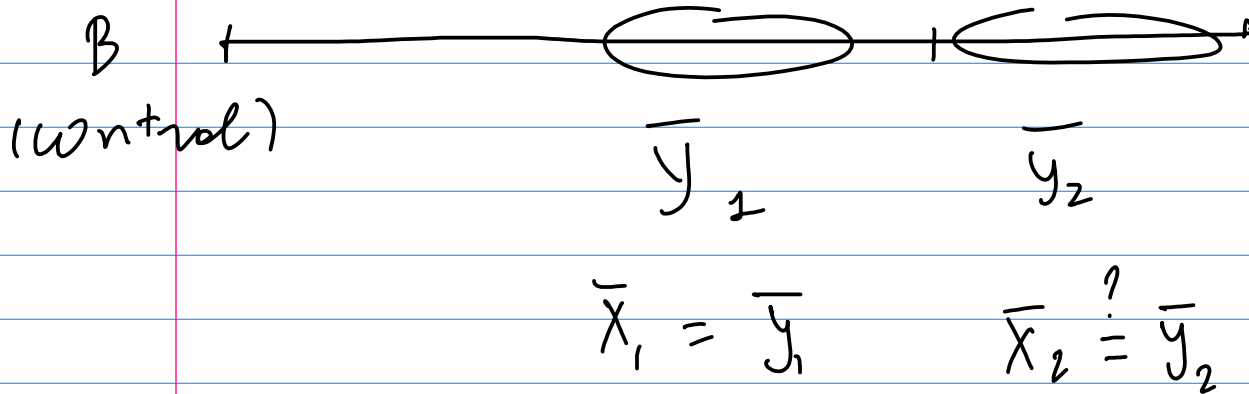
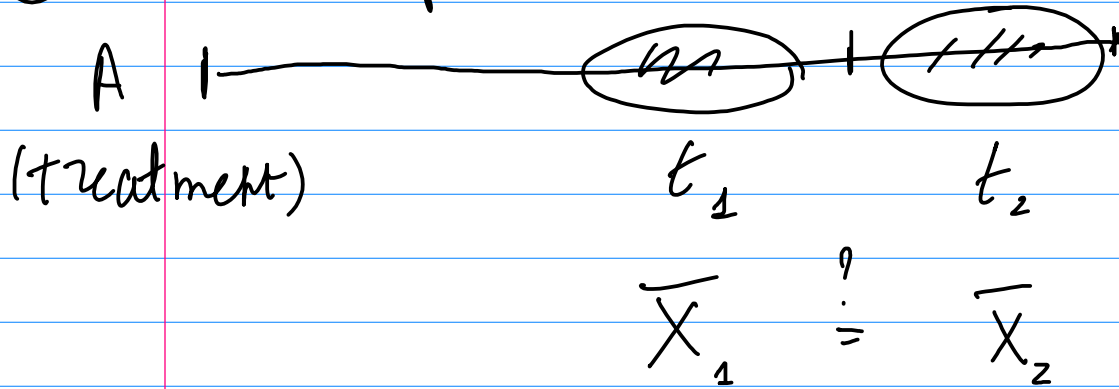
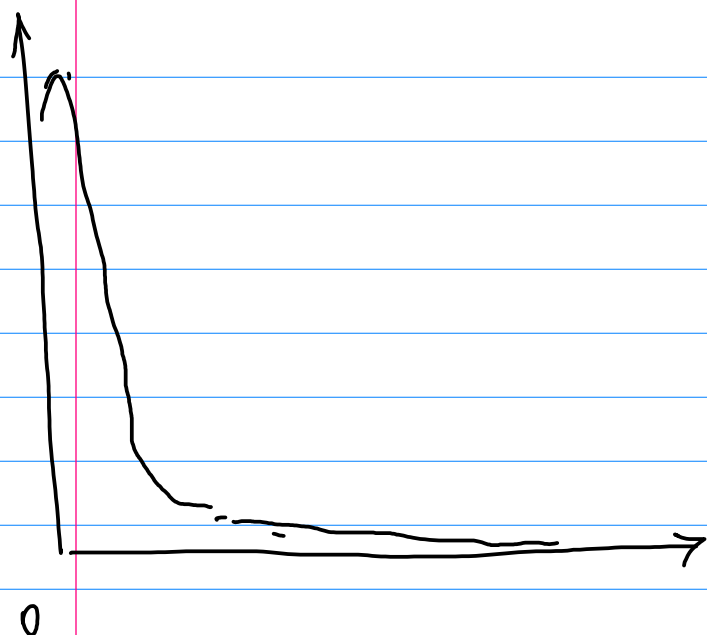


# A/B Testing

① Time effect

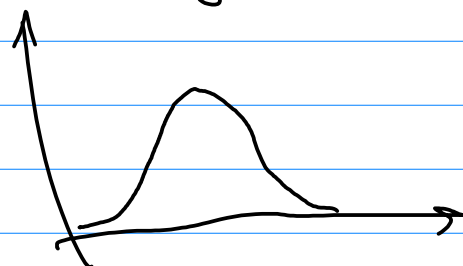


② Skewed data, not normal



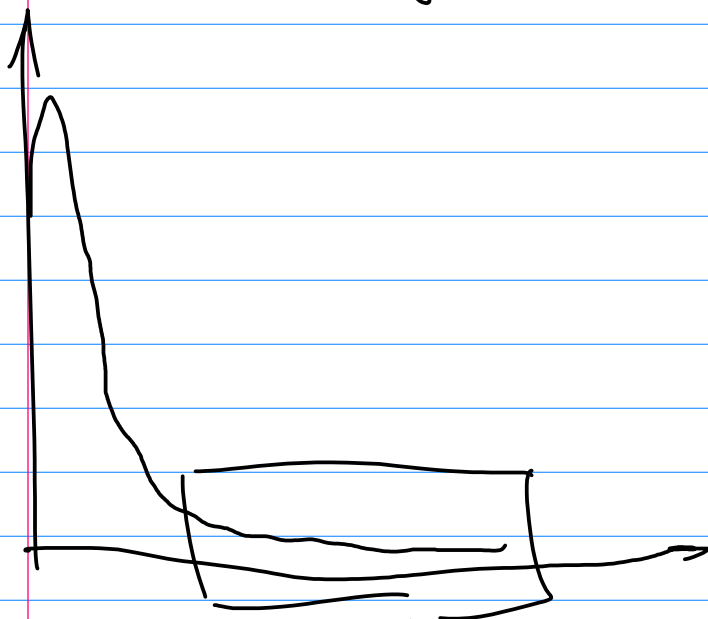
Transform data:

- logarithms

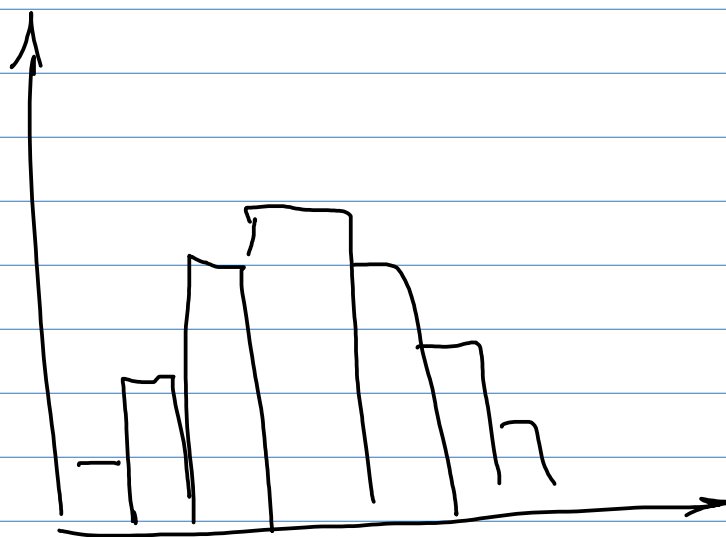


③ Very big data:

- Bucketing



$$a_j^b = \sum_{i=1}^{N_j} x_{ij}$$



$N_j$  - # obs in bucket  $j$

$$\frac{s^2}{N} \approx \frac{s_b^2}{B}$$

$$j = \overline{1, B}$$

# Types of metrics

1) User-level

$$M_1 = \overline{X}$$

e.g.

$$\frac{3+0+5+\dots+10}{1+1+1+\dots+1}$$

2) Ratio

$$M_2 = \frac{\sum X}{\sum Y}$$

$$\frac{2+0+\dots+5}{10+2+\dots+5}$$

$$\frac{X}{Y}, X, Y - \text{r.v.}$$

$$R = X/Y \quad E(R) - ? \quad \text{Var}(R) - ?$$

Linearization

Taylor expansion of  $R = X/Y$   
at  $[E(X), E(Y)]$

$$Z = \frac{E[X]}{E[Y]} + \frac{1}{E[Y]} \left( X - \frac{E[X]}{E[Y]} Y \right)$$

$$E(Z) \approx E[R] \quad \text{Var}(Z) \approx \text{Var}(R)$$

$\Rightarrow$  t-tests or Welch test

Alternative: delta-method

$H_0$

Decision	True	False
	TN (Confidence)	FN (Type I error)
Don't Reject	$1 - \alpha$	$\beta$
Reject	FP (Type I error)	TP (Power)
	$\alpha$	$1 - \beta$

FPR  $\rightarrow$  t-test  
 $\searrow$  Welch-test  
 $\searrow$  Mann-Whitney

1) t-test

2) Welch test

• two independent samples

•  $S_x = S_y$   
 $n_1 \neq n_2$

•  $s_x \neq s_y$

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{s_x^2 + s_y^2}} \sim t_d$$

$$n_1 = n_2 = n \Rightarrow$$

$$S_p = \sqrt{\frac{s_x^2 + s_y^2}{2}}$$

$$t = \frac{\bar{X} - \bar{Y}}{S_p \cdot \sqrt{2/n}} \sim t_{2n-2}$$

3) Non-parametric test

$$U = \sum_i \sum_j \phi(x_i, y_j)$$

$$\phi(x_i, y_j) = \begin{cases} 1, & x_i < y_j \\ 0, & \text{else} \end{cases}$$

Levene's test:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2, \quad k - \# \text{ subgroups}$$

$$H_a: \exists i, j \quad \sigma_i^2 \neq \sigma_j^2$$

equivalent to ANOVA

$$W = \frac{n-k}{k-1} \frac{\sum N_i (\bar{z}_{i.} - \bar{z}_{..})^2}{\sum \sum (z_{ij} - \bar{z}_{i.})^2} \sim F(k-1, n-k)$$