Comparing Two Samples

$X_1, X_2, \ldots, X_n$ is drawn from $N(\mu_X, \sigma^2)$ } independent
$Y_1, Y_2, \ldots, Y_m$ is drawn from $N(\mu_Y, \sigma^2)$

Think of a medical study: X - measurements from those that recieved treatment & Y - measurements from a control group [placebo]

Effect of treatment - characterized by $\mu_X - \mu_Y$

Estimate for $\mu_X - \mu_Y = \bar{X} - \bar{Y}$

Note: $\bar{X} - \bar{Y} \stackrel{d}{=} N\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$
(by independence and properties of normals)

σ^2 - known:

$(1-\alpha)$-% C.I. for $\mu_X - \mu_Y$ can be obtained

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2}\,\sigma\sqrt{\frac{1}{n} + \frac{1}{m}},\ \bar{X} - \bar{Y} + z_{\alpha/2}\,\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

where $P(Z \geq z_{\alpha/2}) = \alpha/2$. $\quad Z \stackrel{d}{=} N(0,1)$.

This is similar to C.I. for μ that we obtained before

σ^2 – unknown

$$S_x^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}, \quad S_y^2 = \frac{\sum_{i=1}^{m}(Y_i - \bar{Y})^2}{m-1}$$

Pooled Estimate

$$S_{pooled}^2 = \frac{(n-1)\,S_x^2 + (m-1)\,S_y^2}{m+n-2}$$

- weighted Average of sample variances with weights proportional to degrees of freedom.
- if m or n is larger then it correctly adjusts.

Result :

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{1}{n} + \frac{1}{m}}\; S_{pooled}} \overset{d}{=} t_{m+n-2}$$

Confidence Interval

$(1-\alpha)\%$ CI for $\mu_x - \mu_y$ is

$$\left((\bar{X} - \bar{Y}) - t_{m+n-\frac{\alpha}{2}}(\alpha/2)\, S_{\bar{x}-\bar{y}},\; (\bar{X} - \bar{Y}) + t_{m+n-\frac{\alpha}{2}}(\alpha/2)\, S_{\bar{x}-\bar{y}} \right)$$

with $S_{\bar{x}-\bar{y}} = S_{pooled}\sqrt{\frac{1}{n}+\frac{1}{m}}$

& $P(T > t_{m+n-\frac{\alpha}{2}}(\alpha/2)) = \frac{\alpha}{2}$.

Testing:

Null Hypothesis: $\mu_x = \mu_y$

Decide on α.

let $T \stackrel{d}{=} t_{m+n-2}$

Frame Alternat:	$\mu_x > \mu_y$	$\mu_x < \mu_y$	$\mu_x \neq \mu_y$
Compute: p-value	$P\left(T > \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}}\right)$	$P\left(T < \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}}\right)$	$P\left(\lvert T\rvert > \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}}\right)$

If p-value $< \alpha$ then reject Null hypothesis.

x ——— x ———

Generalise: Compare treatments

I - treatments

J - population / sample size in each

$Y_{ij} \equiv$ the j^{th} observation of i^{th} treatment.

[Kalyani's talk]

Model $= \mu + \alpha_i + \epsilon_{ij} \longrightarrow$ Random Error independent $N(0, \sigma^2)$

α_i: differential effect $\left(\sum_{i=1}^{n} \alpha_i = 0\right)$

$E[Y_{ij}] \equiv$ Expected response measurement $= \mu + \alpha_i$

If $\alpha_i = 0$ for $i = 1, \ldots I$ $\equiv$ then all treatments have same response measurement Expected

$\alpha_i - \alpha_j$ $\equiv$ difference in Expected values of response measurement between treatment i and j.

Analysis of variance

$$\bar{Y} = \underbrace{\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij}}{IJ}}_{\text{overall}} , \quad \underbrace{\bar{Y}_{i.} = \frac{1}{J} \sum_{j=1}^{J} Y_{ij}}_{\text{under } i^{th} \text{ treatment}}$$

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{ij} - \bar{Y}_{i.})^2 + J \sum_{i=1}^{I} (\bar{Y}_{i.} - \bar{Y})^2$$

Ex: CROSS TERMS are all zero

SS_{Total} = $SS_{\text{within groups}}$ + $SS_{\text{Between treatment Groups}}$

treatment

$SS_{\text{within groups}} \equiv SS_W$

$SS_{\text{Between treatment Groups}} \equiv SS_B$

Independence & Normal assumption

Results: $\bullet$ $E[SS_w] = I(J-1)\sigma^2$

and

$$\frac{SS_w}{\sigma^2} \sim \chi^2_{I(J-1)}$$

Results:

- $E[SS_B] = J\left(\sum_{i=1}^{J} \alpha_i^2\right) + (I-1)\sigma^2$

If $\alpha_1 = \alpha_2 = \cdots = \alpha_J = 0$ then

- $\frac{SS_B}{\sigma^2} \overset{d}{=} \chi^2_{I-1}$

- SS_B is independent of SS_w

- $\frac{SS_B}{I-1} \Big/ \frac{SS_w}{I(J-1)} \overset{d}{=} F(I-1, I(J-1))$ (*)

$\boxed{\text{See H.W.}}$

One can use above to test.

Null hypothesis :- $\alpha_1 = \alpha_2 \cdots = \alpha_I = 0$

Decide on $\underline{\alpha}$

p-value

$$:= P\left(F(I-1, J(I-1)) > \frac{SS_B}{I-1} \Big/ \frac{SS_W}{I(J-1)} \right)$$

If p-value $< \alpha$ then reject null hypothesis.

— x —

- Application in least square estimate as well.