1) what do we want to do
- Associations of X & Y
- Patterns
- Trends

2) Characterise X & Y
- X is cts / discrete / categorical
- Y cts / discrete — regression
  categorical — classification / logistic Regression
  (time / censoring) — Survival Analysis

3) what Model

A) Distribution model:
= represent / capture a pattern in the randomness / variability in the variable Y

↳ $E[Y]$, $Var[Y]$
→ $Y \sim Bin(n, p)$

B) Structure in (X, Y)

- If I change X, how does that affect Y

- Linear Model

$$Y = \underbrace{a + bx}_{B} + \underbrace{\varepsilon}_{A}$$

In $y = a + bx$
$E[y] = a + bx$
$var(y) = var(\varepsilon)$

$a + bx$ is fixed (deterministic)
ε is the only random term (R.V)

and thus Y is a R.V