Chapters 1-6:

Population vs. sample → Parameter vs. statistic
Statistics: Descriptive vs. inferential

Types of variables

Quantitative vs. Qualitative
/ |
Discrete Continuous

Tables, charts & graphs

- frequency tables
- qualitative: bar graph/pie chart
- stem-and-leaf plot/dot plot
- time plot
- histogram (modality)
  - traits: # of modes, tail weight, overall shape (symmetry, skewness)
  - identify skewness by TAIL
- boxplot (skewness)
  - outliers, overall shape (symmetry, skewness)
  - identify skewness inside box or entire graph

Measures of center/spread/position

- center: mean, median, mode
  → Outlier effect? Skewness effect?
- spread: range, variance, standard deviation, IQR
  → Why use squared and ($n - 1$)? Ever negative? Empirical Rule?
- position: min, max, percentiles (quartiles)
  → recall that we INCLUDE the median when determining quartiles
  → 5-number summary, boxplot, types of outliers

Chapters 7-10:

Displaying bivariate data

- scatterplot: visual aid to see form/strength/direction of relationship and/or outliers (large residual, high leverage, influential)
- correlation: numerical aid to see strength/direction of relationship (range?)
  → Warning: assumes linearity, sensitive to outliers

Simple linear regression analysis

- regression line: $\hat{y} = b_0 + b_1 x$
- least-squares estimation gives $b_1 = r\left(\frac{s_y}{s_x}\right)$ and $b_0 = \bar{y} - b_1\bar{x}$
- estimation: interpolation vs. extrapolation (BAD!)
- R-squared: r^2 = proportion of variation in y explained by x
- causation: association does NOT imply causation
- residual plots: observed vs. theoretical appearance
- transformation of a variable can help improve linearity

Chapter 11-13:

- observational/retrospective/prospective study, experiment/controlled clinical trial
  → population and causal inferences (what needs to be present for each?)
- types of bias (response, undercoverage, nonresponse)
- types of sampling: with/without replacement, SRS/stratified/cluster/ voluntary/convenience/systematic
- controlling factors: randomization, blocking, direct control, replication
- more experiment design definitions

Chapters 14-15:

- types of events: marginal, conditional, union, intersection, complement,
  - What common words identify them?
- relating events: dependent vs. disjoint vs. independent
  - Do these relations affect the rules below? If so, how?
  - Do they allow certain rules to be easily extended?
- probability laws:
  - conditional probability: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$
  - complement rule: $P(A^C) = 1 - P(A)$
  - multiplication rule: $P(A \cap B) = P(A \text{ and } B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B)$
  - addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  - total probability rule: $P(A) = P\left(A \cap B\right) + P\left(A \cap B^C\right)$
  - recall examples where we combined a few of these together

Chapter 16-17:

Distributions

- discrete (exact probability or intervals) vs. continuous (only intervals)
  Discrete: If $P(X = a) > 0$, then $P(X \leq a) \neq P(X < a)$
  Continuous: If $P(X = a) = P(X = b) = 0$, then $P(a \leq X \leq b) = P(a < X < b)$
- discrete distributions:
  - determine probability distribution (values of X and corresponding probabilities)
  - mean: $\mu = \sum x_i P(X = x_i)$
  - variance: $\sigma^2 = \sum (x_i - \mu)^2 P(X = x_i)$
- continuous distributions:
  - uniform distribution: finding an area of a rectangle (with a twist!)
  - normal distribution: symmetric, 2 parameters: μ and σ, other properties

Standard Normal Distribution (and its applications)

- $\mu = 0$ and $\sigma = 1$
- Table Z only gives areas to left of value z, conversion to these values required
  → use diagrams, complements, symmetry, etc.
- standardizing: $P(X \le x)$ → $P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P(Z \le z)$
- identifying values for a given probability: $x = \mu + z\sigma$

Combinations and Functions of Random Variables

For any constants a and b,

Means:	*Variances*:
1. $E(a) = a$	1. $V(a) = 0$
2. $E(aX) = aE(X)$	2. $V(aX) = a^2V(X)$
3. $E(aX + b) = aE(X) + b$	3. $V(aX + b) = a^2V(X)$
4. $E(aX \pm bY) = aE(X) \pm bE(Y)$	4. $V(aX \pm bY) = a^2V(X) + b^2V(Y) \pm$ ~~$2ab\text{cov}(X, Y)$~~

$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b$, $E(Y) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n) + b$

If $X_1, X_2, \ldots, X_n$ are independent, $V(Y) = a_1^2V(X_1) + a_2^2V(X_2) + \ldots + a_n^2V(X_n)$

<u>Chapter 18:</u>

Sampling Distributions

- sample proportion:

Rule 1: $\mu_{\hat{p}} = p$. *Rule 2*: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$.

Rule 3: If np and $n(1-p)$ are both ≥ 15, then $\hat{p}$ has an approx. normal dist'n.

All 3 rules → $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \dot{\sim} N(0,1)$

- sample mean:

Rule 1: $\mu_{\bar{y}} = \mu$. *Rule 2*: $\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$.

Rule 3: When the population distribution is normal, the sampling distribution of $\bar{y}$ is also normal for any sample size n.

Rule 4 (CLT): When $n > 30$, the sampling distribution of $\bar{y}$ is well approximated by a normal curve, even when the population distribution is not itself normal.

All 4 rules → If n is large OR the population is normal, $Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \dot{\sim} N(0,1)$