Deterministic vs Stochastic

- Make sure to go over the typed notes too!
- Okay let's recap. We have:
  - populations, population units
  - Sample, subset of population
  - Precise parameter, and statistic that estimates it
  - Types of variables
  - Statistical inferences like point estimation, confidence interval for range of values
  - Hypothesis testing: Assume something is true, see how unlikely your predictor is, accept or reject
  - If something lies outside of an $n\%$ confidence interval that's equivalent to rejecting null hypothesis for that result
- Notational convention:
  - Underline: A vector, ex $\underline{a}, \underline{b}$
  - Uppercase: Matrix, ex A, B
  - hat: Statistic, eg $\hat{\theta}$
  - $\underline{a}^T, A^T$: Transpose
- Unless otherwise specified, vectors are column vectors and transposes are row bois
- Ex 1: $Y = x^2$
- That's deterministic. Same x, you get same Y value, over and over, no matter how many times you do it
- Ex 2: $Y = a + bx, a, b$ are known. a is intercept, b is slope
- a is the default output if 0 is input, and b tells you how much the output changes when the input increases by one unit
- Important: When you talk about slope you *increase* x. It does **not** "change". It **increases**
- This straight line is also deterministic. AKA mathematical
- In general:

$$Y = g(x), x \text{ is non-random }, Y \text{ is also non-random}$$

- Problem: Particle boards are made. To examine this, boards produced at diff temperatures. Strength y, temperature t.

- Will points be exactly on a curve? Probably not. Some randomness/stochastic stuff will happen
- So you make a scatter plot
- In general:

  $Y = g(x)+\epsilon, x$ is non-random, ϵ is random error term, $, Y$ is therefore now random

- Y is the response/dependent variable, x is the predictor/regressor/independent variable/covariate
- Now: Impose assumptions on ϵ: They are IID with mean $E(\epsilon) = 0, V(\epsilon) = \sigma^2 < \infty$
- From that it follows:

$$E(Y) = E(g(x)) + E(\epsilon) = g(x) + 0 = g(x)$$

$$V(Y) = \sigma^2 + V(g(x)) = \sigma^2 + 0 = \sigma^2$$

- So what are we trying to find or achieve? $g(x)$ I'm pretty sure
- Yup $g(x)$
- Your mission, if you choose to accept it: Estimate g
- Let's relax the formal definition
- aaaaaaaaahhhh relaxing
- In general: We might have several regressors like a billion or so

$$x_1, x_2, \ldots, x_k$$

$$Y = g(x_1, x_2, \ldots, x_k) + \epsilon$$

- Simple case: Assume something about g. The parametric form!
- Parametric form $\Rightarrow$ functional form of g is known to us, apart from particular parameters $\underline{\beta}$

$$Y = g(x_1, x_2, \ldots, x_k; \underline{beta}) + \epsilon$$

- For instance $g = a + bx$ where we don't know a, b
- More simple: Functional form of g is linear:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

- That's the multiple linear regression model

- The unknown parameters are:

$$\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^T; \sigma^2$$

$$\Rightarrow (k+2) \text{ unknown parameters}$$

- We will work with this a lot
- The simplest case of multiple linear regression: $k = 1$

$$Y = \beta_0 + \beta_1 x + \epsilon$$

- Intercept, slope, error (with mean 0). So Y is an $RV, E(Y) = \beta_0 + \beta_1(x), V(Y) = \sigma^2 < \infty$
- This is the population regression model, and it's unknown to us because we don't know $\underline{\beta}$ or σ
- So take a sample out of the population and estimate the parameters, and then we can use our estimated regression model, we can do something
- So what are the β boiz? The slope and the intercept?

$$\beta_0 : \text{ Value of } E(Y) \text{ if } x = 0$$

$$\Rightarrow \text{ Mean of distribution of } Y \text{ when } x = 0$$

$$\beta_1 : \text{ Amount of change in } E(Y) \text{ by unit increase in } x$$

- Always mention the mean of Y when thinking about these things!
- Sample regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$$

- Target: Estimate $\beta_0, \beta_1, \sigma^2$ based on sample data
- How? With least-squares ya dummy!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!1