Error & Power in hypothesis testing: Types of Errors:

Type I: Reject H_0 when TRUE — False positive — $P(\text{Rej } H_0 | H_0 \text{ True}) = \alpha$

Type II: Failed to reject H_0 when H_0 False — $P(\text{FTR } H_0 | H_0 \text{ False}) = \beta$
(FTR)

Reality

Decision	H_0 True	H_0 FALSE
FTR H_0	specificity $(1-\alpha)$ True N	Type II False Neg β
Rej H_0	Type I α (FP)	Power, sensitivity (TP) $(1-\beta)$

Prob (Type I) $= \alpha$ (we choose α)

Prob (Type II) = Depends on α, sample size, & diff. b/w hypothesis: $\mu_A \sim \mu_0$

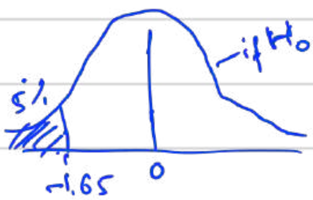
Power of Test:

Hypertensive pop: $\mu = 160$, $\sigma = 36$

Drug test $n = 25$, $\alpha = 0.05$

$H_0: \mu = 160$

$H_A \; \mu \leq 160$



$\alpha = P(\text{Rej } H_0 | H_0 \text{ True})$

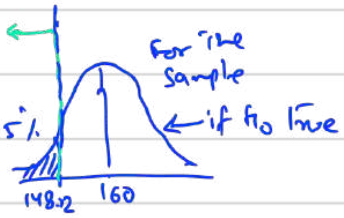
$\beta = P(\text{FTR } H_0 | H_0 \text{ FALSE or } H_A \text{ TRUE})$

Power = Prob. of rejecting null if alternative is true
$= P(\text{Rej } H_0 | H_A \text{ TRUE})$

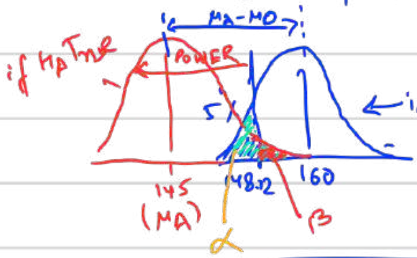
$Z_{stat} < -1.65 \Rightarrow$ Rej H_0

$$Z_{stat} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{\bar{X} - 160}{36/\sqrt{25}}$$

$$-1.65 = \frac{\bar{X} - 160}{36/\sqrt{25}} \Rightarrow \bar{X} = 148.12 \Rightarrow$$



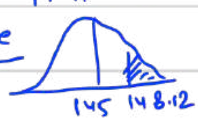
lets overlay if H_A True, let's assume $\mu_A = 145$

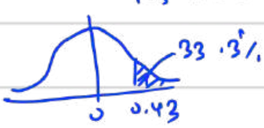


$\beta = P(\bar{X} > 148.12 | \mu_A = 145)$

$$Z_\beta = \frac{148.12 - 145}{36/\sqrt{25}} = 0.43$$

standardize





$\uparrow \alpha - \downarrow \beta$ ($\uparrow$ Power)

$\uparrow n - \downarrow$ s.d. $- \downarrow \beta$ ($\uparrow$ Power)

$\uparrow \mu_A - \mu_0 - \downarrow \beta$ ($\uparrow$ Power)

i.e. large effect of drug

So, Power $= 1 - 33.3$
$= 66.7\%$

So, if H_A is true, drug lowers blood pres. to 145 on average. There's a 33% chance that we fail to reject H_0

BIVARIATE ANALYSIS

Approaches
- Parametric
  - Lot of Assumptions, nice mathematical formulation
  - large sample size
  - higher power
  - sensitive to outliers
- Non parametric
  - Generally work with ranking data

Ex: X - Drug A, Drug B, - Categorical
Y - ΔSBP (Sys. blood pressure change) - Numerical
Analysis tools
- t-Test 2 sample
- one way ANOVA
- Wilcoxin ...

Ex. X - Smoker - Categorical
Y - Cancer - Categorical
Analysis
- Pearson χ^2 test
- Fischer Exact test

Ex: X - Years of education Numeric
Y - Salary Numeric
Analysis
- correlation
- Lin Reg.

X-cat Y-Cat :

Paired / Dependent
- Before / After Treatment
- Left / Right - left side - Right side
- Cross over design Treatment A then after gap Treatment B
- Matching - 2 groups (A, B) then match on other variable (sex, Age
- Twins

Independent
- Different people on Treatment Smoker / Non smoker
- Exposed / Non-exposed
- Male / Female

- ↓ Biological variability
- matching is subjective and not always possible

CONSIDERATIONS:
- Must ensure groups are similar
- Randomize
- Adjustment betn groups
- Simple MATH

- So what are we analysing

	2 Paired	3 or more Paired	2 Independent	3+ independent
Parametric	Paired t-test	Repeated Measure ANOVA	2 sample t-test	One-Way ANOVA
Non Parametric	Wilcoxin signed Rank	Friedman's test	Rank Sum (Mann-Whitney	KRUSKAL WALLIS

Paired t-test: Compares mean of 2 paired groups / dependent groups
- Same individual
- or Diff. individuals matched on Age, gender etc.

Assumption
- independent observation
- Paired
- Large sample
- ΔBP is normally distributed

$H_0 : \mu_A = \mu_B$ or, $H_0 : \mu_A - \mu_B = 0$

$H_1 : \mu_A < \mu_B$ or, $H_1 : (\mu_A - \mu_B) < 0$

Now $\mu_A - \mu_B$ using ΔBP is a single variable

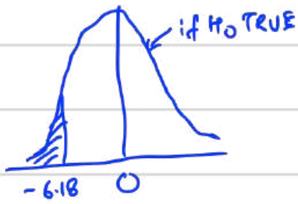
Mean $\bar{d} = \frac{\Sigma \Delta BP}{n} = -6.18$

std. dev. $sd = 8.76$

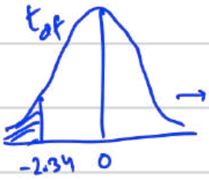
Pair	Before	After	ΔBP
1	135	127	-8
2	142	145	+3
3	137	131	-6
⋮			

Now; we can write $H_0 : \mu_d = 0$ &

$H_1 : \mu_d < 0$



$$t_{STAT} = \frac{\bar{d} - \mu_d}{SE_{\bar{d}}} = \frac{-6.18 - 0}{8.76/\sqrt{11}} = -2.34$$



→ pvalue = 0.0207 = 2.07% If H_0 is true there is 2% chance of getting a value that is less by 6.18

Reject H_0.

Now C.I (this is a 2 sided CI; but for now it's ok)

$95\% = EST \pm t^* SE_{est}$

$= -6.18 \pm (2.22) \frac{8.76}{\sqrt{11}} = (-12.1, -0.30)$ this is CI for ΔBP change

↑ t for 95% — 0 is not there

WILCOXIN SIGNED TEST :
- small sample size
- if we want to look at median change
- do not assume normality

H_0 : median diff $= 0$

H_1 : m.d < 0

If H_0 is true
50% show decrease
50% should show inc.
5.5 individuals of 11

Pair	Before	After	ΔBP	
1	135	127	-8	← Decrease
2	142	145	+3	← Increase
3	137	131	-6	← Decrease
⋮			⋮	
11				

In total 8 dec. 3 inc.

P-value = Prob. of 8 or more dec. while should be 5.5

~ Binomial dist: $P(X \geq 8 \mid X \sim BIN, n = 11, P = 0.5)$ 50%

$= 0.11$

Another Example: say 4 dec., 3 inc. of 7 people. Expect 3.5

-20, +1, -18, +2, -10, +1.5, -14 ← note there are people that show a large dec. in BP after treatment but the extent is ignored, just the count matters - DRAWBACK

WILCOXIN SIGNED RANK TEST:

OBS : $-20, +1, -18, +2, -10, +1.5, -14$
RANK: 7 1 6 3 4 2 5

Tie then average RANK
$-20, +1, +1, +2$
4 1.5 1.5 3 RANK

OBS. $4\downarrow, 3\downarrow$
Exped: $3.5\downarrow, 3.5\uparrow$

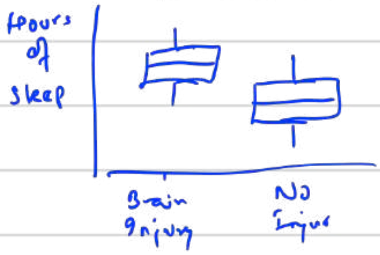
For RANK SUM the RANK: $(1+2+\ldots+7)/2 = 14$ ← for 0.5 or 50%

so, expect: $14\downarrow, 14\uparrow$
Obs : $22\downarrow, 6\uparrow$

P-val: Prob sum$\downarrow \geq 22$ if H_0 True (i.e. 14)

INDEPENDEN 2 Sample t-TEST:

mean of 2 independent
X is categorical
Y- numerical



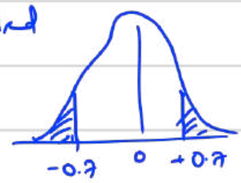
$\bar{Y}_{inj} = 8.1, \; S_{inj} = 0.7, \; n_{inj} = 20$
$\bar{Y}_{no} = 7.4, \; S_{no} = 0.9, \; n_{no} = 25$

We estimate $EST = (\bar{Y}_{inj} - \bar{Y}_{no}) = 8.1 - 7.4 = \underline{0.7}$

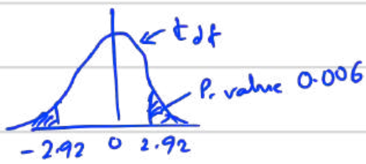
SE_{EST} = for now assume 0.24 assuming pop. mean μ_1 for inj & μ_2 for no injury.

$H_0 : (\mu_1 - \mu_2) = 0$
$H_1 : (\mu - \mu_2) \neq 0$

let's do 2 sided



$t_{STAT} = \frac{(\bar{Y}_1 - \bar{Y}_2) - 0}{SE_{\bar{Y}_1 - \bar{Y}_2}} = \frac{0.7 - 0}{0.24} = 2.92$

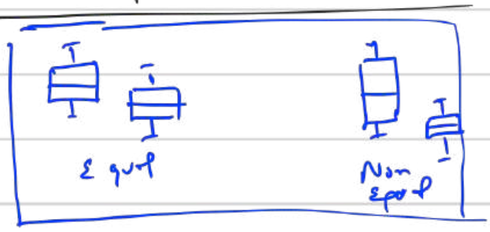


95% CI : $(\bar{Y}_1 - \bar{Y}_2) \pm t^* SE_{(\bar{Y}_1 - \bar{Y}_2)}$

$= 0.7 \pm 2(0.24) = (0.22, 1.18)$ Sleep 0.22 to 1.18 hrs. more

← 0 is not here

Equal Vs. Non Equal Variance Assumption



① $SD_{Large}/SD_{small} > 2 \Rightarrow$ Non Equal

② $H_0: \sigma_1 = \sigma_2$ $H_1: \sigma_1 \neq \sigma_2$
(at pop: level)

③ Tests: Levene's test
Bartlett's test

Eg.

$SD(\bar{Y}_1) = \frac{s_1}{\sqrt{n_1}} \Rightarrow Var(\bar{Y}_1) = \frac{s_1^2}{n_1}$

also called $SE(\bar{Y}_1)$

$SD(\bar{Y}_2) = \frac{s_2}{\sqrt{n_2}} \Rightarrow Var(\bar{Y}_2) = \frac{s_2^2}{n_2}$

Property of variance: $\{ Var(X-Y) = Var(X) + Var(Y)$, if X, Y independent

So,

Non Equal Variance

$$Var(\bar{Y}_1 - \bar{Y}_2) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

$$\Rightarrow SD_{\bar{Y}_1 - \bar{Y}_2} \text{ or } SE_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

d.f: $\min((n_1 - 1, n_2 - 1), n_1 + n_2 - 2)$

Equal Variance: take pooled estimate $\Rightarrow$ weighted average

i.e. 2 difference estimates of same thing, pool them

$$s^2_{pooled} = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$SE_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{s^2_{pooled}}{n_1} + \frac{s^2_{pooled}}{n_2}}$$ with d.f $= n_1 + n_2 - 2$

Bootstrap Hypothesis Testing:

Eg.

X-Feed type	Y weight
M	325
M	257
M	303
C	390
C	260

tests one could do:

- two sample t-test : mean of 2 groups M & C
- Wilcoxon R S test : median of 2 groups (Mann Whitney U test)
- Bootstrapping - small sample size
  - Difficult to measure S.E for test-statistic

H_0: weight same

H_1: not same

Bootstrapping: Pick any number and randomly assign to label

eg. Sample #1

X-Feed type	Y Weight		#1		#2 ...
M	$\bar{Y}_{m\,obs}$	325	260 M	$\bar{Y}_{mB1} = 330$	
M		257	303 M		
M		303	257		
C	$\bar{Y}_{c\,obs}$	390	⋮		
C		260	325 C	$\bar{Y}_{cB1}$	

Let's say

Test-statistic $= |\bar{Y}_c - \bar{Y}_M| = EST$

p-value will be $= \dfrac{\#\,BS\text{-}test\text{-}static \geq obs.test.stat}{total\ B}$

$EST = \bar{Y}_c - \bar{Y}_M = 349.25 - 316 = 33.25$

$EST_{B1} = 352 - 330 = 22$

$EST_{B2} = 379 - 315 = 64 \leftarrow$ of 2 BS only one is ≥ 33.25 & total $B = 2$

——— x ———

Permutation approaches to hypothesis testing:

When
- Small sample
- Parametric approaches not suitable
- test something other than mean/median eg. Range

From above $EST_{mean} = 33.25$

& $|Med_c - Med_M| \rightarrow EST_{median} = 58.5$

Permutation - all possible permutation of data independent of feed type

* Bootstrap is with replacement
Permutation is shuffling. obs. not repeated

$$p\text{-Value} = \frac{\#\,Permutation\ Test\ stat > obs.\ test\ stat\ (EST)}{P \rightarrow (no.\ of\ permutations)}$$