SURV686-HW1

Sagnik Chakravarty

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination.

Signature: Sagnik Chakravarty

Date: 01/15/2025

Table of contents

Question 1

The following data are from Google Trends show the number of times that the term “film noir” was searched
using Google

data:

```
# Creating the Week Column for the film noir dataframe
week <- seq(from = as.Date("2022-10-02"),
            to = as.Date("2022-12-18"),
            by = "week")

# Creating the dataframe
film_noir <- data.frame('Week' = week,
                        'Film noir Searches'=c(68,73,58,59,72,70,77,57,56,76,63,52))

print(film_noir, format ='pdf')
```

```
         Week Film.noir.Searches
1  2022-10-02                 68
2  2022-10-09                 73
3  2022-10-16                 58
4  2022-10-23                 59
5  2022-10-30                 72
6  2022-11-06                 70
7  2022-11-13                 77
8  2022-11-20                 57
9  2022-11-27                 56
10 2022-12-04                 76
11 2022-12-11                 63
12 2022-12-18                 52
```

1.a) Calculate maximum likelihood estimate of p (i.e. the proportion of all 781 searches that occurred in each week). Graph these 12 proportions.

code:

```
library(dplyr)
library(ggplot2)
```

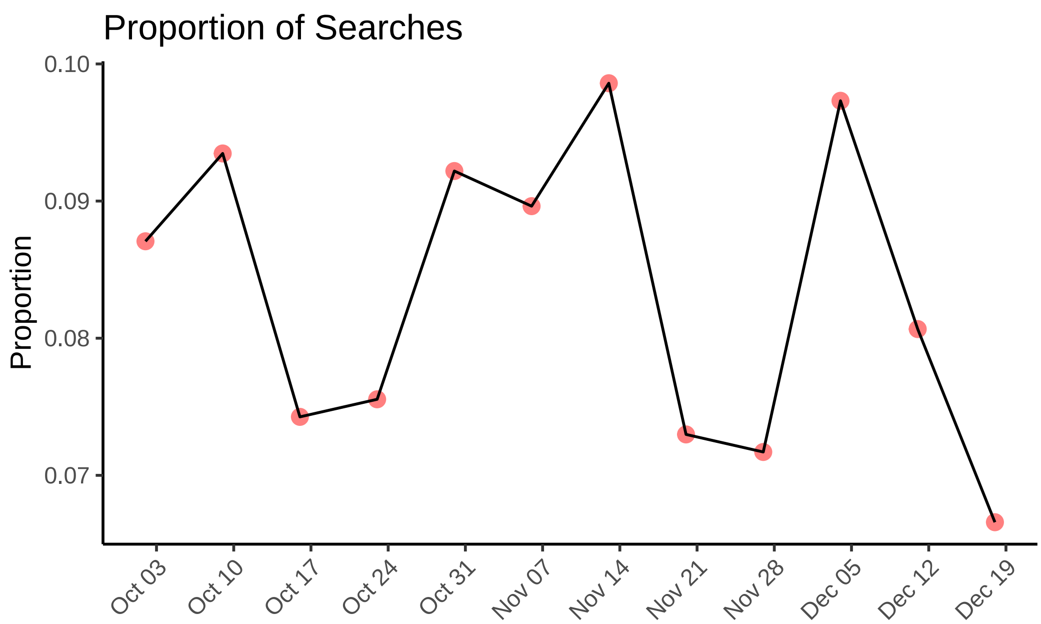
```r
# Checking if the total of the searches are 781 or not
total <- sum(film_noir$Film.noir.Searches)
print(total)
```

```
[1] 781
```

```r
# Creating the Proportion Column
film_noir <- film_noir %>%
  mutate('Proportion(p)'=Film.noir.Searches/total)
print(film_noir, format = 'pdf')
```

```
         Week Film.noir.Searches Proportion(p)
1  2022-10-02                 68    0.08706786
2  2022-10-09                 73    0.09346991
3  2022-10-16                 58    0.07426376
4  2022-10-23                 59    0.07554417
5  2022-10-30                 72    0.09218950
6  2022-11-06                 70    0.08962868
7  2022-11-13                 77    0.09859155
8  2022-11-20                 57    0.07298335
9  2022-11-27                 56    0.07170294
10 2022-12-04                 76    0.09731114
11 2022-12-11                 63    0.08066581
12 2022-12-18                 52    0.06658131
```

```r
# Ploting the proportion for each week
film_noir %>% ggplot(aes(x = Week,
                         y = `Proportion(p)`)) +
  geom_point(col = 'red', size = 2.5, alpha = 0.5) +
  geom_line() +
  scale_x_date(date_labels = "%b %d", date_breaks = "1 week")+
  labs(title = 'Proportion of Searches',
       x = '',
       y = 'Proportion') +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



calculation:

Week	Film noir Searches	Calculation
10/2/22	68	$\frac{68}{781} = 0.08706786$
10/9/22	73	$\frac{73}{781} = 0.09346991$
10/16/22	58	$\frac{58}{781} = 0.07426376$
10/23/22	59	$\frac{59}{781} = 0.07554417$
10/30/22	72	$\frac{72}{781} = 0.09218950$
11/6/22	70	$\frac{70}{781} = 0.08962868$
11/13/22	77	$\frac{77}{781} = 0.09859155$
11/20/22	57	$\frac{57}{781} = 0.07298335$
11/27/22	56	$\frac{56}{781} = 0.07170294$
12/4/22	76	$\frac{76}{781} = 0.09731114$
12/11/22	63	$\frac{63}{781} = 0.08066581$
12/18/22	52	$\frac{52}{781} = 0.06658131$
Total	**781**	

1.b) Write the null hypothesis that the proportion of searches for "film noir" is the same each week. Also, write the alternative hypothesis (i.e., that there has been a change in the proportion of searches each week).

The null and alternative hypotheses are defined as follows:

$$H_0 : p_i = p_j;\ \text{where } i \neq j \,\forall\, i, j \in \text{ weeks in the film noir data, and}$$
$$p_i = \frac{\text{film noir searches in the } i^{th}\text{week}}{\text{total searches}}, \forall i$$

$$H_1 : p_i \neq p_j$$

1.c) Compute the χ^2 and G^2 statistics. What do these tell us?

So to calculate χ^2 we will be calculating the expected value for each week now $E = \frac{\text{Total Searches}}{\text{Number of Weeks}} = \frac{781}{12}$, and we will be substracting this E from each of the observation $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, similarly g value will be calculate using $G^2 = 2\sum O_i ln(\frac{O_i}{E_i})$

code:

```
# Calculating the expected value
expected_value <- total/nrow(film_noir)
print(expected_value)
```

```
[1] 65.08333
```

```
# Calculating the chi square and g square value for each week
film_noir <- film_noir %>%
  mutate(x_square = (Film.noir.Searches - expected_value)^2/expected_value,
         g_square = 2*Film.noir.Searches*log(Film.noir.Searches/expected_value, base = exp

print(film_noir, format = 'pdf')
```

```
        Week Film.noir.Searches Proportion(p)   x_square   g_square
1  2022-10-02                 68    0.08706786 0.13070849   5.962132
2  2022-10-09                 73    0.09346991 0.96297482  16.759477
3  2022-10-16                 58    0.07426376 0.77091336 -13.366157
4  2022-10-23                 59    0.07554417 0.56860862 -11.579465
5  2022-10-30                 72    0.09218950 0.73506189  14.543657
```

```
6   2022-11-06                 70     0.08962868 0.37142552  10.195744
7   2022-11-13                 77     0.09859155 2.18192488  25.893086
8   2022-11-20                 57     0.07298335 1.00394793 -15.118364
9   2022-11-27                 56     0.07170294 1.26771233 -16.835483
10  2022-12-04                 76     0.09731114 1.83109262  23.569856
11  2022-12-11                 63     0.08066581 0.06668801  -4.099255
12  2022-12-18                 52     0.06658131 2.63006829 -23.340177
```

```r
# The total Chi Square
chi_square <- sum(film_noir$x_square)
g_square <- sum(film_noir$g_square)


# getting the critical value
critical_value_0.05 <- qchisq(p = 0.05,
                        df = nrow(film_noir) -1,
                        lower.tail = FALSE)
critical_value_0.1 <- qchisq(p = 0.1,
                        df = nrow(film_noir) -1,
                        lower.tail = FALSE)

cat('Chi Square:\t\t', round(chi_square, 2),
    '\nG square:\t\t', round(g_square, 2),
    '\nCritical value at 0.05: ', round(critical_value_0.05, 2),
    '\nCritical value at 0.1:  ', round(critical_value_0.1, 2))
```

```
Chi Square:      12.52
G square:        12.59
Critical value at 0.05:  19.68
Critical value at 0.1:   17.28
```

Calculations:

Week	Search	Proportion	χ^2 Calculation	G^2 Calculation
10/2/22	68	$\frac{68}{781} =$ 0.08706786	$\frac{(68-65.0833)^2}{65.0833} =$ 0.1307	$2 \times 68 \times \ln(\frac{68}{65.0833}) =$ 5.9621
10/9/22	73	$\frac{73}{781} =$ 0.09346991	$\frac{(73-65.0833)^2}{65.0833} =$ 0.9630	$2 \times 73 \times \ln(\frac{73}{65.0833}) =$ 16.7595

Week	Search	Proportion	χ^2 Calculation	G^2 Calculation
10/16/22	58	$\frac{58}{781} = 0.07426376$	$\frac{(58-65.0833)^2}{65.0833} = 0.7709$	$2 \times 58 \times \ln(\frac{58}{65.0833}) = -13.3662$
10/23/22	59	$\frac{59}{781} = 0.07554417$	$\frac{(59-65.0833)^2}{65.0833} = 0.5686$	$2 \times 59 \times \ln(\frac{59}{65.0833}) = -11.5795$
10/30/22	72	$\frac{72}{781} = 0.09218950$	$\frac{(72-65.0833)^2}{65.0833} = 0.7351$	$2 \times 72 \times \ln(\frac{72}{65.0833}) = 14.5437$
11/6/22	70	$\frac{70}{781} = 0.08962868$	$\frac{(70-65.0833)^2}{65.0833} = 0.3714$	$2 \times 70 \times \ln(\frac{70}{65.0833}) = 10.1957$
11/13/22	77	$\frac{77}{781} = 0.09859155$	$\frac{(77-65.0833)^2}{65.0833} = 2.1819$	$2 \times 77 \times \ln(\frac{77}{65.0833}) = 25.8931$
11/20/22	57	$\frac{57}{781} = 0.07298335$	$\frac{(57-65.0833)^2}{65.0833} = 1.0039$	$2 \times 57 \times \ln(\frac{57}{65.0833}) = -15.1184$
11/27/22	56	$\frac{56}{781} = 0.07170294$	$\frac{(56-65.0833)^2}{65.0833} = 1.2677$	$2 \times 56 \times \ln(\frac{56}{65.0833}) = -16.8355$
12/4/22	76	$\frac{76}{781} = 0.09731114$	$\frac{(76-65.0833)^2}{65.0833} = 1.8311$	$2 \times 76 \times \ln(\frac{76}{65.0833}) = 23.5699$
12/11/22	63	$\frac{63}{781} = 0.08066581$	$\frac{(63-65.0833)^2}{65.0833} = 0.0667$	$2 \times 63 \times \ln(\frac{63}{65.0833}) = -4.0993$
12/18/22	52	$\frac{52}{781} = 0.06658131$	$\frac{(52-65.0833)^2}{65.0833} = 2.6301$	$2 \times 52 \times \ln(\frac{52}{65.0833}) = -23.3402$

χ^2 = 0.1307 + 0.9630 + 0.7709 + 0.5686 + 0.7351 + 0.3714 + 2.1819 + 1.0039 + 1.2677 + 1.8311 + 0.0667 + 2.6301 = 12.52

G^2 = 5.9621 + 16.7595 - 13.3662 - 11.5795 + 14.5437 + 10.1957 + 25.8931 - 15.1184 - 16.8355 + 23.5699 - 4.0993 - 23.3402 = 12.59

Interpretation:

The computed values for the Chi-squared statistic (χ^2=12.52) and the Likelihood Ratio statistic (G^2=12.59) are both **less than the critical value** of 19.68 at a 95% significance level (α = 0.05) with df=11. Additionally, these values are also below the critical value of 17.28 at a 90% significance level ($\alpha = 0.10$).

1. **Failing to Reject the Null Hypothesis H_0**

   - Since both test statistics are less than the critical value, we fail to reject the null hypothesis $H_0 : p_i = p_j$ at both 95and 90% confidence levels.
   - This indicates that there is **no statistically significant evidence** to suggest that the proportions of searches for "film noir" differ across weeks.

2. **Practical Implications:**

   - From a practical perspective, this result implies that there is no clear trend or pattern in search interest for "film noir" during the given time period (October to December 2022). The search behavior appears stable across weeks.

3. **Comparison of Test Statistics:**

   - The similarity between $\chi^2 = 12.52$ and $G^2 = 12.59$ suggests that both tests lead to consistent conclusions, reinforcing the robustness of the result.

Question 2

A graduate student decided to track the number of steps they took each day for a week. The student took a walk every afternoon. The student also walked to class and other places. The student wanted to know if they were walking about the same number of steps each day. Here are the data on steps tracked

data:

```r
library(lubridate)
# Creating the walk dataframe containing the days and the steps taken
walk <- data.frame(Days = wday(1:7,
                               label = TRUE,
                               abbr = TRUE,
                               week_start = 1),
                   Steps = c(3358, 2894, 2346, 2981, 2956, 2239, 3974))

print(walk, format = 'pdf')
```

```
  Days Steps
1  Sun  3358
2  Mon  2894
3  Tue  2346
4  Wed  2981
5  Thu  2956
6  Fri  2239
7  Sat  3974
```

```r
# Calculating the total steps taken
total <- sum(walk$Steps)
cat('Total Steps:\t', total)
```

```
Total Steps:     20748
```

The student wants to be walking about the same number of steps each day. Hence, the null hypothesis is that the
number of steps are equally likely to be walked on each day, or that the daily proportion of each weeks total
steps is the same:

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7$$

$$H_1 : p_1 \neq p_2 \neq p_3 \neq p_4 \neq p_5 \neq p_6 \neq p_7$$

2.a) Graph the proportions of all steps taken on each day of the week

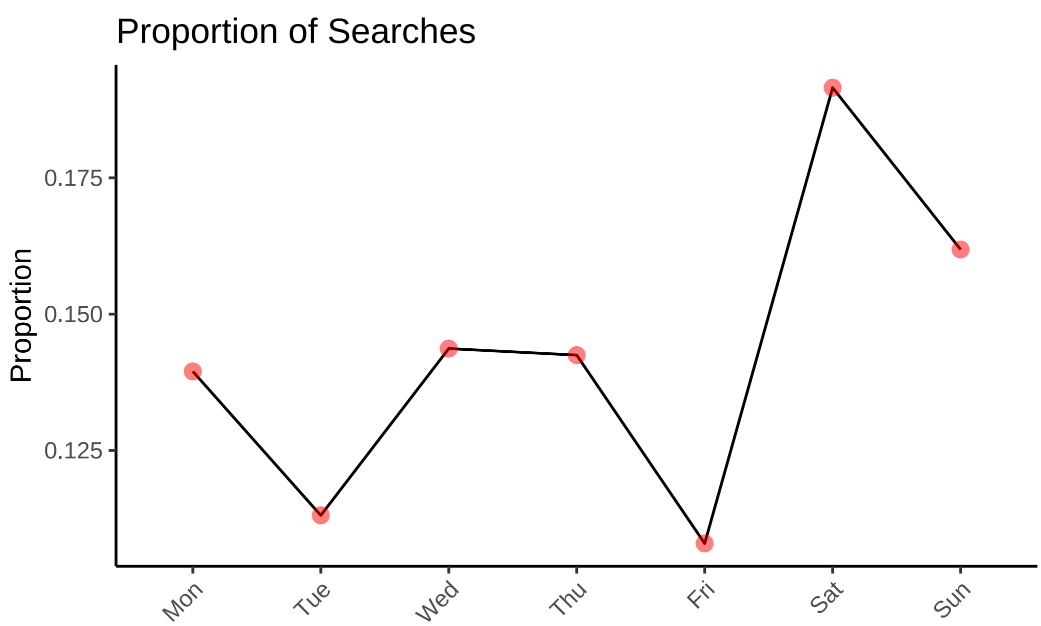
Code:

```r
# Calculating the proportion of steps taken
walk <- walk %>% mutate(Proportion = Steps/total)
print(walk, format = 'pdf')
```

```
  Days Steps Proportion
1  Sun  3358  0.1618469
2  Mon  2894  0.1394833
3  Tue  2346  0.1130711
4  Wed  2981  0.1436765
5  Thu  2956  0.1424716
```

```
6   Fri   2239   0.1079140
7   Sat   3974   0.1915365
```

```
# Ploting the proportion
walk %>%
  ggplot(aes(x = Days, y = Proportion, group = 1)) +
  geom_line() +
  geom_point(col = 'red', size = 2.5, alpha = 0.5) +
  labs(title = 'Proportion of Searches',
       x = '',
       y = 'Proportion') +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Calculation:

Day	Steps	Proportion
Sun	3358	$\frac{3358}{20748} = 0.1618469$
Mon	2894	$\frac{2894}{20748} = 0.1394833$
Tue	2346	$\frac{2346}{20748} = 0.1130711$
Wed	2981	$\frac{2981}{20748} = 0.1436765$

Day	Steps	Proportion
Thu	2956	$\frac{2956}{20748} = 0.1424716$
Fri	2239	$\frac{2239}{20748} = 0.1079140$
Sat	3974	$\frac{3974}{20748} = 0.1915365$
Total	**20748**	

2. b)Calculate the maximum likelihood estimate of p, as well as the maximum likelihood estimate of $\hat{V}(\hat{p})$. Note that the latter [$\hat{V}(\hat{p})$] is a matrix of variances and covariances

The MLE is the proportions for the steps hence its:

$\hat{p} = [0.1618469, 0.1394833, 0.1130711, 0.1436765, 0.1424716, 0.1079140, 0.1915365]$

$\hat{V}(\hat{p})$: is calculated as follows:

$$\hat{V}(\hat{p}) = \frac{1}{n}[diag(\tilde{p}) - \tilde{p}.\tilde{p}']$$

Code:

```
mle <- walk$Proportion
p <- matrix(mle, ncol = 1)

# Calculating the covariance matrix
variance_matrix <- 1/total*(diag(mle)-p%*%t(p))
print(variance_matrix, format = 'pdf')
```

```
              [,1]          [,2]          [,3]          [,4]          [,5]
[1,]  6.538100e-06 -1.088054e-06 -8.820231e-07 -1.120763e-06 -1.111364e-06
[2,] -1.088054e-06  5.785026e-06 -7.601474e-07 -9.658992e-07 -9.577987e-07
[3,] -8.820231e-07 -7.601474e-07  4.833529e-06 -7.829991e-07 -7.764325e-07
[4,] -1.120763e-06 -9.658992e-07 -7.829991e-07  5.929900e-06 -9.865922e-07
[5,] -1.111364e-06 -9.577987e-07 -7.764325e-07 -9.865922e-07  5.888443e-06
[6,] -8.417945e-07 -7.254774e-07 -5.881030e-07 -7.472869e-07 -7.410198e-07
[7,] -1.494101e-06 -1.287650e-06 -1.043824e-06 -1.326359e-06 -1.315236e-06
              [,6]          [,7]
[1,] -8.417945e-07 -1.494101e-06
[2,] -7.254774e-07 -1.287650e-06
[3,] -5.881030e-07 -1.043824e-06
[4,] -7.472869e-07 -1.326359e-06
[5,] -7.410198e-07 -1.315236e-06
```

```
[6,]  4.639897e-06 -9.962154e-07
[7,] -9.962154e-07  7.463384e-06
```

Calculation:

$$\hat{V}(\hat{p}) = \frac{1}{20748}\left(\begin{bmatrix} 0.1618469 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1394833 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1130711 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1436765 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1424716 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1079140 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1915365 \end{bmatrix} - \begin{bmatrix} 0.1618469 \\ 0.1394833 \\ 0.1130711 \\ 0.1436765 \\ 0.1424716 \\ 0.1079140 \\ 0.1915365 \end{bmatrix} [0.16184690.13948330.11307110.14367650.14247160.10791400.1915365]\right)$$

$$\implies \frac{1}{20748}\left(\begin{bmatrix} 0.1618469 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1394833 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1130711 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1436765 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1424716 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1079140 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1915365 \end{bmatrix} - \begin{bmatrix} 0.02619443 & 0.02257495 & 0.01830022 & 0.02325360 & 0.02305858 & 0.01746555 & 0.03099960 \\ 0.02257495 & 0.01945560 & 0.01577154 & 0.02004048 & 0.01987241 & 0.01505221 & 0.02671615 \\ 0.01830022 & 0.01577154 & 0.01278508 & 0.01624567 & 0.01610942 & 0.01220196 & 0.02165725 \\ 0.02325360 & 0.02004048 & 0.01624567 & 0.02064294 & 0.02046982 & 0.01550471 & 0.02751930 \\ 0.02305858 & 0.01987241 & 0.01610942 & 0.02046982 & 0.02029815 & 0.01537468 & 0.02728851 \\ 0.01746555 & 0.01505221 & 0.01220196 & 0.01550471 & 0.01537468 & 0.01164543 & 0.02066948 \\ 0.03099960 & 0.02671615 & 0.02165725 & 0.02751930 & 0.02728851 & 0.02066948 & 0.03668624 \end{bmatrix}\right)$$

$$\implies \frac{1}{20748}\begin{bmatrix} 0.02619443 & 0.02257495 & 0.01830022 & 0.02325360 & 0.02305858 & 0.01746555 & 0.03099960 \\ 0.02257495 & 0.01945560 & 0.01577154 & 0.02004048 & 0.01987241 & 0.01505221 & 0.02671615 \\ 0.01830022 & 0.01577154 & 0.01278508 & 0.01624567 & 0.01610942 & 0.01220196 & 0.02165725 \\ 0.02325360 & 0.02004048 & 0.01624567 & 0.02064294 & 0.02046982 & 0.01550471 & 0.02751930 \\ 0.02305858 & 0.01987241 & 0.01610942 & 0.02046982 & 0.02029815 & 0.01537468 & 0.02728851 \\ 0.01746555 & 0.01505221 & 0.01220196 & 0.01550471 & 0.01537468 & 0.01164543 & 0.02066948 \\ 0.03099960 & 0.02671615 & 0.02165725 & 0.02751930 & 0.02728851 & 0.02066948 & 0.03668624 \end{bmatrix}$$

$$\implies \begin{bmatrix} 6.5381\times10^{-6} & -1.0881\times10^{-6} & -8.8202\times10^{-7} & -1.1208\times10^{-6} & -1.1114\times10^{-6} & -8.4179\times10^{-7} & -1.4941\times10^{-6} \\ -1.0881\times10^{-6} & 5.7850\times10^{-6} & -7.6015\times10^{-7} & -9.6590\times10^{-7} & -9.5780\times10^{-7} & -7.2548\times10^{-7} & -1.2877\times10^{-6} \\ -8.8202\times10^{-7} & -7.6015\times10^{-7} & 4.8335\times10^{-6} & -7.8300\times10^{-7} & -7.7643\times10^{-7} & -5.8810\times10^{-7} & -1.0438\times10^{-6} \\ -1.1208\times10^{-6} & -9.6590\times10^{-7} & -7.8300\times10^{-7} & 5.9299\times10^{-6} & -9.8659\times10^{-7} & -7.4729\times10^{-7} & -1.3264\times10^{-6} \\ -1.1114\times10^{-6} & -9.5780\times10^{-7} & -7.7643\times10^{-7} & -9.8659\times10^{-7} & 5.8884\times10^{-6} & -7.4102\times10^{-7} & -1.3152\times10^{-6} \\ -8.4179\times10^{-7} & -7.2548\times10^{-7} & -5.8810\times10^{-7} & -7.4729\times10^{-7} & -7.4102\times10^{-7} & 4.6399\times10^{-6} & -9.9622\times10^{-7} \\ -1.4941\times10^{-6} & -1.2877\times10^{-6} & -1.0438\times10^{-6} & -1.3264\times10^{-6} & -1.3152\times10^{-6} & -9.9622\times10^{-7} & 7.4634\times10^{-6} \end{bmatrix}$$

2.c) Calculate the maximum likelihood estimate of the proportion of steps taken on the weekend (Sunday and Saturday, $p_1 + p_7$) and the maximum likelihood estimate of the variance of the proportion of steps taken on the weekend

code:

```r
# Calculating the mle at the weekend
weekend_mle <- mle[1] + mle[7]

# Calculating the mle of variance at the weekend
weekend_variance <- variance_matrix[1,1] + variance_matrix[7,7] + 2*variance_matrix[1,7]

# Displaying the result
cat('MLE estimate for steps in weekend:\t\t\t', weekend_mle,
    '\nMLE estimate for the variance of steps in weekend:\t', weekend_variance)
```

```
MLE estimate for steps in weekend:            0.3533835
MLE estimate for the variance of steps in weekend:   1.101328e-05
```

Calculation:

The MLE for the weekend is: $p_1 + p_7 = 0.1618469 + 0.1915365 = 0.3533835$

The variance is given as

$$
\begin{aligned}
V(p_1+p_7) = V(p_1) + V(p_2) + 2 \times cov(p_1, p_2) \\
&\implies 6.5381e-06 + 7.463384e-06 + 2 \times -1.494101e-06 \\
&\implies 1.400148e-05 - 2.988201e-06 \\
&\implies 1.101328e-05
\end{aligned}
$$

2. d) Test the null and alternate hypothesis by computing both the χ^2 and G^2 statistics. What do you conclude?

<u>**code:**</u>

```
# Calculating the expected steps
expected_steps <- total/7
print(expected_steps)
```

```
[1] 2964
```

```
# Calculating the x square and g square for each observation
walk <- walk %>%
  mutate(x_square = (Steps - expected_steps)^2/expected_steps,
         g_square = 2*Steps*log(Steps/expected_steps, base = exp(1)))

print(walk, format = 'pdf')
```

```
  Days Steps Proportion     x_square    g_square
1  Sun  3358  0.1618469  52.37381916   838.19610
2  Mon  2894  0.1394833   1.65317139  -138.33366
3  Tue  2346  0.1130711 128.85425101 -1097.12078
4  Wed  2981  0.1436765   0.09750337    34.09732
5  Thu  2956  0.1424716   0.02159244   -15.97839
6  Fri  2239  0.1079140 177.33636977 -1256.12544
7  Sat  3974  0.1915365 344.16329285  2330.61935
```

```
# Calculating the chi square and g square value
chi_square <- sum(walk$x_square)
g_square <- sum(walk$g_square)

# Calculating the critical values at 95 and 90% level of significance
critical_value_0.05 <- qchisq(p = 0.05,
```

```
                              df = nrow(walk) -1,
                              lower.tail = FALSE)
critical_value_0.1 <- qchisq(p = 0.1,
                              df = nrow(walk) -1,
                              lower.tail = FALSE)

# Displaying the results
cat('Chi Square:\t\t', round(chi_square, 2),
    '\nG square:\t\t', round(g_square, 2),
    '\nCritical value at 0.05: ', round(critical_value_0.05, 2),
    '\nCritical value at 0.1:  ', round(critical_value_0.1, 2))
```

```
Chi Square:      704.5
G square:        695.35
Critical value at 0.05:  12.59
Critical value at 0.1:   10.64
```

Calculation:

Since we are considering all the proportion to be same for the null hypothesis the expected proportion are as follows

$E = \frac{\text{total steps}}{\text{total days}} = \frac{20748}{7} = 2964$

Then we calculate the chi square and g square values same as we did in question 1

Days	**Steps**	**Proportion (P)**	x^2	G^2
Sun	3358	$\frac{3358}{20748} = 0.1618469$	$\frac{(3358-2964)^2}{2964} = 52.37381916$	$2 \times 3358 \ln\left(\frac{3358}{2964}\right) = 838.1961$
Mon	2894	$\frac{2894}{20748} = 0.1394833$	$\frac{(2894-2964)^2}{2964} = 1.65317139$	$2 \times 2894 \ln\left(\frac{2894}{2964}\right) = -138.33366$
Tue	2346	$\frac{2346}{20748} = 0.1130711$	$\frac{(2346-2964)^2}{2964} = 128.854251$	$2 \times 2346 \ln\left(\frac{2346}{2964}\right) = -1097.12078$
Wed	2981	$\frac{2981}{20748} = 0.1436765$	$\frac{(2981-2964)^2}{2964} = 0.09750337$	$2 \times 2981 \ln\left(\frac{2981}{2964}\right) = 34.09732$
Thu	2956	$\frac{2956}{20748} = 0.1424716$	$\frac{(2956-2964)^2}{2964} = 0.02159244$	$2 \times 2956 \ln\left(\frac{2956}{2964}\right) = -15.97839$
Fri	2239	$\frac{2239}{20748} = 0.107914$	$\frac{(2239-2964)^2}{2964} = 177.3363698$	$2 \times 2239 \ln\left(\frac{2239}{2964}\right) = -1256.12544$
Sat	3974	$\frac{3974}{20748} = 0.1915365$	$\frac{(3974-2964)^2}{2964} = 344.1632929$	$2 \times 3974 \ln\left(\frac{3974}{2964}\right) = 2330.61935$

$\chi^2 = 52.37 + 1.65 + 128.85 + 0.1 + 0.02 + 177.34 + 344.16 = 704.5$

$G^2 = 838.2 + -138.33 + -1097.12 + 34.1 + -15.98 + -1256.13 + 2330.6 = 695.35$

Interpretation:

- Both the χ^2 and G^2 statistics far exceed the critical values at both significance levels (704.5>12.59) and 695.35>12.59).

- This provides strong evidence to **reject the null hypothesis** (H0).

Practical Implication

- There is a statistically significant difference in the number of steps taken across different days of the week.
- The walking behavior varies significantly by day, and the student does not walk the same number of steps each day.

Question 3

The following table is based on a study of aspirin use and myocardial infarction. The data are similar to actual data

data:

```
mycoridal <- data.frame(Group = c('Placebo', "Aspirin"),
                        Yes = c(173, 83),
                        No = c(9879, 9970))
print(mycoridal, format = 'pdf')
```

```
    Group Yes   No
1 Placebo 173 9879
2 Aspirin  83 9970
```

3.a) About 1.27% $(n_{11} + n_{21})/(n_{11} + n_{21} + n_{12} + n_{22})$ **had myocardial infarction(MI). Since this was a designed experiment, 50% were assigned to take a placebo. If the use of aspirin or placebo was independent of risk of myocardial infarction (i.e. if the risk of myocardial infarction was no different whether you took placebo or aspirin), what would the expected counts be in each cell (n11, n12, n21, and n22)?**

code:

```
# Observed counts
n11 <- 173  # Placebo, Yes (MI)
n12 <- 9879 # Placebo, No (MI)
n21 <- 83   # Aspirin, Yes (MI)
n22 <- 9970 # Aspirin, No (MI)

# Calculate totals
grand_total <- n11 + n12 + n21 + n22
```

```r
row_total_placebo <- n11 + n12
row_total_aspirin <- n21 + n22
col_total_yes <- n11 + n21
col_total_no <- n12 + n22

# Calculate expected counts
E11 <- (row_total_placebo * col_total_yes) / grand_total
E12 <- (row_total_placebo * col_total_no) / grand_total
E21 <- (row_total_aspirin * col_total_yes) / grand_total
E22 <- (row_total_aspirin * col_total_no) / grand_total

# Display results
cat("Expected Counts:\n")
```

```
Expected Counts:
```

```r
cat("Placebo, Yes (MI):", round(E11, 2), "\n")
```

```
Placebo, Yes (MI): 127.99
```

```r
cat("Placebo, No (MI):", round(E12, 2), "\n")
```

```
Placebo, No (MI): 9924.01
```

```r
cat("Aspirin, Yes (MI):", round(E21, 2), "\n")
```

```
Aspirin, Yes (MI): 128.01
```

```r
cat("Aspirin, No (MI):", round(E22, 2), "\n")
```

```
Aspirin, No (MI): 9924.99
```

calculation:

Formula to calculate expected count as:

$$E_{ij} = \frac{\text{Row Total}_i \times \text{Col Total}_j}{\text{Grand Total}}$$

where:

- E_{ij} : is the expected count for the cell in row i and column j.
- Row Total_i : is the total count for row i (e.g., Placebo or Aspirin).
- Col Total_j : is the total count for column j (e.g., Yes or No for myocardial infarction).
- Grand Total: is the total number of participants.

1. Grand Total: $n_{11} + n_{21} + n_{12} + n_{22}$ = 173 + 9879 + 83 + 9970 = 20105
2. Row Totals:
    1. Row Total(Placebo): $n_{11} + n_{12}$ = 173 + 9879 = 10052
    2. Row Total(Aspirin): $n_{21} + n_{22}$ = 83 + 9970 = 10053
3. Col Totals:
    1. Col Total(Yes): $n_{11} + n_{21}$ = 173 + 83 = 256
    2. Col Total(No): $n_{12} + n_{22}$ = 9879 + 9970 = 19849
4. Compute Expected Value
    1. E_{11} Placebo Yes: $E_{11} = \frac{10052 \times 256}{20105} = \frac{2573312}{20105} = 127.99$
    2. E_{12} Placebo No: $E_{12} = \frac{10052 \times 19849}{20105} = \frac{199522148}{20105} = 9924.01$
    3. E_{21} Aspirin Yes: $E_{21} = \frac{10053 \times 256}{20105} = \frac{2573568}{20105} = 128.01$
    4. E_{22} Aspirin No: $E_{22} = \frac{10053 \times 19849}{20105} = \frac{199541997}{20105} = 9924.99$