

1. Descriptive Statistics (Week 2)

Context: Used to organize, summarize, and present data.

- **Mean (Arithmetic):**

- **Population Mean (μ):** $\mu = \frac{\sum_{i=1}^N x_i}{N}$

- **Sample Mean (\bar{x}):** $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

- **Definition:** The average of the observations, computed by summing all observations and dividing by the number of observations.

- **Measures of Variability:**

- **Range:** Range = Largest observation – Smallest observation

- **Population Variance (σ^2):** $\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$

- **Sample Variance (s^2):** $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$

- **Note:** Divided by $n - 1$ to be an unbiased estimator.

- **Sample Standard Deviation (s):** $s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$

- **Coefficient of Variation (CV):** $CV = \frac{s}{\bar{x}}$ (Sample) or $CV = \frac{\sigma}{\mu}$ (Population)

- **Measures of Relative Standing:**

- **Percentile (P -th):** The value for which P percent of observations are less than that value and $(100 - P)$ percent are greater.

- **Location of Percentile (L_P):** $L_P = (n + 1) \frac{P}{100}$

- **Interquartile Range (IQR):** $IQR = Q_3 - Q_1$ (Spread of the middle 50%)

2. Probability Basics (Week 3)

Context: Calculating the likelihood of events.

- **Complement Rule:** $P(A^c) = 1 - P(A)$
- **Addition Rule (Union):**

- General: $P(A \cup B) = P(A) + P(B) - P(AB)$
- Mutually Exclusive Events: $P(A \cup B) = P(A) + P(B)$
- $P(A|B) = \frac{P(AB)}{P(B)}$ if $P(B) > 0$
- **Conditional Probability:**
- **Multiplication Rule (Intersection):**
 - General: $P(AB) = P(B)P(A|B)$ or $P(AB) = P(A)P(B|A)$
 - Independent Events: $P(AB) = P(A)P(B)$
- **Independence:** Two events A and B are independent if $P(AB) = P(A)P(B)$ or equivalently $P(A|B) = P(A)$.
- Bayes' Theorem (Formula):

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

3. Random Variables and Probability Distributions (Week 4 & 5)

Context: Modeling discrete and continuous outcomes.

General Formulas (Discrete)

- **Expected Value (Mean) μ :** $E[X] = \sum xP(x)$
- **Variance σ^2 :** $V(X) = \sum (x - \mu)^2 P(x) = E[X^2] - (E[X])^2$

Discrete Distributions

- **Binomial Distribution:**
 - PMF: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ for $x = 0, 1, \dots, n$
 - Mean: $E[X] = np$
 - Variance: $Var(X) = np(1 - p)$
- **Poisson Distribution:**

- PMF: $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$ for $x = 0, 1, 2, \dots$
- Mean: $E[X] = \lambda$
- Variance: $Var(X) = \lambda$

Continuous Distributions

- **Uniform Distribution (on a, b):**

- PDF: $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$
- Mean: $E[X] = \frac{a+b}{2}$
- Variance: $Var(X) = \frac{(b-a)^2}{12}$

- **Normal Distribution:**

- PDF: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$
- Standardizing: $Z = \frac{X - \mu}{\sigma}$

- **Exponential Distribution:**

- PDF: $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$
- Mean: $E[X] = \frac{1}{\lambda}$
- Variance: $Var(X) = \frac{1}{\lambda^2}$

4. Sampling Distributions (Week 6)

Context: Properties of sample statistics (like the sample mean).

- **Sampling Distribution of the Sample Mean (\bar{X}):**

- Mean: $\mu_{\bar{x}} = \mu$
- Standard Deviation (Standard Error): $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

- **Sampling Distribution of a Proportion (\hat{P}):**

- Mean: $E(\hat{P}) = p$

- Standard Error: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
 - **Central Limit Theorem:** If samples of size n are drawn from a population with mean μ and variance σ^2 , the sampling distribution of \bar{X} is approximately normal if n is large ($n \geq 30$), regardless of the population shape.
-

5. Estimation (Week 7)

Context: Estimating population parameters using confidence intervals.

- Confidence Interval for Mean (σ known):

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Confidence Interval for Mean (σ unknown):

$$\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

Uses Student's t-distribution with $n - 1$ degrees of freedom.

- Confidence Interval for Proportion (p):

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

6. Hypothesis Testing (Week 8)

Context: Testing claims about population parameters.

- **Test Statistics for Mean:**

- σ known: $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

- σ unknown: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

- Test Statistic for Proportion:

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

- **Errors:**
 - **Type I Error (α):** Rejecting a true null hypothesis.
 - **Type II Error (β):** Failing to reject a false null hypothesis.
-

7. Comparing Two Means and Proportions (Week 9)

Context: Inferences about the difference between two population parameters.

- **Difference of Two Means (Independent Samples, σ_1, σ_2 unknown):**
 - Equal Variances Assumed (Pooled):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.

- Unequal Variances (Separate):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- **Matched Pairs (Paired t-test):**

$$t = \frac{\bar{d} - \mu_D}{s_D / \sqrt{n_D}}$$

where \bar{d} and s_D are the mean and standard deviation of the differences.

- **Difference of Two Proportions:**

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where \hat{p} is the pooled proportion $\frac{x_1 + x_2}{n_1 + n_2}$.

8. Linear Regression and Correlation (Week 11)

Context: Analyzing relationships between variables.

- Sample Correlation Coefficient (r):

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n - 1) s_x s_y}$$

- Simple Linear Regression Equation:

$$\hat{y} = b_0 + b_1 x$$

- **Least Squares Estimators:**

- Slope: $b_1 = \frac{s_{xy}}{s_x^2} = r \frac{s_y}{s_x}$

- Y-Intercept: $b_0 = \bar{y} - b_1 \bar{x}$

- **Coefficient of Determination (R^2):** $R^2 = r^2$ (measures the proportion of variation in Y explained by X).
-

9. Chi-Squared Tests (Week 9/11 topic)

Context: Analysis of nominal/categorical data.

- Chi-Squared Statistic (χ^2):

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

where f_o is the observed frequency and f_e is the expected frequency.

- Contingency Table Expected Frequency:

$$e_{ij} = \frac{(\text{Row } i \text{ Total}) \times (\text{Column } j \text{ Total})}{\text{Sample Size}}$$