Homework 5

Han Nguyen - TXN200004

November 2, 20204

Problem 1

Light Bulb Defects - Geometric Distribution

Max owns a light bulb manufacturing company where 3 out of every 75 bulbs are defective.

```r
# Define the probability of defect
p <- 3/75  # probability of finding a defective bulb
q <- 1 - p  # probability of finding a non-defective bulb

# Display probabilities
cat("Probability of defect (p):", p, "\n")
```

```
## Probability of defect (p): 0.04
```

```r
cat("Probability of non-defect (q):", q, "\n")
```

```
## Probability of non-defect (q): 0.96
```

a)

Question: What is the probability that Max will find the first faulty light bulb on the 6th one that he tested?

Derivation:

This follows a geometric distribution. The probability of finding the first defective bulb on the k-th trial is:

$$P(X = k) = (1 - p)^{k-1} \cdot p$$

For $k = 6$:

$$P(X = 6) = (1 - \frac{3}{75})^{6-1} \cdot \frac{3}{75} = (\frac{72}{75})^5 \cdot \frac{3}{75} = (0.96)^5 \cdot 0.04$$

```r
# Probability of first defect on 6th trial
k <- 6
prob_1a <- (q^(k-1)) * p

cat("P(X = 6) =", round(prob_1a, 3), "\n")
```

```
## P(X = 6) = 0.033
```

Answer: The probability is **0.033**

b)

Question: What is the probability of taking at least four trials to find the first defective light bulb?

Derivation:

We need to find $P(X \geq 4)$, which equals $1 - P(X < 4) = 1 - P(X \leq 3)$.

Alternatively, using the complement rule:

$$P(X \geq 4) = P(\text{first 3 trials are non-defective}) = (1-p)^3 = (0.96)^3$$

```r
# Using complement of CDF
prob_1b <- 1 - pgeom(2, prob = p)  # pgeom(k-1) gives P(X <= k)
cat("P(X >= 4):", round(prob_1b, 3), "\n")
```

```
## P(X >= 4): 0.885
```

Answer: The probability is **0.885**

c)

Question: What is the probability of taking at most 10 trials to find the first defective light bulb?

Derivation:

We need to find $P(X \leq 10)$, which is the cumulative probability:

$$P(X \leq 10) = \sum_{k=1}^{10} P(X = k) = 1 - P(X > 10) = 1 - (1-p)^{10}$$

```r
# Using complement
prob_1c<- 1 - q^10
cat("P(X <= 10):", round(prob_1c, 3), "\n")
```

```
## P(X <= 10): 0.335
```

Answer: The probability is **0.335**

Problem 2

Yahtzee Simulation - Binomial Distribution

In this simplified Yahtzee game, we roll 5 fair six-sided dice and count the number of ones. We repeat this process 10,000 times.

```r
# Load required library
library(ggplot2)

# Set seed for reproducibility
set.seed(20220707)

# Simulate rolling 5 dice 10,000 times
n_simulations <- 10000
n_dice <- 5

# For each simulation, roll 5 dice and count the number of ones
X <- replicate(n_simulations, {
  dice_rolls <- sample(1:6, size = n_dice, replace = TRUE)
  sum(dice_rolls == 1)  # Count how many ones
})

# Display first few values
cat("First 20 values of X:", head(X, 20), "\n")
```

```
## First 20 values of X: 2 0 3 2 2 1 3 1 1 0 0 1 1 0 0 0 1 0 0 1
```

```r
cat("Summary statistics:\n")
```

```
## Summary statistics:
```

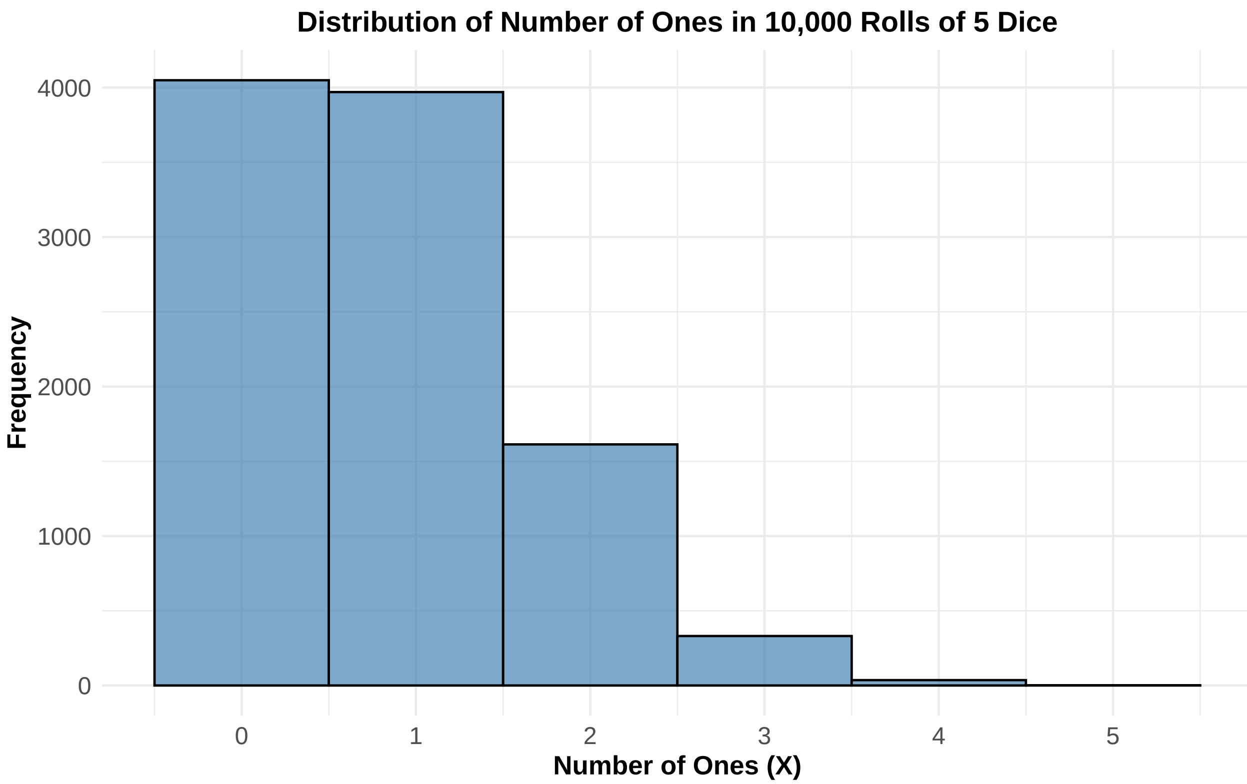
```r
summary(X)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  1.0000  0.8338  1.0000  5.0000
```

Histogram of X

```r
# Create a data frame for ggplot
data_X <- data.frame(X = X)

# Create histogram using ggplot
ggplot(data_X, aes(x = X)) +
  geom_histogram(binwidth = 1, fill = "steelblue", color = "black", alpha = 0.7) +
  scale_x_continuous(breaks = 0:5) +
  labs(
    title = "Distribution of Number of Ones in 10,000 Rolls of 5 Dice",
    x = "Number of Ones (X)",
    y = "Frequency"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.text = element_text(size = 11),
    axis.title = element_text(size = 12, face = "bold")
  )
```



Sample Mean and Sample Variance

Formulas:

The **sample mean** is calculated as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_{10000}}{10000}$$

The **sample variance** is calculated as:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{\sum_{i=1}^{10000}(x_i - \bar{x})^2}{10000 - 1}$$

```r
# Calculate sample mean
sample_mean <- sum(X) / length(X)

# Calculate sample variance
sample_variance <- sum((X - sample_mean)^2) / (length(X) - 1)

# Display results
cat("Sample Mean (x-bar) =", round(sample_mean, 3), "\n\n")
```

```
## Sample Mean (x-bar) = 0.834
```

```r
cat("Sample Variance (s^2) =", round(sample_variance, 3), "\n\n")
```

```
## Sample Variance (s^2) = 0.705
```

Results:

- **Sample Mean:** $\bar{x} = 0.834$
- **Sample Variance:** $s^2 = 0.705$

Note: This follows a binomial distribution with $n = 5$ trials and $p = 1/6$ probability of success (rolling a one). The theoretical mean is $np = 5 \times \frac{1}{6} \approx 0.833$ and theoretical variance is $np(1-p) = 5 \times \frac{1}{6} \times \frac{5}{6} \approx 0.694$. We can see that the sample mean and sample variance are very close to the theoretical values.

Problem 3

Traffic Congestion - Poisson Distribution

On average, 180 cars per hour pass a specified point on a road during morning rush hour. Congestion occurs if more than 5 cars pass in any one minute.

```r
# Calculate lambda for one minute
# Average cars per hour = 180
# Average cars per minute (lambda)
lambda <- 180 / 60
cat("Average cars per minute (lambda):", lambda, "\n")
```

```
## Average cars per minute (lambda): 3
```

Probability of Congestion

Question: What is the probability that congestion will occur in any minute (i.e., more than 5 cars pass)?

Solution:

The number of cars passing in one minute follows a Poisson distribution with $\lambda = \frac{180}{60} = 3$.

We need to find $P(X > 5)$ where $X \sim \text{Poisson}(\lambda = 3)$.

Using the complement rule:

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{k=0}^{5} \frac{e^{-\lambda}\lambda^k}{k!}$$

```r
# Probability of congestion (more than 5 cars)
# P(X > 5) = 1 - P(X <= 5)
prob_congestion <- 1 - ppois(5, lambda = lambda)

cat("P(X > 5) = P(congestion) =", round(prob_congestion, 3), "\n")
```

```
## P(X > 5) = P(congestion) = 0.084
```

Answer: The probability of congestion occurring in any one minute is **0.084**

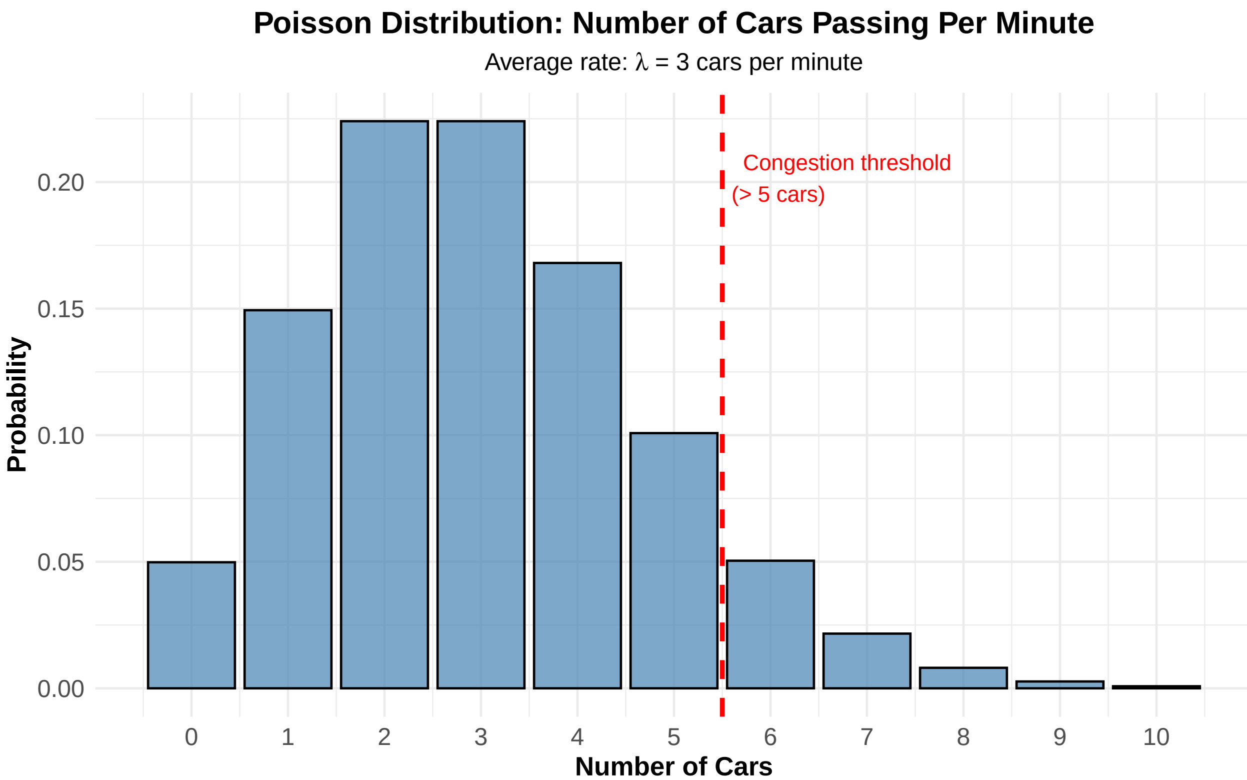
Bar Chart of Poisson Probabilities

Question: Create a bar chart showing the probability distribution for 0 to 10 cars passing in one minute.

```r
# Create data for the bar chart
cars <- 0:10
probabilities <- dpois(cars, lambda = lambda)

# Create data frame
poisson_data <- data.frame(
  Cars = cars,
  Probability = probabilities
)

# Create bar chart using ggplot
ggplot(poisson_data, aes(x = Cars, y = Probability)) +
  geom_bar(stat = "identity", fill = "steelblue", color = "black", alpha = 0.7) +
  geom_vline(xintercept = 5.5, linetype = "dashed", color = "red", linewidth = 1) +
  annotate("text", x = 5.5, y = max(probabilities) * 0.9,
           label = "Congestion threshold\n(> 5 cars)",
           color = "red", hjust = -0.1, size = 3.5) +
  scale_x_continuous(breaks = 0:10) +
  labs(
    title = "Poisson Distribution: Number of Cars Passing Per Minute",
    subtitle = expression(paste("Average rate: ", lambda, " = 3 cars per minute")),
    x = "Number of Cars",
    y = "Probability"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    plot.subtitle = element_text(hjust = 0.5, size = 11),
    axis.text = element_text(size = 11),
    axis.title = element_text(size = 12, face = "bold")
  )
```




```r
# Display the probability table
cat("\nProbability Distribution Table:\n")
```

```
##
## Probability Distribution Table:
```

```r
print(poisson_data, digits = 3)
```

```
##    Cars Probability
## 1     0     0.04979
## 2     1     0.14936
## 3     2     0.22404
## 4     3     0.22404
## 5     4     0.16803
## 6     5     0.10082
## 7     6     0.05041
## 8     7     0.02160
## 9     8     0.00810
## 10    9     0.00270
## 11   10     0.00081
```

Interpretation:

- The red dashed line shows the congestion threshold (5 cars)
- Cars to the right of this line (6, 7, 8, 9, 10, . . .) represent congestion scenarios
- The distribution is centered around $\lambda = 3$ cars per minute
- The probability of exactly 3 cars is highest at 0.224

Problem 4

University Entrance Test - Normal Distribution

Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100.

```
# Define the parameters of the normal distribution
mu <- 500      # mean
sigma <- 100   # standard deviation

cat("Distribution: X ~ N(mu =", mu, ", sigma =", sigma, ")\n")
```

```
## Distribution: X ~ N(mu = 500 , sigma = 100 )
```

a) Probability of Scoring 585 or Less

Question: What is the probability that someone will score 585 or less on this national test?

Solution:

We need to find $P(X \leq 585)$ where $X \sim N(500, 100)$.

This can be calculated using the cumulative distribution function (CDF):

$$P(X \leq 585) = \Phi\left(\frac{585 - 500}{100}\right) = \Phi(0.85)$$

where Φ is the standard normal CDF.

```
# Score threshold
score <- 585

# Calculate probability using pnorm
prob_585_or_less <- pnorm(score, mean = mu, sd = sigma)
cat("P(X <= 585) =", round(prob_585_or_less, 3), "\n")
```

```
## P(X <= 585) = 0.802
```

Answer: The probability of scoring 585 or less is **0.802**

Interpretation: This means approximately 80.2% of test-takers score 585 or below.

b) Quartiles of the Distribution

Question: Find the lower quartile (Q1), median (Q2), and upper quartile (Q3) of the normal distribution.

Solution:

For a normal distribution: - **Lower Quartile (Q1):** 25th percentile, $P(X \leq Q_1) = 0.25$ - **Median (Q2):** 50th percentile, $P(X \leq Q_2) = 0.50$ - **Upper Quartile (Q3):** 75th percentile, $P(X \leq Q_3) = 0.75$

We use the quantile function (inverse CDF):

$$Q_p = \mu + \sigma \cdot \Phi^{-1}(p)$$

```r
# Calculate quartiles using qnorm
Q1 <- qnorm(0.25, mean = mu, sd = sigma)  # Lower quartile (25th percentile)
Q2 <- qnorm(0.50, mean = mu, sd = sigma)  # Median (50th percentile)
Q3 <- qnorm(0.75, mean = mu, sd = sigma)  # Upper quartile (75th percentile)

# Display results
cat("Lower Quartile (Q1, 25th percentile):", round(Q1, 3), "\n")
```

```
## Lower Quartile (Q1, 25th percentile): 432.551
```

```r
cat("Median (Q2, 50th percentile):", round(Q2, 3), "\n")
```

```
## Median (Q2, 50th percentile): 500
```

```r
cat("Upper Quartile (Q3, 75th percentile):", round(Q3, 3), "\n")
```

```
## Upper Quartile (Q3, 75th percentile): 567.449
```

Answers:

- **Lower Quartile (Q1):** 432.551
- **Median (Q2):** 500
- **Upper Quartile (Q3):** 567.449

Interpretation:

- 25% of students score below 432.551
- 50% of students score below 500 (the median)
- 75% of students score below 567.449

Problem 5

Coin Flips - Binomial Distribution

Suppose we flip a fair coin 10 times.

```r
# Define parameters
n <- 10      # number of coin flips
p <- 0.5     # probability of heads (fair coin)

cat("Distribution: X ~ Binomial(n =", n, ", p =", p, ")\n")
```

```
## Distribution: X ~ Binomial(n = 10 , p = 0.5 )
```

Event A: Seven or More Heads

Question: What is the probability of observing seven or more heads?

Derivation:

We need to find $P(X \geq 7)$ where $X \sim \text{Binomial}(n = 10, p = 0.5)$.

The binomial probability formula is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Therefore:

$$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

$$P(X \geq 7) = \sum_{k=7}^{10} \binom{10}{k} (0.5)^k (0.5)^{10-k} = \sum_{k=7}^{10} \binom{10}{k} (0.5)^{10}$$

Breaking it down: - $P(X = 7) = \binom{10}{7}(0.5)^{10} = 120 \times (0.5)^{10}$ - $P(X = 8) = \binom{10}{8}(0.5)^{10} = 45 \times (0.5)^{10}$ - $P(X = 9) = \binom{10}{9}(0.5)^{10} = 10 \times (0.5)^{10}$ - $P(X = 10) = \binom{10}{10}(0.5)^{10} = 1 \times (0.5)^{10}$

$$P(X \geq 7) = \frac{120 + 45 + 10 + 1}{1024} = \frac{176}{1024}$$

```r
# Event A: P(X >= 7)
# Direct calculation: P(X=7) + P(X=8) + P(X=9) + P(X=10)
prob_A <- sum(dbinom(7:10, size = n, prob = p))

cat("Event A: P(X >= 7) =", round(prob_A, 4), "\n")
```

```
## Event A: P(X >= 7) = 0.1719
```

```r
cat("Rounded to 2 decimal places:", round(prob_A, 2), "\n")
```

```
## Rounded to 2 decimal places: 0.17
```

Answer: The probability of observing seven or more heads is **0.17**

Event B: Three or Less Heads

Question: What is the probability of observing three or less heads?

Derivation:

We need to find $P(X \leq 3)$ where $X \sim \text{Binomial}(n = 10, p = 0.5)$.

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$P(X \leq 3) = \sum_{k=0}^{3} \binom{10}{k} (0.5)^{10}$$

Breaking it down: - $P(X=0) = \binom{10}{0}(0.5)^{10} = 1 \times (0.5)^{10}$ - $P(X=1) = \binom{10}{1}(0.5)^{10} = 10 \times (0.5)^{10}$ - $P(X=2) = \binom{10}{2}(0.5)^{10} = 45 \times (0.5)^{10}$ - $P(X=3) = \binom{10}{3}(0.5)^{10} = 120 \times (0.5)^{10}$

$$P(X \leq 3) = \frac{1+10+45+120}{1024} = \frac{176}{1024}$$

```
# Event B: P(X <= 3)
# Direct calculation: P(X=0) + P(X=1) + P(X=2) + P(X=3)
prob_B <- sum(dbinom(0:3, size = n, prob = p))

cat("Event B: P(X <= 3) =", round(prob_B, 4), "\n")
```

```
## Event B: P(X <= 3) = 0.1719
```

```
cat("Rounded to 2 decimal places:", round(prob_B, 2), "\n")
```

```
## Rounded to 2 decimal places: 0.17
```

Answer: The probability of observing three or less heads is **0.17**

Comparison

Both Event A and Event B have the **same probability** of **0.17**, so they have equal chance to happen. This makes sense due to the **symmetry of the binomial distribution** when $p = 0.5$. The probability of getting 7 or more heads equals the probability of getting 3 or fewer heads (which is equivalent to getting 7 or more tails).

Problem 6

Hit-and-Run Taxi-Cab - Bayes' Theorem

An eyewitness observes a hit-and-run taxi-cab accident in New Territories, Hong Kong where 95% of the cabs are green and the rest are red. The witness is 80% sure about the cab color.

Given Information

```
# Prior probabilities (base rates)
P_Green <- 0.95   # P(cab is Green)
P_Red <- 0.05     # P(cab is Red)

# Conditional probabilities (witness accuracy)
P_SaysGreen_given_Green <- 0.80  # P(witness says Green | cab is Green)
P_SaysRed_given_Red <- 0.80      # P(witness says Red | cab is Red)

# Complement probabilities (witness errors)
P_SaysRed_given_Green <- 0.20    # P(witness says Red | cab is Green)
P_SaysGreen_given_Red <- 0.20    # P(witness says Green | cab is Red)
```

Derivation Using Bayes' Theorem

Question: What is the probability that the cab actually was red given that the witness said it was red?

We need to find: $P(\text{Red} \mid \text{says Red})$

Bayes' Theorem:

$$P(\text{Red} \mid \text{says Red}) = \frac{P(\text{says Red} \mid \text{Red}) \times P(\text{Red})}{P(\text{says Red})}$$

Step 1: Calculate the numerator

$$P(\text{says Red} \mid \text{Red}) \times P(\text{Red}) = 0.80 \times 0.05 = 0.04$$

Step 2: Calculate the denominator using the Law of Total Probability

The probability that the witness says "Red" can come from two scenarios: 1. The cab is actually red AND the witness correctly identifies it as red 2. The cab is actually green AND the witness incorrectly identifies it as red

$$P(\text{says Red}) = P(\text{says Red} \mid \text{Red}) \times P(\text{Red}) + P(\text{says Red} \mid \text{Green}) \times P(\text{Green})$$

$$P(\text{says Red}) = (0.80 \times 0.05) + (0.20 \times 0.95)$$

$$P(\text{says Red}) = 0.04 + 0.19 = 0.23$$

Step 3: Apply Bayes' Theorem

$$P(\text{Red} \mid \text{says Red}) = \frac{0.04}{0.23} = \frac{4}{23} \approx 0.1739$$

```
# Step 1: Calculate numerator
numerator <- P_SaysRed_given_Red * P_Red

# Step 2: Calculate denominator using Law of Total Probability
P_SaysRed <- (P_SaysRed_given_Red * P_Red) + (P_SaysRed_given_Green * P_Green)

# Step 3: Apply Bayes' Theorem
P_Red_given_SaysRed <- numerator / P_SaysRed
cat("Bayes' Theorem:\n")
```

```
## Bayes' Theorem:
```

```
cat("P(Red | says Red) = Numerator / Denominator\n")
```

```
## P(Red | says Red) = Numerator / Denominator
```

```
cat("                  =", numerator, "/", P_SaysRed, "\n")
```

```
##                    = 0.04 / 0.23
```

```r
cat("                    =", round(P_Red_given_SaysRed, 4), "\n\n")
```

```
##                      = 0.1739
```

```r
cat("Rounded to 2 decimal places:", round(P_Red_given_SaysRed, 2), "\n")
```

```
## Rounded to 2 decimal places: 0.17
```

Answer: The probability that the cab actually was red given that the witness said it was red is **0.17** or approximately **17.39%**.

Interpretation

This result is counterintuitive. Even though the witness is 80% accurate, the probability that the cab was actually red is only about 17%. This demonstrates the **base rate fallacy** - we cannot ignore the prior probability (base rate) when making inferences. The low base rate of red cabs (5%) overwhelms the witness's 80% accuracy.

Problem 7

Monopoly Dice Simulation

Each player in Monopoly rolls a pair of dice and moves the same number of spaces. We simulate rolling a pair of dice 100 times and let X be the total value shown on the two dice.

Simulation

```r
# Load required library
library(ggplot2)

# Set seed for reproducibility
set.seed(20191031)

# Simulate rolling two dice 100 times
n_rolls <- 100

# Roll first die 100 times
die1 <- sample(1:6, size = n_rolls, replace = TRUE)

# Roll second die 100 times
die2 <- sample(1:6, size = n_rolls, replace = TRUE)

# Calculate total value X for each roll
X <- die1 + die2

# Display first 20 values
cat("First 20 values of X (sum of two dice):\n")
```

```
## First 20 values of X (sum of two dice):
```

```r
cat(head(X, 20), "\n\n")
```

```
## 8 4 8 6 9 2 6 8 6 7 9 2 10 8 12 6 2 6 6 7
```

```r
cat("Summary statistics:\n")
```

```
## Summary statistics:
```

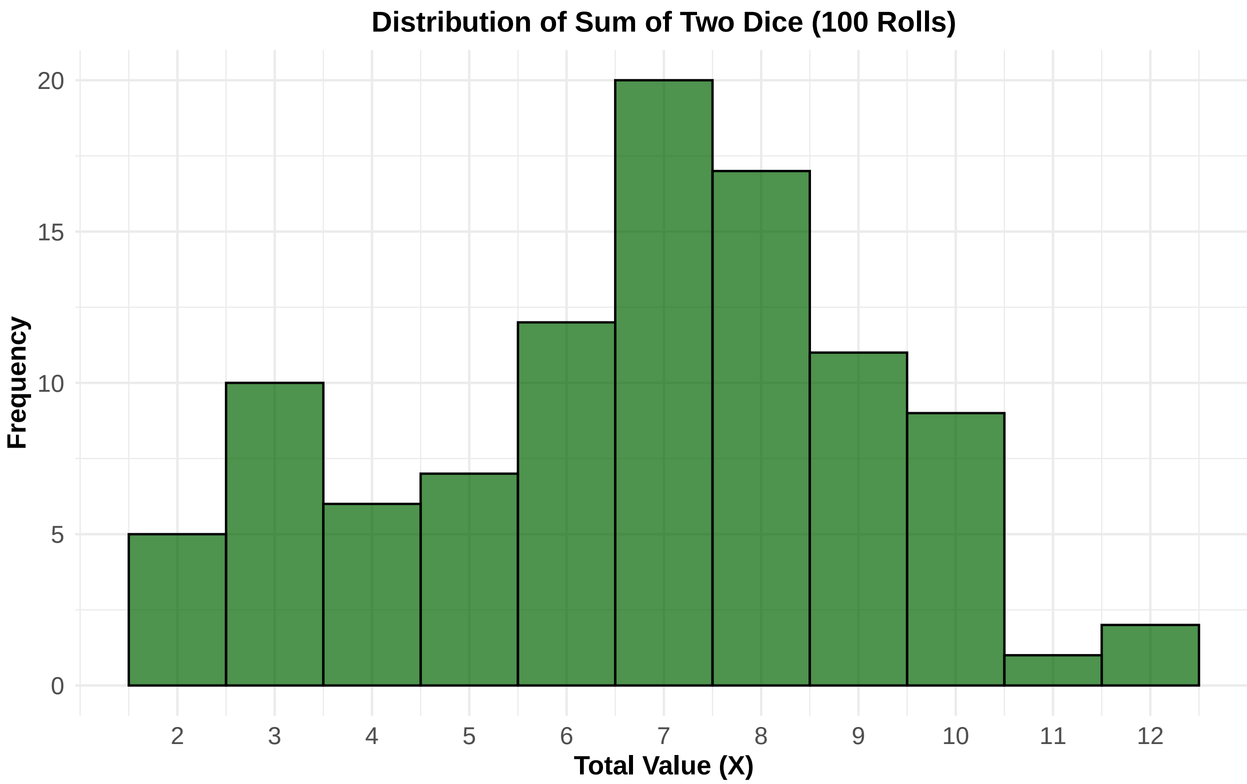
```r
summary(X)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    5.00    7.00    6.71    8.00   12.00
```

Histogram of X

```r
# Create a data frame for ggplot
data_X <- data.frame(X = X)

# Create histogram using ggplot
ggplot(data_X, aes(x = X)) +
  geom_histogram(binwidth = 1, fill = "darkgreen", color = "black", alpha = 0.7) +
  scale_x_continuous(breaks = 2:12) +
  labs(
    title = "Distribution of Sum of Two Dice (100 Rolls)",
    x = "Total Value (X)",
    y = "Frequency"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.text = element_text(size = 11),
    axis.title = element_text(size = 12, face = "bold")
  )
```



Sample Mean and Sample Variance

Formulas:

The **sample mean** is calculated as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_{100}}{100}$$

The **sample variance** is calculated as:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{\sum_{i=1}^{100}(x_i - \bar{x})^2}{100 - 1}$$

```r
# Calculate sample mean
sample_mean <- sum(X) / length(X)

# Calculate sample variance
sample_variance <- sum((X - sample_mean)^2) / (length(X) - 1)

# Display results
cat("Sample Mean:\n")
```

```
## Sample Mean:
```

```r
cat("x-bar = (1/n) * sum(x_i)\n")
```

```
## x-bar = (1/n) * sum(x_i)
```

```r
cat("      = (1/", length(X), ") * ", sum(X), "\n")
```

```
##       = (1/ 100 ) *  671
```

```r
cat("      =", round(sample_mean, 4), "\n\n")
```

```
##       = 6.71
```

```r
cat("Sample Variance:\n")
```

```
## Sample Variance:
```

```r
cat("s^2 = [1/(n-1)] * sum((x_i - x-bar)^2)\n")
```

```
## s^2 = [1/(n-1)] * sum((x_i - x-bar)^2)
```

```r
cat("    = [1/", length(X) - 1, "] * ", round(sum((X - sample_mean)^2), 4), "\n")
```

```
##     = [1/ 99 ] *  578.59
```

```r
cat("    =", round(sample_variance, 4), "\n\n")
```

```
##     = 5.8443
```

```r
cat("Rounded to 2 decimal places:\n")
```

```
## Rounded to 2 decimal places:
```

```r
cat("Sample Mean (x-bar) =", round(sample_mean, 2), "\n")
```

```
## Sample Mean (x-bar) = 6.71
```

```r
cat("Sample Variance (s^2) =", round(sample_variance, 2), "\n")
```

```
## Sample Variance (s^2) = 5.84
```

Results:

- **Sample Mean:** $\bar{x} = 6.71$
- **Sample Variance:** $s^2 = 5.84$

Note: Theoretically, when rolling two fair dice, the expected value is $E[X] = E[D_1] + E[D_2] = 3.5 + 3.5 = 7$, and the variance is $\text{Var}(X) = \text{Var}(D_1) + \text{Var}(D_2) = \frac{35}{12} + \frac{35}{12} = \frac{35}{6} \approx 5.83$.

Problem 8

Left-Handed Students - Binomial Distribution

Currently, there are 54 enrolled students in STAT 3355. It is known that 13.1% of the population in U.S. are left-handed. Assume the 54 students are independent samples with equal probability of being left-handed.

```r
# Define parameters
n <- 54          # number of students
p <- 0.131       # probability of being left-handed

cat("Distribution: X ~ Binomial(n =", n, ", p =", p, ")\n")
```

```
## Distribution: X ~ Binomial(n = 54 , p = 0.131 )
```

```r
cat("Expected number of left-handed students: E[X] = n*p =", n*p, "\n")
```

```
## Expected number of left-handed students: E[X] = n*p = 7.074
```

Probability of 10 or Fewer Left-Handed Students

Question: What is the probability that 10 or fewer left-handed students in this class?

Derivation:

Let X be the number of left-handed students. Then $X \sim \text{Binomial}(n = 54, p = 0.131)$.

We need to find $P(X \leq 10)$:

$$P(X \leq 10) = \sum_{k=0}^{10} P(X = k) = \sum_{k=0}^{10} \binom{54}{k} (0.131)^k (1 - 0.131)^{54-k}$$

We can use the cumulative distribution function (CDF) `pbinom()` to calculate this:

```r
# Calculate P(X <= 10)
prob_10_or_fewer <- pbinom(10, size = n, prob = p)

cat("P(X <= 10) =", round(prob_10_or_fewer, 4), "\n")
```

```
## P(X <= 10) = 0.9113
```

```r
cat("Rounded to 2 decimal places:", round(prob_10_or_fewer, 2), "\n")
```

```
## Rounded to 2 decimal places: 0.91
```

Answer: The probability of observing 10 or fewer left-handed students in this class is **0.91** or approximately **91.13%**.

Bar Chart of Probabilities

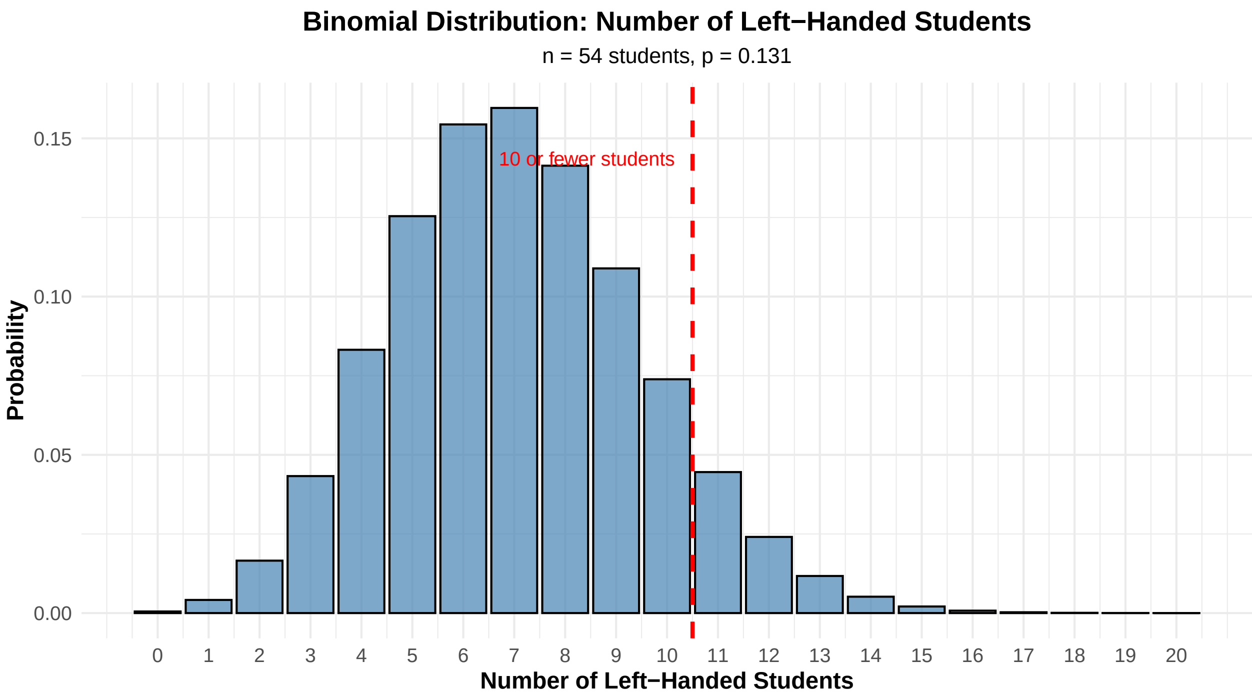
Question: Create a bar chart showing the probability distribution for 0 to 20 left-handed students.

```r
# Load required library
library(ggplot2)

# Create data for the bar chart
students <- 0:20
probabilities <- dbinom(students, size = n, prob = p)

# Create data frame
lefthanded_data <- data.frame(
  Students = students,
  Probability = probabilities
)

# Create bar chart using ggplot
ggplot(lefthanded_data, aes(x = Students, y = Probability)) +
  geom_bar(stat = "identity", fill = "steelblue", color = "black", alpha = 0.7) +
  geom_vline(xintercept = 10.5, linetype = "dashed", color = "red", linewidth = 1) +
  annotate("text", x = 10.5, y = max(probabilities) * 0.9,
           label = "10 or fewer students",
           color = "red", hjust = 1.1, size = 3.5) +
  scale_x_continuous(breaks = 0:20) +
  labs(
    title = "Binomial Distribution: Number of Left-Handed Students",
    subtitle = expression(paste("n = 54 students, p = 0.131")),
    x = "Number of Left-Handed Students",
    y = "Probability"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    plot.subtitle = element_text(hjust = 0.5, size = 11),
    axis.text = element_text(size = 10),
    axis.title = element_text(size = 12, face = "bold")
  )
```




```r
# Display the probability table
cat("\nProbability Distribution Table:\n")
```

```
##
## Probability Distribution Table:
```

```r
print(lefthanded_data, digits = 4)
```

```
##    Students Probability
## 1         0   5.094e-04
## 2         1   4.147e-03
## 3         2   1.657e-02
## 4         3   4.329e-02
## 5         4   8.320e-02
## 6         5   1.254e-01
## 7         6   1.544e-01
## 8         7   1.596e-01
## 9         8   1.414e-01
## 10        9   1.089e-01
## 11       10   7.388e-02
## 12       11   4.455e-02
## 13       12   2.407e-02
## 14       13   1.172e-02
## 15       14   5.174e-03
## 16       15   2.080e-03
## 17       16   7.643e-04
## 18       17   2.575e-04
## 19       18   7.981e-05
## 20       19   2.279e-05
## 21       20   6.013e-06
```

```r
# Calculate and display cumulative probability up to 10
cat("\nCumulative probability P(X <= 10) =", round(sum(probabilities[1:11]), 4), "\n")
```

```
##
## Cumulative probability P(X <= 10) = 0.9113
```

Problem 9

Cereal Box Heights - Normal Distribution

Cereal is sold by weight not volume. This introduces variability in the volume due to settling. As such, the height to which a cereal box is filled is random. Suppose the heights for a certain type of cereal box have a normal distribution with mean 12 and variance 0.52 in units of inches.

```r
# Define the parameters of the normal distribution
mu <- 12           # mean
variance <- 0.52   # variance
sigma <- sqrt(variance)   # standard deviation

cat("Distribution: X ~ N(mu =", mu, ", sigma^2 =", variance, ")\n")
```

```
## Distribution: X ~ N(mu = 12 , sigma^2 = 0.52 )
```

```r
cat("Standard deviation: sigma =", round(sigma, 4), "\n")
```

```
## Standard deviation: sigma = 0.7211
```

Probability of Height 10.7 Inches or Less

Question: What is the probability that a randomly chosen cereal box has heights of 10.7 inches or less?

Derivation:

We need to find $P(X \leq 10.7)$ where $X \sim N(\mu = 12, \sigma^2 = 0.52)$.

This can be calculated using the cumulative distribution function (CDF):

$$P(X \leq 10.7) = \Phi\left(\frac{10.7 - \mu}{\sigma}\right) = \Phi\left(\frac{10.7 - 12}{\sqrt{0.52}}\right) = \Phi\left(\frac{-1.3}{\sqrt{0.52}}\right)$$

where Φ is the standard normal CDF.

Calculating the z-score:

$$z = \frac{10.7 - 12}{\sqrt{0.52}} = \frac{-1.3}{0.7211} \approx -1.8028$$

```r
# Height threshold
height <- 10.7

# Calculate z-score
z_score <- (height - mu) / sigma
cat("Z-score: z = (x - mu) / sigma\n")
```

```
## Z-score: z = (x - mu) / sigma
```

```r
cat("        z = (", height, "-", mu, ") /", round(sigma, 4), "\n")
```

```
##         z = ( 10.7 - 12 ) / 0.7211
```

```r
cat("        z =", round(z_score, 4), "\n\n")
```

```
##         z = -1.8028
```

```r
# Calculate probability using pnorm
prob_10_7_or_less <- pnorm(height, mean = mu, sd = sigma)

cat("P(X <= 10.7) =", round(prob_10_7_or_less, 4), "\n")
```

```
## P(X <= 10.7) = 0.0357
```

```r
cat("Rounded to 2 decimal places:", round(prob_10_7_or_less, 2), "\n")
```

```
## Rounded to 2 decimal places: 0.04
```

Answer: The probability that a randomly chosen cereal box has heights of 10.7 inches or less is **0.04** or approximately **3.57%**.

Quartiles of the Distribution

Question: Find the lower quartile (Q1), median (Q2), and upper quartile (Q3) of the normal distribution.

Derivation:

For a normal distribution: - **Lower Quartile (Q1):** 25th percentile, $P(X \leq Q_1) = 0.25$ - **Median (Q2):** 50th percentile, $P(X \leq Q_2) = 0.50$ - **Upper Quartile (Q3):** 75th percentile, $P(X \leq Q_3) = 0.75$

We use the quantile function (inverse CDF) to find these values:

$$Q_p = \mu + \sigma \cdot \Phi^{-1}(p)$$

where Φ^{-1} is the inverse of the standard normal CDF.

```r
# Calculate quartiles using qnorm
Q1 <- qnorm(0.25, mean = mu, sd = sigma)  # Lower quartile (25th percentile)
Q2 <- qnorm(0.50, mean = mu, sd = sigma)  # Median (50th percentile)
Q3 <- qnorm(0.75, mean = mu, sd = sigma)  # Upper quartile (75th percentile)

Q1
```

```
## [1] 11.51362
```

```
Q2
```

```
## [1] 12
```

```
Q3
```

```
## [1] 12.48638
```

Answers:

- **Lower Quartile (Q1):** 11.51 inches
- **Median (Q2):** 12 inches
- **Upper Quartile (Q3):** 12.49 inches