Probability Basics/Distributions, Sampling, Linear Regression

Population Measures

$$\text{Mean } \mu = \frac{1}{n}\sum x_i$$

$$\text{Variance } \sigma^2 = \frac{1}{n}\sum (x_i - \overline{x})^2$$

$$\text{Standard Deviation } \sigma = \sqrt{\frac{1}{n}\sum (x_i - \overline{x})^2}$$

Sampling

$$\text{Sample mean } \overline{x} = \frac{1}{n}\sum x_i$$

$$\text{Sample variance } s_x^2 = \frac{1}{n-1}\sum (x_i - \overline{x})^2$$

$$\text{Std. Deviation } s_x = \sqrt{\frac{1}{n-1}\sum (x_i - \overline{x})^2}$$

$$\text{z-score } z = \frac{x - \mu}{\sigma}$$

$$\text{Correlation } r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{(x_i - \overline{x})}{s_x}\right)\left(\frac{(y_i - \overline{y})}{s_y}\right)$$

Linear Regression

$$\text{Line } \hat{y} = a + bx$$

$$b = r\frac{s_y}{s_x}, a = \overline{y} - b\overline{x}$$

$$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y})^2}$$

$$SE_b = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

$$\text{To test } H_0 : b = 0, \text{ use } t = \frac{b}{SE_b}$$

$$CI = b \pm t^* SE_b$$

Counting and Probabilities

$${}_nP_x = \frac{n!}{(n-x)!} \quad \text{Permutations}$$

$${}_nC_x = \frac{n!}{x!(n-x)!} \quad \text{Combinations}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{Conditional Probability}$$

$$P(A \cap B) = P(A \mid B)P(B) \quad \text{Probability of an Intersection}$$

Discrete Probability Distributions

$$P_x(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \quad \text{Binomial Probability}$$

$$P_x(x) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{Poisson Probability}$$

Continuous Probability Distributions

Random Variable ~ Distribution (mean, variance)

Standard Normal	$Z \sim N(0,1)$
Normal	$X \sim N(\mu, \sigma^2)$
Binomial	$X \sim \text{Binomial}\,[np, np(1-p)]$
Sample Mean	$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
By CLT, if n ≥30,	$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
Sample Proportion	$\overline{p} \sim \left(p, \frac{p(1-p)}{n}\right)$