Hi !

Everything is ok?
Can you hear me?
See the notes?

$\boxed{\text{Small number of observations } n!}$

→ Welch's test
→ Mann-Whitney test

→ for small n !!

Welch's test

two samples
→ test H_0
→ construct CI

$X_1 \dots X_{n_X}$ - eXperimental
$Y_1 \dots Y_{n_Y}$ - control sample.

$$X_i \sim iid\,(\mu_X, \sigma_X^2)$$
$$Y_i \sim iid\,(\mu_Y, \sigma_Y^2)$$

independent

X_i and Y_i may not be iid!

No equal variances assumed!

$$\text{"}t\text{"} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{se(\bar{X} - \bar{Y})}$$

$$se(\bar{X} - \bar{Y}) = \sqrt{\frac{\hat{\sigma}_X^2}{n_X} + \frac{\hat{\sigma}_Y^2}{n_Y}}$$

use $N(0;1)$ when $n_X >> 0$ and $n_Y >> 0$.

| which distribution should I use for small n_Y n_X?

idea!

$t \sim$ law

is not $N(0;1)$
is not t_{n-1}
n_x and n_y "
is not $t_{n_x+n_y-2}$

estimate the law of distribution!

Some basic facts about χ^2, t distribution

$\boxed{\text{Reminder.}}$

by def $\chi^2_d = Z_1^2 + Z_2^2 + \dots + Z_d^2$ with $Z_i \sim$ iid $N(0;1)$

$$E(\chi^2_d) = E(Z_1^2 + \dots\ Z_d^2) = d \cdot E(Z_1^2) = d$$

$$E(Z_1) = 0 \quad E(Z_1^2) = 1$$

$$Var(\chi^2_d) = Var(Z_1^2 + \dots + Z_d^2) = d \cdot Var(Z_1^2) =$$
$$= d \cdot \left(E(Z_1^4) - (E(Z_1^2))^2\right) = d \cdot (3 - 1^2) = 2d$$

$$E(Z_1^{2k}) = (2k-1) \cdot (2k-3) \cdot \dots \cdot 1$$

Student t distribution.

$$t_d = \frac{N(0;1)}{\sqrt{\chi^2_d/d}} \quad \text{indep-t.}$$

let's assume that $\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{se(\bar{X} - \bar{Y})} \overset{?}{\sim} t_d$

close to

we need only to estimate d with $\hat{d}$

$\boxed{\text{let's estimate } d \text{ under ideal assumptions.}}$

Add. assumption. $X_1 \dots X_{n_X} \sim N(\mu_X; \sigma_X^2)$

$Y_1 \dots Y_{n_Y} \sim N(\mu_Y; \sigma_Y^2)$

use it later !!

$$E\left[\bar{X} - \bar{Y} - (\mu_X - \mu_Y)\right] = 0$$

$$Var\left(\bar{X} - \bar{Y} - \underbrace{(\mu_X - \mu_Y)}_{const.}\right) = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y} - 2 \cdot 0$$

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{Var(\bar{X} - \bar{Y})}} \sim N(0;1) \qquad \xrightarrow[n_X, n_Y \to \infty]{CLT} N(0;1)$$

$$t = \frac{N(0;1)}{\sqrt{\chi_d^2 / d}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \cdot \frac{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}{se(\bar{X} - \bar{Y})}$$

$$se(\bar{X} - \bar{Y}) = \sqrt{\frac{\hat{\sigma}_X^2}{n_X} + \frac{\hat{\sigma}_Y^2}{n_Y}}$$

$$\frac{1}{\sqrt{\chi_d^2 / d}} = \sqrt{\frac{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}{\frac{\hat{\sigma}_X^2}{n_X} + \frac{\hat{\sigma}_Y^2}{n_Y}}}$$

$$\frac{\chi_d^2}{d} \sim \frac{\frac{\hat{\sigma}_X^2}{n_X} + \frac{\hat{\sigma}_Y^2}{n_Y}}{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

let's match $E(\cdot)$ $Var(\cdot)$!!

LHS

$$E\left(\frac{\chi_d^2}{d}\right) = \frac{d}{d} = 1$$

reminder: $E(\hat{\sigma}_X^2) = \sigma_X^2$, $\hat{\sigma}_X^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$

RHS

$$E\left(\frac{\frac{\hat{\sigma}_X^2}{n_X} + \frac{\hat{\sigma}_Y^2}{n_Y}}{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right) = \frac{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}{-//-} = 1 \quad !!$$

LHS

$$\mathrm{Var}\left(\frac{\chi^2_d}{d}\right) = \frac{1}{d^2}\mathrm{Var}\left(\chi^2_d\right) = \frac{1}{d^2}\cdot 2d = \frac{2}{d}$$

RHS

$$\mathrm{Var}\left(\frac{\frac{\hat\sigma_x^2}{n_x}+\frac{\hat\sigma_y^2}{n_y}}{\underbrace{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}_{b}}\right) = \frac{1}{b^2}\left(\frac{1}{n_x^2}\cdot\underline{\mathrm{Var}\left(\hat\sigma_x^2\right)} + \frac{1}{n_y^2}\cdot\mathrm{Var}\left(\hat\sigma_y^2\right)\right)$$

$$\mathrm{Var}\left(\hat\sigma_x^2\right) = \mathrm{Var}\left(\frac{\chi^2_{n_x-1}}{n_x-1}\cdot\sigma_x^2\right) = \frac{\sigma_x^4}{(n_x-1)^2}\cdot\mathrm{Var}\left(\chi^2_{n_x-1}\right) = \frac{\sigma_x^4}{(n_x-1)^2}\cdot 2(n_x-1) = \frac{2\sigma_x^4}{n_x-1}$$

ideal situation $X_i \sim \text{ind } N(\mu_x; \sigma_x^2)$

$$\frac{\hat\sigma_x^2}{\sigma_x^2}\cdot(n-1) \sim \chi^2_{n-1}$$

LHS

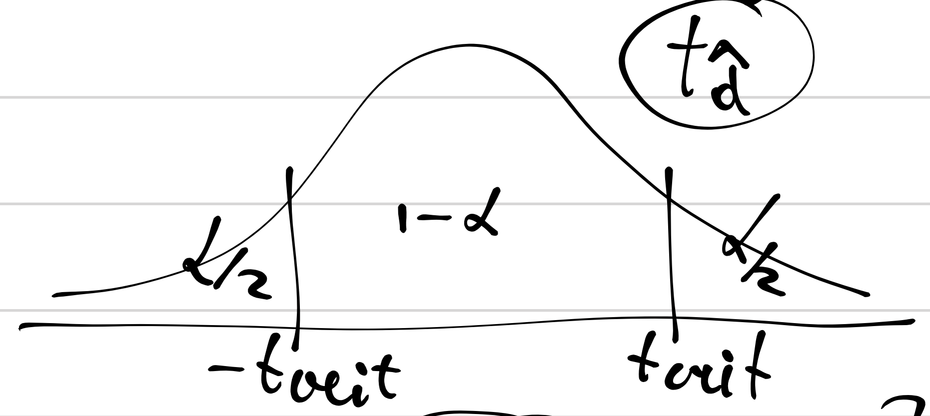
$$\frac{2}{d} = \frac{1}{\left(\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}\right)^2}\cdot\left(\frac{2\sigma_x^4}{n_x^2(n_x-1)} + \frac{2\sigma_y^4}{n_y^2(n_y-1)}\right)$$

$$\boxed{\hat d = \frac{\left(\frac{\hat\sigma_x^2}{n_x}+\frac{\hat\sigma_y^2}{n_y}\right)^2}{\frac{\hat\sigma_x^4}{n_x^2(n_x-1)} + \frac{\hat\sigma_y^4}{n_y^2(n_y-1)}}}$$

welch-Sattertwaite degrees of freed apls ...

$t_{\hat d}$

$\hat d$ may be 4.56



95% CI for $\mu_x - \mu_y$

$$\left[\bar X - \bar Y - t_{crit}\cdot se(\bar X - \bar Y);\ \bar X - \bar Y + t_{crit}\cdot se(\bar X - \bar Y)\right]$$

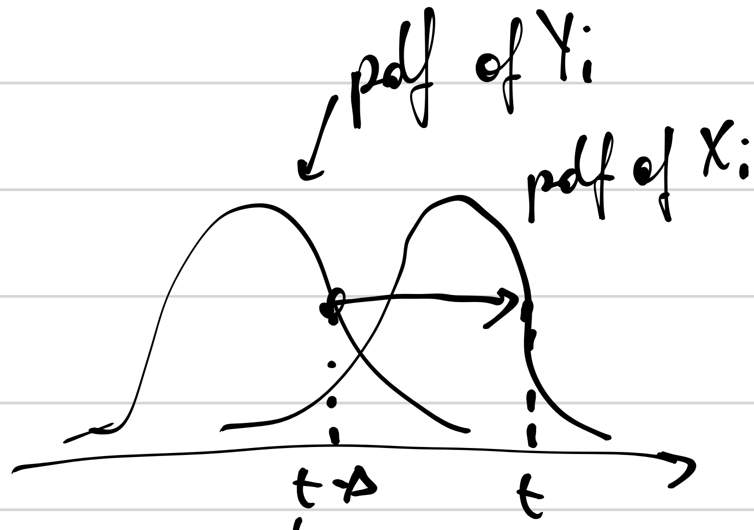
(1, 2, 3, 42 - -)

Mann-Whitney test

Mann-Whitney U test

(+) very low n_X, n_Y

(−) no CI !!



Assumtions.

$X_1, \dots, X_{n_X}$

$Y_1, \dots, Y_{n_Y}$

✓ all observ are iid.

✓ $Y_i \sim pdf_Y(t)$

✓ $X_i \sim pdf_X(t)$

$$\boxed{pdf_X(t) = pdf_Y(t-\Delta)}$$

$H_0: \Delta = 0$

$H_A: \Delta > 0$ / $\Delta < 0$ / $\Delta \neq 0$

the idea of test. (n_X, n_Y are small)

we cannot work with $\bar{X} - \bar{Y}$!

→ CLT is not applicable

→ n_X, n_Y are not enough to estimate the law.

$X_1 \quad Y_2 \quad X_2 \quad Y_3 \quad Y_1 \quad X_3 \quad X_4$

Assume that H_0 is satisf. | $X_i, Y_j \sim$ iid !

$n_X = 4$
$n_Y = 3$

$$P(XYXYYXX) = P(XXXXYYY) = \dots = P(\text{any order}).$$

YYYY XXXXX — „effect"

XYXYXYX — „no effect"

$$U_x = \sum_{i,j} I(X_i > Y_j)$$

all poss. pairs

$$U_x(\underset{0}{X}\,\underset{}{Y}\,\underset{1}{X}\,YY\,\underset{3}{X}\,\underset{3}{X}) \overset{?}{=} +1+3+3=7$$

Ex

$n_x = 2$
$n_y = 3$

a) $U_x \in ?$ $\{0, \ldots, 6\}$

XXYYY   YYYXX
(3 3)

$U_x = 5$
$H_0 : \Delta = 0$
$H_A : \Delta > 0$

b) p-value $= P(U_x \geq 5 \mid H_0)$?

c) conclusion !

|all orders| 5 letters 2X and 3Y

$$C_5^2 = \frac{5!}{2!\,3!} = \frac{5\cdot 4}{2} = 10$$

|favor. orders.| $U_x \geq 5$ |5 or 6.|

$U_x = 6$ YYYXX
$U_x = 5$ YYXYX (2 3)

$$\text{p-value} = \frac{2}{10} = 0.2$$

$\alpha = 0.05$ $\Rightarrow$ H_0 is not rejected

Another method to calculate U_x

theorem

way 1 $\quad U_X = \sum_{i,j} I(X_i > Y_j)$

way 2 $\quad U_X = \sum_i \mathrm{Rank}(X_i) - \frac{n_X(n_X+1)}{2}$

way 1 $\quad \underset{0}{X}\, Y\, \underset{1}{X}\, Y\, Y\, \underset{3}{X}\, \underset{3}{X} \qquad U_X = 7$

way 2 $\quad \underset{1}{X}\, \underset{2}{Y}\, \underset{3}{X}\, \underset{4}{Y}\, \underset{5}{Y}\, \underset{6}{X}\, \underset{7}{X} \qquad U_X = 1+3+6+7 - \frac{4(4+1)}{2} =$

$= 17 - 10 = 7$

base of ind

$n_X + n_Y = 1$

	way 1	way 2
X	$U_X = 0$	$U_X = 1 - \frac{1\cdot(1+1)}{2} = 0$
Y	$U_X = 0$	$U_X = 0 - \frac{0\cdot 1}{2} = 0$

for k letters it works.

$X\,Y \ldots X$

$n_X + n_Y = k$

$U_X^{I} = U_X^{II}$

before:

$\sum I(X_i > Y_j) = \sum \mathrm{Rank}(X_i) - \frac{n_X(n_X+1)}{2}$ (old value)

$+Y$ will not chang, will not change

$XY \ldots X$ $+X$ LHS ↑ by n_Y

$$\text{RHS} \uparrow \text{ by } +(n_X+n_Y+1) - \left(\frac{(n_X+1)(n_X+2)}{2} - \frac{n_X(n_X+1)}{2}\right) =$$

$$= \frac{2\cancel{n_X} + 2n_Y + \cancel{2} - \cancel{n_X n_X} - \cancel{2} - 3\cancel{n_X} + \cancel{n_X n_X} + \cancel{n_X}}{2} =$$

$$= n_Y$$

theor: $\quad E(U_X) = \frac{n_X \cdot n_Y}{2} \qquad \mathrm{Var}(U_0) = \frac{n_X n_Y (n_X + n_Y + 1)}{12}$

$$\frac{U_X - E(U_0)}{\sqrt{\mathrm{Var}(U_0)}} \to N(0;1)$$