $Z > K^{***}$ where K^{**} and K^{***} are such that $P(\text{Type I error}) = \alpha$

$\Rightarrow P(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$

$\Rightarrow P(Z > K^{**} \cup Z < K^{***} \mid H_0 \text{ is true}) = \alpha$

$$\Rightarrow \underbrace{P(Z > K^{**} \mid H_0 \text{ true})}_{\alpha/2} + \underbrace{P(Z < K^{***} \mid H_0 \text{ true})}_{\alpha/2} = \alpha$$

$K^{**} = \tau_{\alpha/2}$ and $K^{***} = \tau_{1-\alpha/2} = -\tau_{\alpha/2}$

Test Rule : Reject H_0 at a los iff

$Z_{obs} > \tau_{\alpha/2}$ or $Z_{obs} < -\tau_{\alpha/2}$

ie $$|Z_{obs}| = \frac{|\bar{x} - \bar{Y}|}{\sqrt{\frac{\sigma_{X_0}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_2}}} > \tau_{\alpha/2}$$

(I) b) Variances are unknown, i) $\mu_X = \mu_Y$ Equality of means | 31st March 2023

DS Stat

We can not use the test statistic defined in (a). Since σ_X^2 and σ_Y^2 are unknown, we need to estimate them.

However, we make an assumption here, i.e, though σ_X^2 & σ_Y^2 are unknown, they are equal say $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ where σ^2 is unknown. This is called the assumption of homoscedasticity. To conclude the test statistic we need to estimate σ^2 based on the sample data $(X_1, \ldots, X_{n_1})$ and $(Y_1, \ldots, Y_{n_2})$ an apt estimation of σ^2 by the pooled estimator.

$$\hat{\sigma}^2 = \frac{(n_1-1)s_X^2 + (n_2-1)s_Y^2}{(n_1+n_2-2)}$$

where s_X, s_Y are sample variances of X and Y.

An apt test statistic for testing H_0 is given by $T = \dfrac{(\bar{X}-\bar{Y})}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$. Under H_0, $T \sim t_{n_1+n_2-2}$

Critical region : In line with the same logic as in (a).

Case I : $H_1 : \mu_X - \mu_Y > 0$ test rule : reject H_0 at α l.o.s iff $T_{obs} > t_{n_1+n_2-2;\alpha}$

Case II : $H_1 : \mu_X - \mu_Y < 0$ test rule : reject H_0 at α l.o.s iff $T_{obs} < t_{n_1+n_2-2;\alpha}$

Case III : $H_1 : \mu_X - \mu_Y \neq 0$ test rule reject H_0 at α l.o.s. iff $|T_{obs}| > t_{n_1+n_2-2;\alpha/2}$