STAT7203: Assignment 1

Alexander White

43218307

1. Comparative Study

(a) **Experimental Units:**
Tomato Plants

(b) **Factors:**
Irrigation frequency and fertilizer type.

(c) **Levels of Each Factor:**
Irrigation frequency: 2
Fertilizer type: 3

(d) **Treatments:**
Daily and Organic
Daily and Synthetic
Daily and None
Every other day and Organic
Every other day and Synthetic
Every other day and None

(e) **Type of Response Variable:**
The response variable is quantitative

2. Visualization

(a) **Sample Statistics:**
R Code:

```
#STAT7203 Assignment 1 Coding

#setwd('Desktop/STAT7203')
# Read in file and check it
invest <- read.csv('investment.csv')
head(invest)

# Calculate mean for each type
avg_balanced <- mean(invest$Balanced, na.rm = TRUE)
avg_growth <- mean(invest$Growth, na.rm = TRUE)

# Calculate median for each type
med_balanced <- median(invest$Balanced, na.rm = TRUE)
med_growth <- median(invest$Growth, na.rm = TRUE)

# Calculate standard deviation for each type
std_balanced <- sd(invest$Balanced, na.rm = TRUE)
std_growth <- sd(invest$Growth, na.rm = TRUE)

# Printing the results
cat("Balanced Fund - Mean:",
    avg_balanced,
    " Median:",
    med_balanced,
```

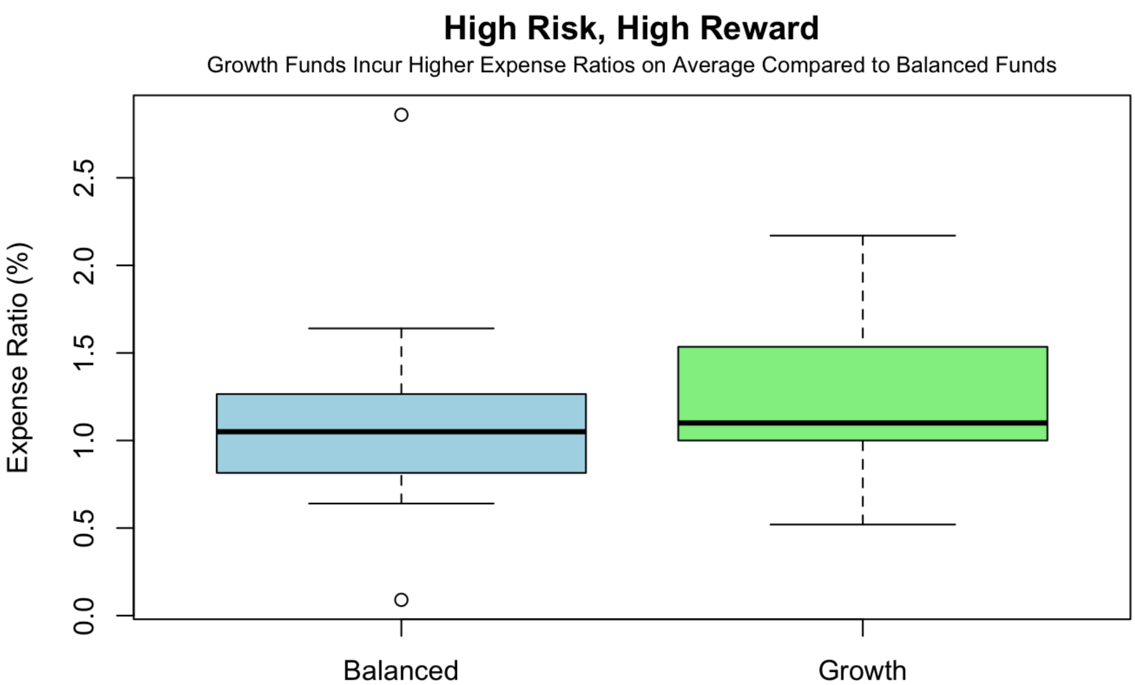
```
    " Std Dev:",
    std_balanced,
    "\n")
cat("Growth Fund - Mean:",
    avg_growth,
    " Median:",
    med_growth,
    " Std Dev:",
    std_growth,
    "\n")

# Comparative boxplot
boxplot(invest$Balanced, invest$Growth,
        names = c("Balanced", "Growth"),
        main = "High Risk, High Reward",
        ylab = "Expense Ratio (%)",
        col = c("lightblue", "lightgreen"))

# Adding a subtitle for further context
mtext("Growth Funds Perform Better on Average", side = 3, line = 0.5, cex =
```

Output:
Balanced Fund - Mean: 1.1205 Median: 1.05 Std Dev: 0.5360428
Growth Fund - Mean: 1.2435 Median: 1.1 Std Dev: 0.4484626

(b) **Comparative Boxplot:**



Interesting Features:

(i) The Balanced funds group has two significant outliers. The Growth funds group has no significant outliers.

(ii) The Growth funds group has a slightly larger IQR than the Balanced funds, which indicates more variability in the middle 50 percent of the data.

(iii) The Balanced funds have a much larger range compared to the Growth funds, primarily due to the outliers, which leads to greater overall variability.

(iv) However, without the 2 outliers, the Balanced funds group has a smaller range, which indicates that Balanced funds may be more consistent in their expense ratios.

(v) The median expense ratio in both the Balanced and Growth funds groups are quite similar, being 1.05 and 1.1 respectively.

Overall, it appears that the growth funds are generally more volatile in their expense ratios. This could be due to the nature of growth funds investing in higher risk investments.

3. Counting

(a) **Probability that the first GG song heard is the fifth song played:**
The GG song must be at the 5th position. Therefore the first four songs must non-GG songs.
Arranging the first 4 songs can be done this many ways:

$$\text{Number of ways} = \binom{90}{4} \times 4!$$

The GG song will then be played 5th.
The next 95 songs can then be played in 95! ways.
The total number of favourable outcomes is:

$$\text{Favourable outcomes} = \binom{90}{4} \times 4! \times 95!$$

The total number of possible outcomes for all 100 songs:

$$\text{Total outcomes} = 100!$$

Therefore, the probability that the first GG song is the fifth song that is played:

$$P(\text{GG song is 5th}) = \frac{\binom{90}{4} \times 4! \times 95!}{100!}$$

Simplifying:

$$P(\text{GG song is 5th}) = \frac{90 \times 89 \times 88 \times 87}{100 \times 99 \times 98 \times 97 \times 96} \approx 0.00679$$

(b) **Probability that at least one of the first five songs played is by the GG:**
Can use the complement rule:

$$P(\text{At least one GG song}) = 1 - P(\text{No GG song in the first five})$$

Probability that none of the first 5 songs are GG songs:

$$P(\text{No GG song in the first 5}) = \frac{90}{100} \times \frac{89}{99} \times \frac{88}{98} \times \frac{87}{97} \times \frac{86}{96}$$

Therefore, the probability that at least one of the first five songs is a GG song is:

$$P(\text{At least one GG song}) = 1 - \left(\frac{90}{100} \times \frac{89}{99} \times \frac{88}{98} \times \frac{87}{97} \times \frac{86}{96}\right) \approx 0.416$$

4. Conditional Probability

Let voting for Tastycrats be represented by T.
Let voting for Non-Tastycrats be represented by T^c.
Let participation in the exit poll be represented by P.

Therefore:

- $P(T) = 0.52$
- $P(T^c) = 1 - P(T) = 0.48$
- $P(P \mid T) = 0.60$
- $P(P \mid T^c) = 0.80$

Need to find $P(T \mid P)$, the person voted for the Tastycrats given they participated in the exit poll.
Can use Bayes' Theorem to compute this:

$$P(T \mid P) = \frac{P(P \mid T) \times P(T)}{P(P \mid T) \times P(T) + P(P \mid T^c) \times P(T^c)}$$

Using given values:

$$= \frac{(0.60 \times 0.52)}{(0.60 \times 0.52) + (0.8 \times 0.48)}$$

$$\approx 0.4483$$

Therefore, the probability that a person who participated in the exit poll voted for the Tastycrats is approximately 0.4483 or 44.83%.

5. Discrete Random Variable

Let the probability mass function of Y be defined as:

$$f(y) = P(Y = y)$$

The possible values of Y are $1, 2, 3, 4$.

Calculating f(y) for each possible value:
$P(Y = 1)$:

$$P(Y = 1) = \frac{2}{5}$$

$P(Y = 2)$:

$$P(Y = 2) = \frac{3}{5} \times \frac{2}{4} = \frac{6}{20} = \frac{3}{10}$$

$P(Y = 3)$:

$$P(Y = 3) = \frac{3}{5} \times \frac{2}{4} \times \frac{2}{3} = \frac{12}{60} = \frac{1}{5}$$

$P(Y = 4)$:

$$P(Y = 4) = \frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} \times 1 = \frac{6}{60} = \frac{1}{10}$$

Therefore, the probability mass function of Y is given below.

y	1	2	3	4
$P(Y = y)$	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$

6. Discrete Random Variable

(a) **Probability Mass Function**

Let X be the random variable corresponding to the number of dice that show the same number.
The possible values of X are 0, 2, and 3. $P(X = 0)$:

$$P(X = 0) = \frac{5}{6} \times \frac{4}{6} = \frac{20}{36}$$

$P(X = 2)$:

$$P(X = 2) = \binom{3}{2} \times \frac{1}{6} \times \frac{5}{6} = \frac{15}{36}$$

$P(X = 3)$:

$$P(X = 3) = 6 \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{6}{216} = \frac{1}{36}$$

x	0	2	3
$P(X = x)$	$\frac{20}{36}$	$\frac{15}{36}$	$\frac{1}{36}$

(b) **Probability of Winning Any Money**

The probability that the player wins any money.
Player only wins money when $X = 2$ or $X = 3$:

$$P(\text{Win}) = P(X = 2) + P(X = 3) = \frac{15}{36} + \frac{1}{36} = \frac{16}{36} \approx 44.44$$

(c) **Expected Profit**

The player will be profitable if their payout exceeds the buy-in cost of \$2.
Therefore, Profit = Payout − \$2:

$$E[\text{Profit}] = \left(\frac{20}{36} \times -2\right) + \left(\frac{15}{36} \times -1\right) + \left(\frac{1}{36} \times 8\right)$$

$$E[\text{Profit}] = -\frac{40}{36} - \frac{15}{36} + \frac{8}{36} = -\frac{47}{36} \approx -\$1.305$$

Therefore, the expected profit is actually negative: $-\$1.305$

(d) **Variance of the Profit** Variance is calculated using the formula:

$$Var(X) = E(X^2) - (E(X))^2$$

$$E[\text{Profit}^2] = \frac{20}{36} \times 4 + \frac{15}{36} \times 1 + \frac{1}{36} \times 64 = \frac{159}{36} = \frac{53}{12} \approx 4.416$$

$$\text{Var}[\text{Profit}] = \frac{53}{12} - \left(-\frac{47}{36}\right)^2 = 4.416 - 1.704 \approx 2.712$$

Therefore, the variance in profit for the player is \$2.712

7. Continuous Random Variable

Let X be the amount of time a statistics book is checked out by a random student. Suppose X has probability density function:

$$f(x) = \begin{cases} \frac{1}{\log(4)} \cdot \frac{1}{1+x}, & 0 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

(a) **Probability of Student Being Charged a Fine:**

We want to find the probability that the book is loaned for over 2 hours, and within the limits $2 < x < 3$:

$$P(X > 2) = \int_2^3 \frac{1}{\log(4)} \cdot \frac{1}{1+x}\, dx = \frac{1}{\log(4)} \int_2^3 \frac{1}{1+x}\, dx$$

$$= \frac{1}{\log(4)} \left[\log(1+x)\right]_2^3$$

$$= \frac{1}{\log(4)}(\log(4) - \log(3)) = \frac{1}{\log(4)} \left(\log\left(\frac{4}{3}\right)\right)$$

$$= \frac{\log\left(\frac{4}{3}\right)}{\log(4)} \approx 0.21$$

Therefore, the probability that a student checking out the book will be charged a fine is approximately 21%

(b) **Probability that the Fine is at Least \$3.00:**

Probability that the student is fined and then given they are fined, that the fine is at least \$3, $P(X > 2.75 \mid X > 2)$:

$$P(X > 2.75 \mid X > 2) = \frac{P(X > 2.75)}{P(X > 2)}$$

$$= \frac{\frac{1}{\log(4)} \left[\log(1+x)\right]_{2.75}^3}{\frac{\log\left(\frac{4}{3}\right)}{\log(4)}}$$

$$= \frac{\frac{1}{\log(4)}\left(\log(4) - \log(2.75)\right)}{\frac{\log\left(\frac{4}{3}\right)}{\log(4)}}$$

$$= \frac{\log\left(\frac{4}{3.75}\right)}{\log\left(\frac{4}{3}\right)} \approx 0.2243$$

Thus, the probability that the fine is at least \$3 is 22.43%.

8. Independence

(a) **Independence of A_1 and A_3:**

Using independent events formula; if $P(A_1 \cap A_3) = P(A_1) \times P(A_3)$, the events are independent.

$$P(A_1) = \frac{2}{4} = \frac{1}{2}, \quad P(A_3) = \frac{2}{4} = \frac{1}{2}$$

$$P(A_1 \cap A_3) = \frac{1}{4} \quad \text{because there's 4 slips of paper and you choose 1}$$

Since:

$$P(A_1) \times P(A_3) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = P(A_1 \cap A_3)$$

Therefore, the events A_1 and A_3 are independent.

(b) **Independence of A_1, A_2, and A_3:**

We need to check if $P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3)$.

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \quad \text{as there's still 4 slips of paper and you choose 1}$$

$$P(A_1) \times P(A_2) \times P(A_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

Since $\frac{1}{4} \neq \frac{1}{8}$, the events A_1, A_2, and A_3 are not jointly independent.

9. Expectation

(a) Proving $E(\max\{X, Y\}) = E(X) + E(Y) - E(\min\{X, Y\})$:

Using the identity:

$$\max\{X, Y\} + \min\{X, Y\} = X + Y$$

Take expectation of both sides and knowing the property of linearity of expectation:

$$E(\max\{X, Y\}) + E(\min\{X, Y\}) = E(X) + E(Y)$$

Rearranging:

$$E(\max\{X,Y\}) = E(X) + E(Y) - E(\min\{X,Y\})$$

(b) **Find the moment generating function of X for a discrete uniform distribution:**

A random variable X has a discrete uniform distribution on the set $\{a, a+1, \dots, b\}$ with probability mass function of X is given by:

$$P(X = x) = \frac{1}{b-a+1}, \quad \text{for } x = a, a+1, \dots, b.$$

The moment generating function $M_X(t)$ of a random variable X is defined as:

$$M_X(t) = E[e^{tX}] = \sum_{x=a}^{b} e^{tx} \cdot P(X = x)$$

Substituting the pmf of X into the sum:

$$\begin{aligned} M_X(t) &= \sum_{x=a}^{b} e^{tx} \cdot \frac{1}{b-a+1} \\ &= \frac{1}{b-a+1} \sum_{x=a}^{b} e^{tx} \end{aligned}$$

$\sum_{x=a}^{b} e^{tx}$ is a geometric series.
The first term of this series is e^{ta}, and the common ratio is e^t.
The number of terms in this series is $b-a+1$.
The sum of a geometric series $\sum_{x=0}^{n-1} r^x$ is given by:

$$S_n = \frac{r^n - 1}{r - 1}$$

Applying this:

$$\sum_{x=a}^{b} e^{tx} = e^{ta} \sum_{k=0}^{b-a} e^{tk} = e^{ta} \cdot \frac{e^{t(b-a+1)} - 1}{e^t - 1}$$

Substituting back into the expression for the moment generating function:

$$M_X(t) = \frac{1}{b-a+1} \cdot \frac{e^{ta}\left(e^{t(b-a+1)} - 1\right)}{e^t - 1}$$

The moment generating function $M_X(t)$ can be written as:

$$M_X(t) = \frac{e^{ta}\left(e^{t(b-a+1)} - 1\right)}{(b-a+1)(e^t - 1)}$$

Thus, this is the moment generating function for a discrete uniform random variable X on the set $\{a, a+1, \dots, b\}$.