Ch1 Statistics: The Art and Science of Learning from Data
Design, Description, Inference
Parameter – numerical summary of a population
Statistic - numerical summary of a sample

Ch2 Exploring Data with Graphs and Numerical Summaries
Categorical Variables:
  Summarize with Counts and Percentages
  Graphs – Bar Charts and Pie Charts
Quantitative Variables:
  Graphs – Dotplots, Histograms, Stemplots, Boxplots
  Measures of Center – Mean, Median, Mode
  Measures of Spread – Range, IQR, Variance, Standard Deviation
  Choosing best measures of center or spread for a particular shape distribution
  How outliers affect measures of center and spread.
  Empirical Rule (68% – 95% – 99.7%)
  Quartiles and Percentiles
  Five Number Summary
  Z-score

Ch3 Association: Contingency, Correlation and Regression
Contingency Tables: Conditional Proportions
Correlation (r):
  Measures strength and direction of linear association between 2 quantitative vars
  positive, negative
  strong, weak
  number between -1 and +1, no units
Regression:
  Equation to predict y from x
  x=explanatory (or predictor) variable
  y=response variable
  Regression Equation:
  \[ \hat{y} = a + bx \]
  \[ b = r \frac{s_y}{s_x} \] slope – average change in y for a one-unit change in x
  \[ a = \bar{y} - bx \] y-intercept – expected value of y when x=0, BUT we only interpret
  if x=0 makes sense and is close to the values of x observed in data
  Find the equation using the data summaries
  Use line for making predictions
  Residuals = observed y – predicted y (prediction errors)
  Least Squares Method: finds the line that minimizes the sum of squared residuals
  \[ R^2 = (r)^2 \] proportion of the variability in y that is explained by the regression on x
Cautions:
- Extrapolation
- Influential Outlier
- Correlation (or Association) does not imply Causation
- Simpson’s Paradox – a lurking variable can reverse the association between two categorical variables in a Contingency Table

Chapter 4: Gathering Data
Experiments vs Observational Studies
Simple Random Sample

Surveys:
- Margin of Error
- Sampling Bias: Undercoverage, Volunteer Samples, Convenience Samples
- Nonresponse Bias
- Response Bias

Experiments:
- Control: Placebos, Blind Study, Lurking Variables, Matched Pairs (Blocks)
- Randomization
- Replication

Observational Studies:
- Cross-sectional Studies
- Retrospective Studies
- Prospective Studies

Chapter 5: Probability in our Daily Lives
Randomness
Probability
Independent Trials
Sample Space
- Complement of an Event: \( P(A^c) = 1 - P(A) \)
- Disjoint Events A and B: \( P(A \text{ or } B) = P(A) + P(B) \)
- Conditional Probability: \( P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} \)
- Independent Events A and B:
  - Definition: \( P(A \mid B) = P(A) \)
  - Multiplication Rule: \( P(A \text{ and } B) = P(A) \times P(B) \)

P(at least one)
Problems of sensitivity and specificity
Chapter 6: Probability Distributions

Discrete Random Variable:
Finite number of possible values
Probability Distribution: list, graph or formula with all possible values of X and their probabilities

Population Mean $\mu = \sum xP(x)$

Continuous Random Variables:
Infinite number of possible values
Probabilities are areas under a density curve (smooth) with a total area of 1
Assign probabilities to intervals, not individual values of X

Normal Probability Distributions:
Bell-shaped curves, indexed by their mean: $\mu$ and standard deviation: $\sigma$
Follows Empirical Rule
$z$-score: $z = \frac{x - \mu}{\sigma}$
Empirical Rule
Using the Z table
area to the left, to the right, in between value of x for top 5%, bottom 20%, central 50%, etc

Binomial Distribution:
Each of n trials can have two possible outcomes: success or failure
Probability of success for each trial is the same: p (independent events)
Binomial Random Variable X counts the number of successes
Mean: $\mu = np$ and Standard Deviation $\sigma = \sqrt{np(1-p)}$

Binomial Formula:

$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0,1,2...n \quad \text{where } \binom{n}{x} = \frac{n!}{x!(n-x)!}$