Chapter 1

- Observational Study v.s. Experimental Study
- Blinding
- Random sampling (simple random sample, random cluster sampling, stratified random sampling)

Chapter 2

- Variable:
  - Categorical: nominal or ordinal
  - Numeric: continuous or discrete
- Frequency table and histogram
- Shapes of distributions:
  - Symmetric and bell-shaped
  - Skewed to the right
  - Skewed to the left
  - How to tell if a distribution is left skewed or right skewed?
    - Mean v.s. median
    - Normal quantile plot
  - Transformations to correct skewness
  - Skewness of Binomial distribution

- Descriptive statistics:
  - mean and median
  - standard deviation, IQR, and range
  - percentile and quantile
  - robustness
- Boxplots
- Effect of transformation of variables
  - How does mean, median, SD, IQR change under linear transformation $Y' = aY + b$?

Chapter 3

- Probability rules (1) to (8)
  - Disjoint v.s. independent

    - Conditional probability
- Random variables:
    - Discrete:
        - pmf, summation equals 1
        - Mean and variance
        - Binomial distribution
    - Continuous:
        - pdf, integral equals 1
        - Areas under the curve
        - Normal distribution
    - Discrete v.s. continuous: the probability of single point

- Rules for means of random variables (1) to (4)
- Binomial distribution $B(n, p)$:
    - Binary outcomes, independent trials, n is fixed, same value of p
    - pmf
    - Properties of binomial coefficients
    - The use of complement to calculate probability
    - Mean and variance

Chapter 4

- Normal distribution $N(\mu, \sigma^2)$:
    - Note that the second argument denotes the variance, rather than the SD
    - pdf
    - Areas under a normal curve: standardization and Z table
    - Inverse reading of Z table
    - Empirical rule for normal distribution
    - Assessing normality: normal quantile plot (only R implementation required)

Book Problems

For a particular disease, the probability of the disease is 0.02. If someone has the disease, the probability they test positive is 0.9. If they do not have the disease, the probability they test negative is 0.99.

- What is the probability that if someone tests negative, they do not have the disease?

$$
\begin{aligned}
P(D^C|-) &= \frac{P(- \cap D^C)}{P(-)} \\
&= \frac{P(-|D^C)P(D^C)}{P(-|D^C)P(D^C) + P(-|D)P(D)} \\
&= \frac{0.99 \times (1 - 0.02)}{0.99 \times (1 - 0.02) + (1 - 0.9) \times 0.04} \\
&= 0.9959.
\end{aligned}
$$

Ben's favorite cereal is running a promotion that says 1-in-2 boxes of the cereal contain a prize. Suppose that Ben is going to buy 5 boxes of this cereal, and let X represents the number of prizes he wins in these boxes. Assume that these boxes represent a random sample, and assume that prizes are independent between boxes.

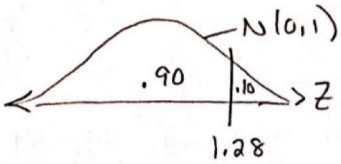
- What is the probability that he wins at most 4 prize in the boxes?

$$
\begin{aligned}
P(X \leq 4) &= 1 - P(X > 4) \\
&= 1 - P(X = 5) \\
&= 1 - \binom{5}{5} 0.5^5 (1 - 0.5)^{5-5} \\
&= 1 - 0.5^5 \\
&= 0.96875.
\end{aligned}
$$

Body fat percentage for males has a population mean of 13%, with a standard deviation of 5%. Assume the population is normally distributed.

- What is the probability that a single male has a body fat percentage above 15%?
- Find the 90% percentile for a single males body fat percentage.

$$
\begin{aligned}
Pr\{Y > 15\} &= Pr\left\{\frac{Y-\mu}{\sigma} > \frac{15-13}{5}\right\} = Pr\{Z > 0.4\} \\
&= 1 - Pr\{Z < 0.4\} \\
&= 1 - 0.6554 = 0.3446
\end{aligned}
$$



$z = 1.28$

$Y = \mu + z\sigma$

$= 13 + 1.28(5)$

$= 19.4$

The 90% percentile for a single males body fat percentage is 19.4%.