Statistical Testing

1) Sample size
2) Effect size
3) Significance level
4) Power

$H_0 : \rho = 0$

$\hat{\rho}_P$ – Pearson correlation

$$\hat{\rho}_P = \frac{\widehat{cov}(x,y)}{\sqrt{\widehat{Var}(x) \cdot \widehat{Var}(y)}}$$

$\hat{\rho}_s$ – Spearman correlation

$$\hat{\rho}_s = \frac{\widehat{cov}(rank(x), rank(y))}{\sqrt{\widehat{Var}(rank(x)) \cdot \widehat{Var}(rank(y))}}$$

$[H_0 : \mu_a = \mu_b]$ $\quad \theta = \frac{\mu_a - \mu_b}{\sigma}$

$$\Rightarrow d = \frac{\bar{x}_a - \bar{x}_b}{s}$$

Cohen's d

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

$se(\mu_a - \mu_b)$

$H_0 : Med_A = Med_B$

t – test :

$H_0 : \mu_a = \mu_b$

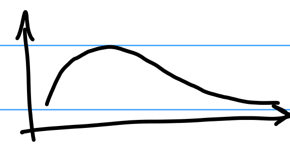
CLT

sample is large enough?

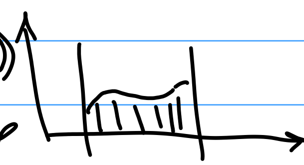
$X_a \sim N(\mu_a, \sigma^2)$; $X_b \sim N(\mu_b, \sigma^2)$

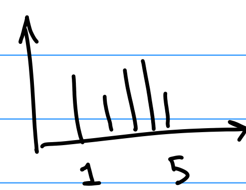
$\Rightarrow$ 20 obs.

X_a



X_b





$\Rightarrow$ > 200 obs

$\Rightarrow$ Welch test

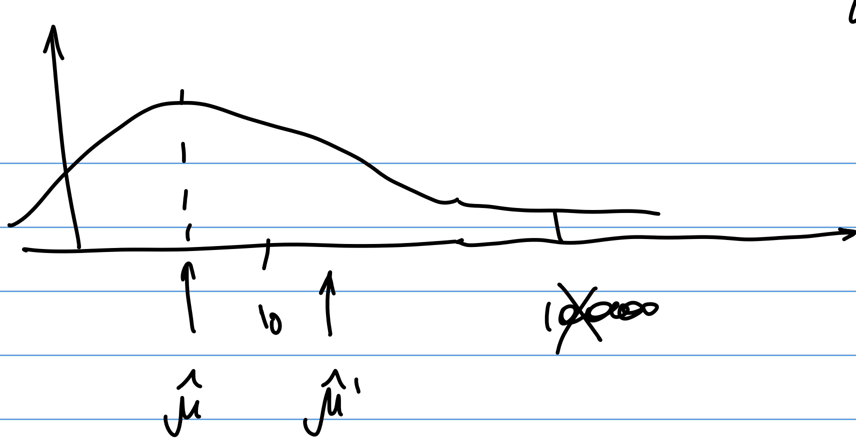
$H_a : \mu_a = \mu_b,$

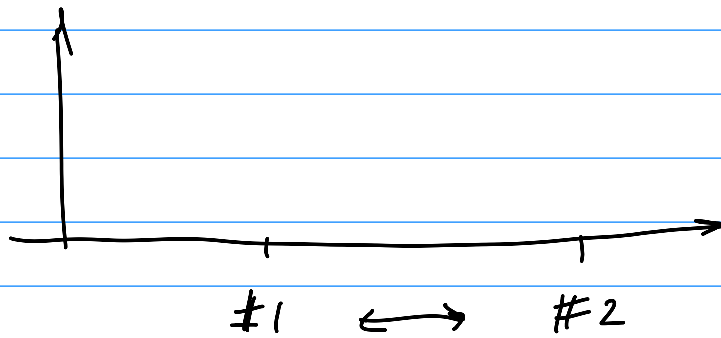
$\sigma_a \neq \sigma_b$

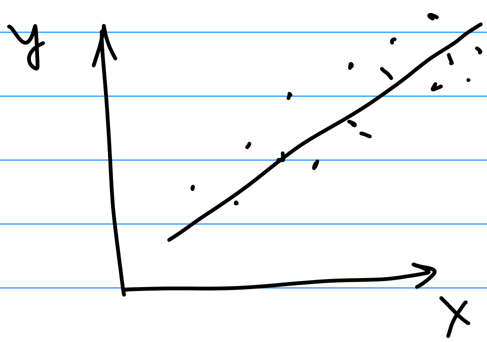
$\Rightarrow$ MW U-test

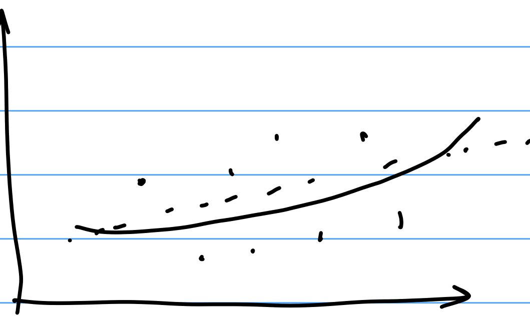
robust to skewed data

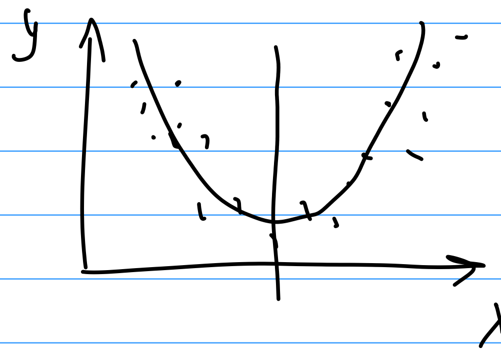
& outliers











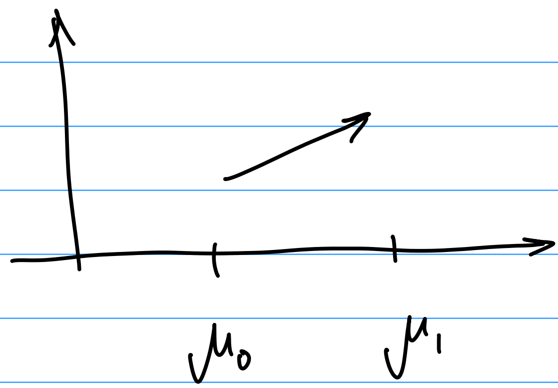
Example : $y_i = \alpha + \beta \cdot D_i + \varepsilon_i$

$$D_i = \begin{cases} 1, & \text{if treatment was given} \\ 0, & \text{if no treatment was given} \end{cases}$$

treatment group

control group

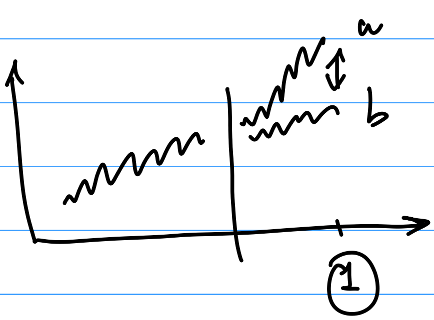
↑ treatment variable



μ_1 vs μ_0

increase ?

- inflation
- economic growth
- wage increased



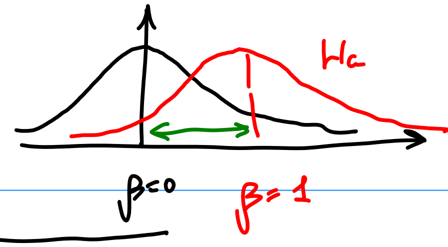
$$Var(\hat{\beta}) = \frac{1}{p(1-p)} \cdot \frac{\sigma^2}{N}$$

↑ p – fraction of obs in treatment group

σ^2 – Var of y

N – sample size

$H_0: \beta = 0$



$$MDE = \left(t_{1-k} + t_\alpha\right) \cdot \sqrt{\frac{1}{p(1-p)} \cdot \frac{\sigma^2}{N}}$$

power sign. level sample size

$$\text{Power} = \frac{\beta_E}{\sqrt{\frac{1}{p(1-p)} \cdot \frac{\sigma^2}{N}}} - t_\alpha = \underbrace{t_{1-k}}_{\text{power}}$$

effect size

$$N = \left[\frac{\sigma \cdot \left(t_{(1-k)} + t_\alpha\right) \cdot \sqrt{\frac{1}{p(1-p)}}}{\beta_E}\right]^2$$