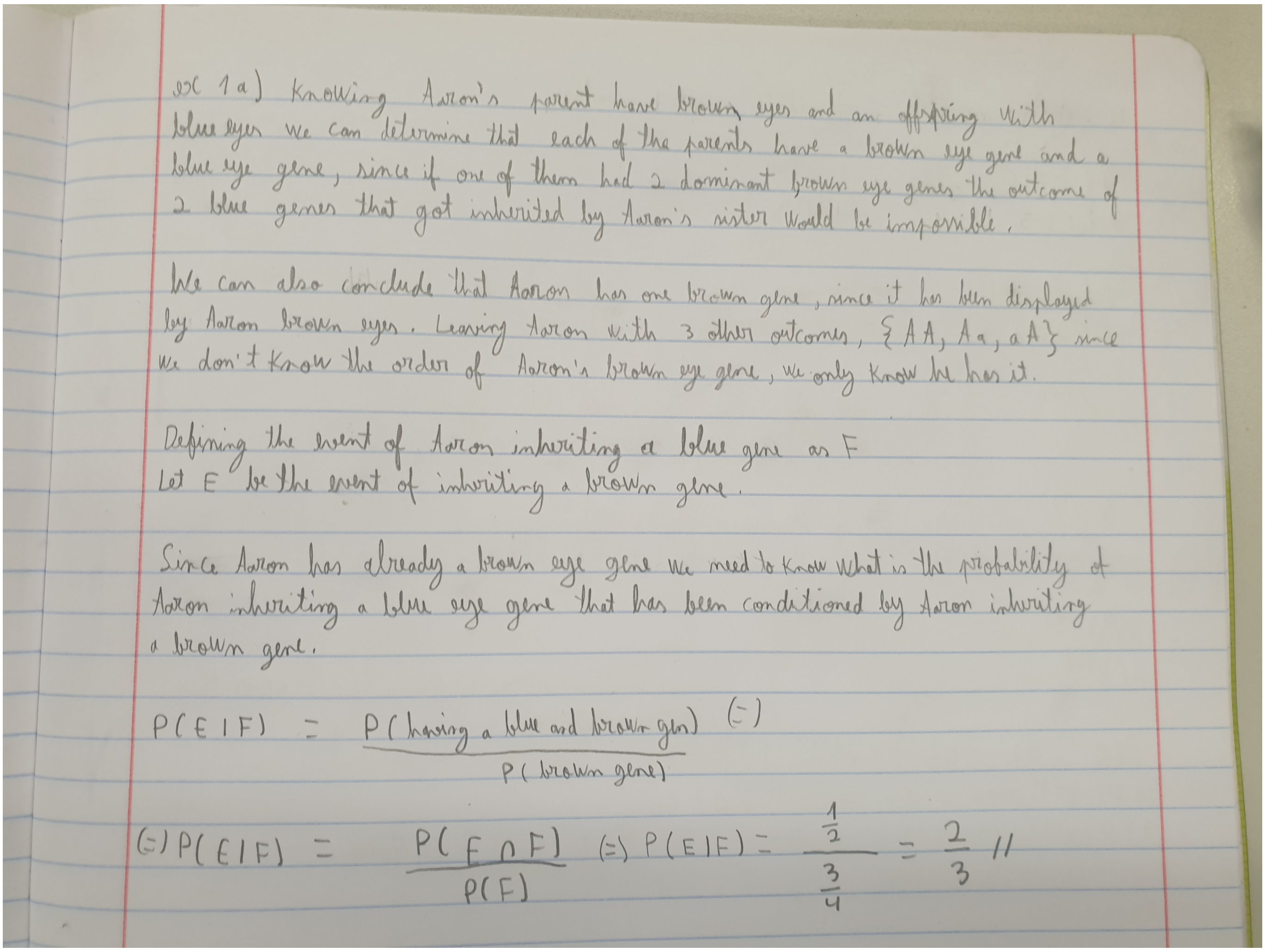
Stats

Table of contents


```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6      v dplyr   1.0.10
v tibble  3.1.8      v stringr 1.4.1
v readr   2.1.3      v forcats 0.5.2
v purrr   0.3.4
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2

Rows: 111 Columns: 4
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (4): radiation, temperature, wind, ozone

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

0.1 Question 1 a)



ex 1 a) Knowing Aaron's parent have brown eyes and an offspring with blue eyes we can determine that each of the parents have a brown eye gene and a blue eye gene, since if one of them had 2 dominant brown eye genes the outcome of 2 blue genes that got inherited by Aaron's sister would be impossible.

We can also conclude that Aaron has one brown gene, since it has been displayed by Aaron brown eyes. Leaving Aaron with 3 other outcomes, $\{AA, Aa, aA\}$ since we don't know the order of Aaron's brown eye gene, we only know he has it.

Defining the event of Aaron inheriting a blue gene as F
Let E be the event of inheriting a brown gene.

Since Aaron has already a brown eye gene we need to know what is the probability of Aaron inheriting a blue eye gene that has been conditioned by Aaron inheriting a brown gene.

$$P(E \mid F) = \frac{P(\text{having a blue and brown gen})}{P(\text{brown gene})} \Leftrightarrow$$

$$\Leftrightarrow P(E \mid F) = \frac{P(E \cap F)}{P(F)} \Leftrightarrow P(E \mid F) = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3} \ //$$

Figure 1: Exercise 1 a)

0.2 Question 1 b)

0.3 Question 1 c)

Part 1 of exercise 1 c)

Part 2 of exercise 1 c)

0.4 Question 2 a)

Part 1 of exercise 2 a)

Part 2 of exercise 2 a)

1 b) In the previous exercise we have proven that a $\frac{2}{3}$ chance of blue eyes which is the 2 outcomes $\{aA, Aa\}$ out of possible sample of $\{AA, aA, aA\}$. Since Aaron's wife has blue eyes she can only have 1 gene combination $\{aa\}$.
Notice that out of the 3 possible Aaron genes, they are all equally likely.

		Aaron's genes		
		A A	A a	a A
Aaron's wife genes	a	Aa aA	Aa aa	aa Aa
	a	aA Aa	aA aa	aa aA

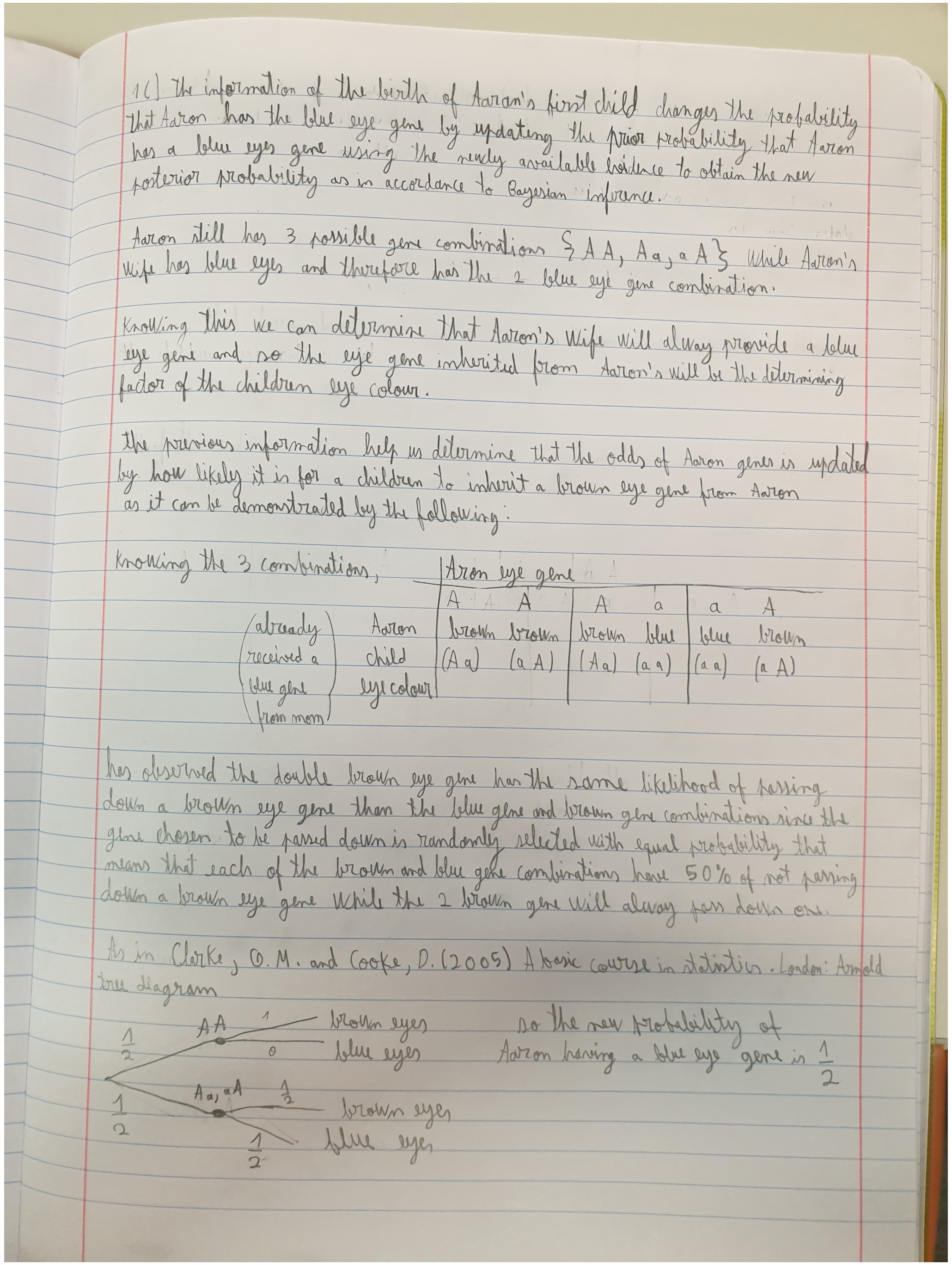
Again since the gene is chosen randomly with equal probability each of 12 possible combination listed above $\{Aa, aA, aA, Aa, Aa, aa, aA, aa, aa, Aa, aa, aA\}$ each of these have a equal probabability of $\frac{1}{12}$.

Knowing all the outcomes and their respective probability we can sum up the number of blue eye combinations defined by the $\{aa\}$ gene combination.

As observed from the listed outcomes 4 outcomes lead to a probability for blue eyes of

$$P(\text{blue eyes}) = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{4}{12} = \frac{1}{3} \quad //$$

Figure 2: Exercise 1 b)



1(c) The information of the birth of Aaron's first child changes the probability that Aaron has the blue eye gene by updating the prior probability that Aaron has a blue eyes gene using the newly available evidence to obtain the new posterior probability as in accordance to Bayesian inference.

Aaron still has 3 possible gene combinations {A A, A a, a A} while Aaron's wife has blue eyes and therefore has the 2 blue eye gene combination.

Knowing this we can determine that Aaron's wife will alway provide a blue eye gene and so the eye gene inherited from Aaron's will be the determining factor of the children eye colour.

the previous information help us determine that the odds of Aaron genes is updated by how likely it is for a children to inherit a brown eye gene from Aaron as it can be demonstrated by the following:

Knowing the 3 combinations,

		Aron eye gene					
		A	A	A	a	a	A
(already received a blue gene from mom)	Aaron child eye colour	brown (A a)	brown (a A)	brown (A a)	blue (a a)	blue (a a)	brown (a A)

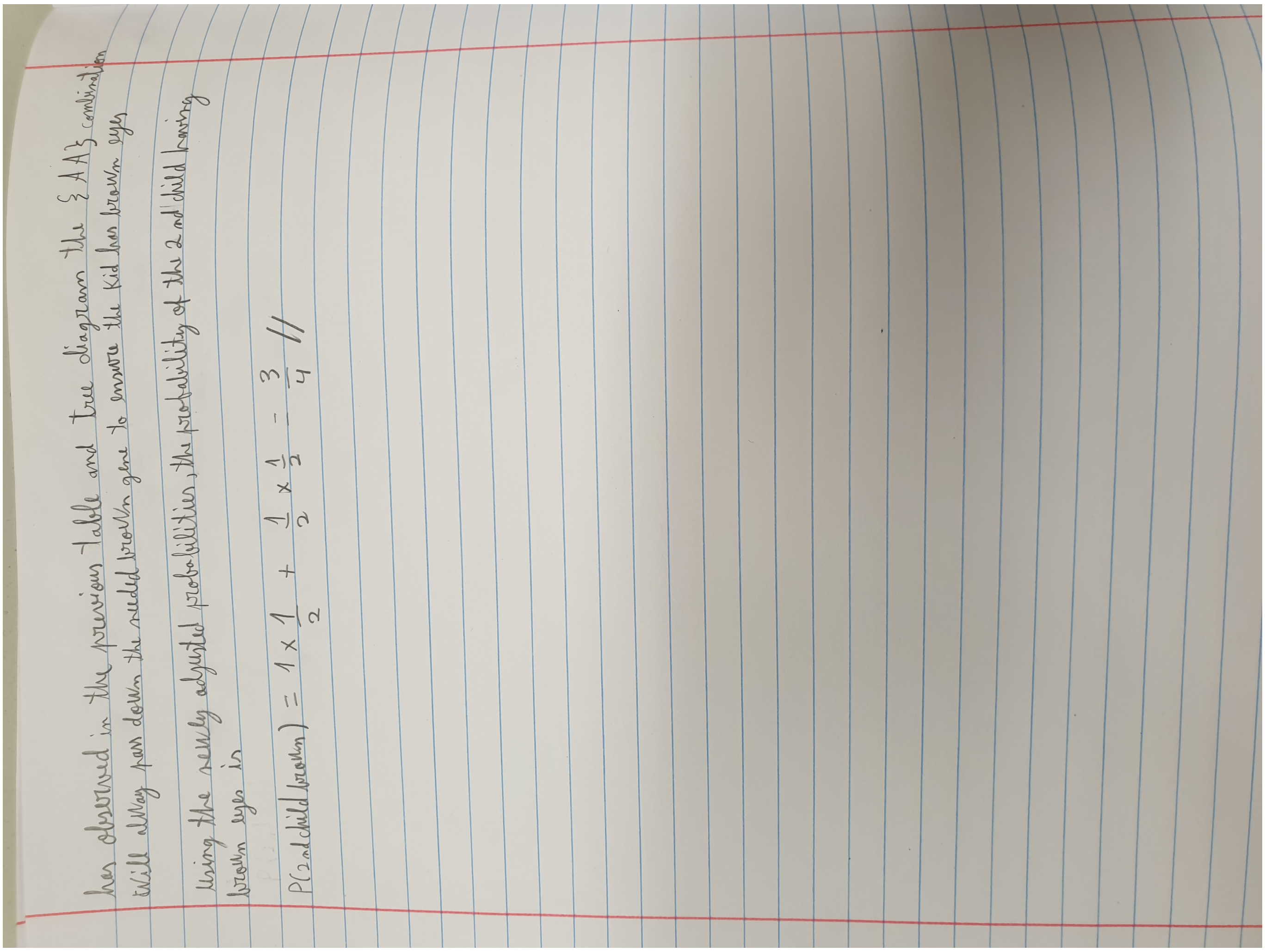
has observed the double brown eye gene has the same likelihood of passing down a brown eye gene than the blue gene and brown gene combinations since the gene chosen to be passed down is randomly selected with equal probability that means that each of the brown and blue gene combinations have 50% of not passing down a brown eye gene while the 2 brown gene will alway pass down one.

As in Clarke, G. M. and Cooke, D. (2005) A basic course in statistics. London: Arnold

tree diagram

so the new probability of Aaron having a blue eye gene is $\frac{1}{2}$

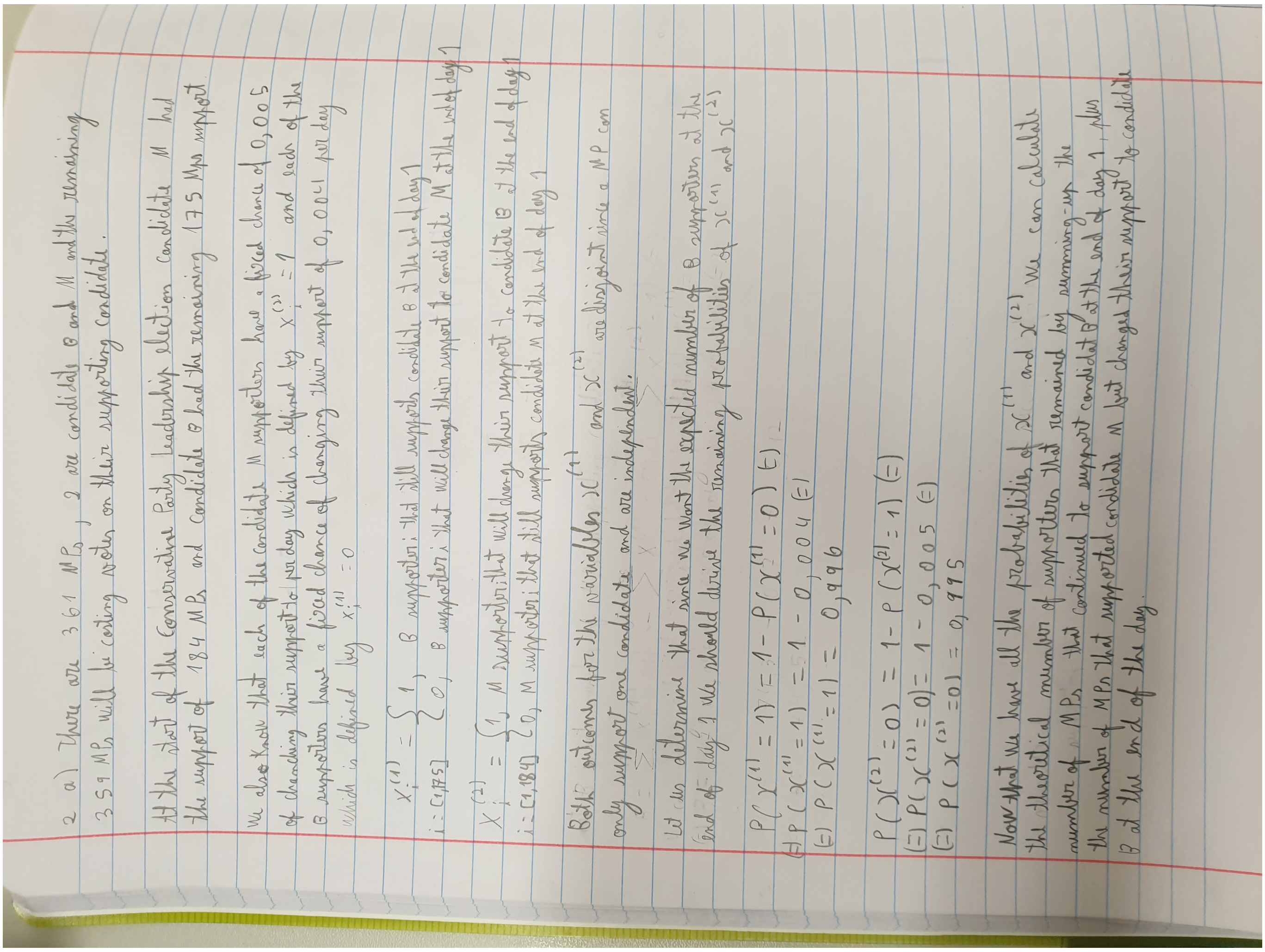
Figure 3: Exercise 1 c) part 1


has observed in the previous table and tree diagram the $\{AA\}$ combination will alWay pass down the needed brown gene to ensure the Kid has brown eyes

Using the newly adjusted probabilities, the probability of the 2nd child having brown eyes is

$$P(\text{2nd child brown}) = 1 \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{3}{4} \ //$$

Figure 4: Exercise 1 c) part 2



2 a) There are 361 MPs, 2 are candidate B and M and the remaining 359 MPs will be casting votes on their supporting candidate.

At the start of the Conservative Party leadership election candidate M had the support of 184 MPs and candidate B had the remaining 175 MPs support.

We also know that each of the candidate M supporters have a fixed chance of 0,005 of changing their support to per day which is defined by $X_i^{(2)} = 1$ and each of the B supporters have a fixed chance of changing their support of 0,004 per day which is defined by $X_i^{(1)} = 0$

$$X_i^{(1)} = \begin{cases} 1, & \text{B supporter}_i \text{ that still supports candidate B at the end of day 1} \\ 0, & \text{B supporter}_i \text{ that will change their support to candidate M at the end of day 1} \end{cases} \quad i = [1,175]$$

$$X_i^{(2)} = \begin{cases} 1, & \text{M supporter}_i \text{ that will change their support to candidate B at the end of day 1} \\ 0, & \text{M supporter}_i \text{ that still supports candidate M at the end of day 1} \end{cases} \quad i = [1,184]$$

Both outcomes for the variables $X^{(1)}$ and $X^{(2)}$ are disjoint since a MP can only support one candidate and are independent.

Let us determine that since we want the expected number of B supporters at the end of day 1 we should derive the remaining probabilities of $X^{(1)}$ and $X^{(2)}$

$P(X^{(1)} = 1) = 1 - P(X^{(1)} = 0) \Leftrightarrow$
$\Leftrightarrow P(X^{(1)} = 1) = 1 - 0{,}004 \Leftrightarrow$
$\Leftrightarrow P(X^{(1)} = 1) = 0{,}996$

$P(X^{(2)} = 0) = 1 - P(X^{(2)} = 1) \Leftrightarrow$
$\Leftrightarrow P(X^{(2)} = 0) = 1 - 0{,}005 \Leftrightarrow$
$\Leftrightarrow P(X^{(2)} = 0) = 0{,}995$

Now that we have all the probabilities of $X^{(1)}$ and $X^{(2)}$ we can calculate the theoretical number of supporters that remained by summing-up the number of MPs that continued to support candidate B at the end of day 1, plus the number of MPs that supported candidate M but changed their support to candidate B at the end of the day.

Figure 5: Exercise 2 a) part 1

Let B be the number of MPs supporting candidate B

$$B = \sum_{i}^{175} (X_i^{(1)} = 1) + \sum_{i}^{184} (X_i^{(2)} = 1) \Leftrightarrow$$

$$\Leftrightarrow B = 175 \times 0{,}996 + 184 \times 0{,}005 \Leftrightarrow$$

$$\Leftrightarrow B = 174{,}3 + 0{,}92$$

$$\Leftrightarrow B = 175{,}22 \; //$$

Figure 6: Exercise 2 a) part 2

0.5 Question 2 C)

```
# set.seed(26041999)

candidateBWins = 0

## Here we set the number of times we will estimate
largeNumberOfExtimations = 1e+06

for (i in 1:largeNumberOfExtimations) {

    candidateB = 175  ## candidate B initial votes/support
    candidateM = 184  ## candidate M initial votes/support

    ## increment for each day from the initial day to the final day,
    ## the 14th
    for (i in 1:14) {

        ## will be used to hold the new value of votes that continue to
        ## support the same candidate
        candidateBAuxCalc = 0
        candidateMAuxCalc = 0

        ## here we will estimate how many candidates still support the
        ## same candidate by generating a vector where 1 means they
        ## still support the same candidate after the end of each day
        ## the 0 represent them changing the support and vote from one
        ## candidate to another
        candidateBAuxCalc = sum(sample(c(1, 0), candidateB, replace = TRUE,
            prob = c(0.996, 0.004)))
        candidateMAuxCalc = sum(sample(c(1, 0), candidateM, replace = TRUE,
            prob = c(0.995, 0.005)))

        ## here we calculate the number of votes that will be exchanged
        ## by the candidates, so the votes that are subtracted from
        ## what they had previously
        votesMovedFromCandidaBToCandidateM = candidateB - candidateBAuxCalc
        votesMovedFromCandidaMToCandidateB = candidateM - candidateMAuxCalc

        ## here calculate that the new current votes for each candidate
        ## by adding the current votes from the mps that still support
        ## the same candidate plus the number of votes of the mps that
        ## exchanged support for their candidate
        candidateB = candidateBAuxCalc + votesMovedFromCandidaMToCandidateB
        candidateM = candidateMAuxCalc + votesMovedFromCandidaBToCandidateM

    }
```

```
    ## check if the candidate B did win the election by holding the
    ## majority of the votes
    if (candidateB > candidateM) {
        candidateBWins = candidateBWins + 1
    }

}

## probability estimation of candidate B winning after 14 days of
## campaign by holding the majority of the votes

probabilityB = candidateBWins/largeNumberOfExtimations

probabilityB
```

```
[1] 0.359058
```

As we can see the probability of candidate B winning after 14 days is approximately 0.35

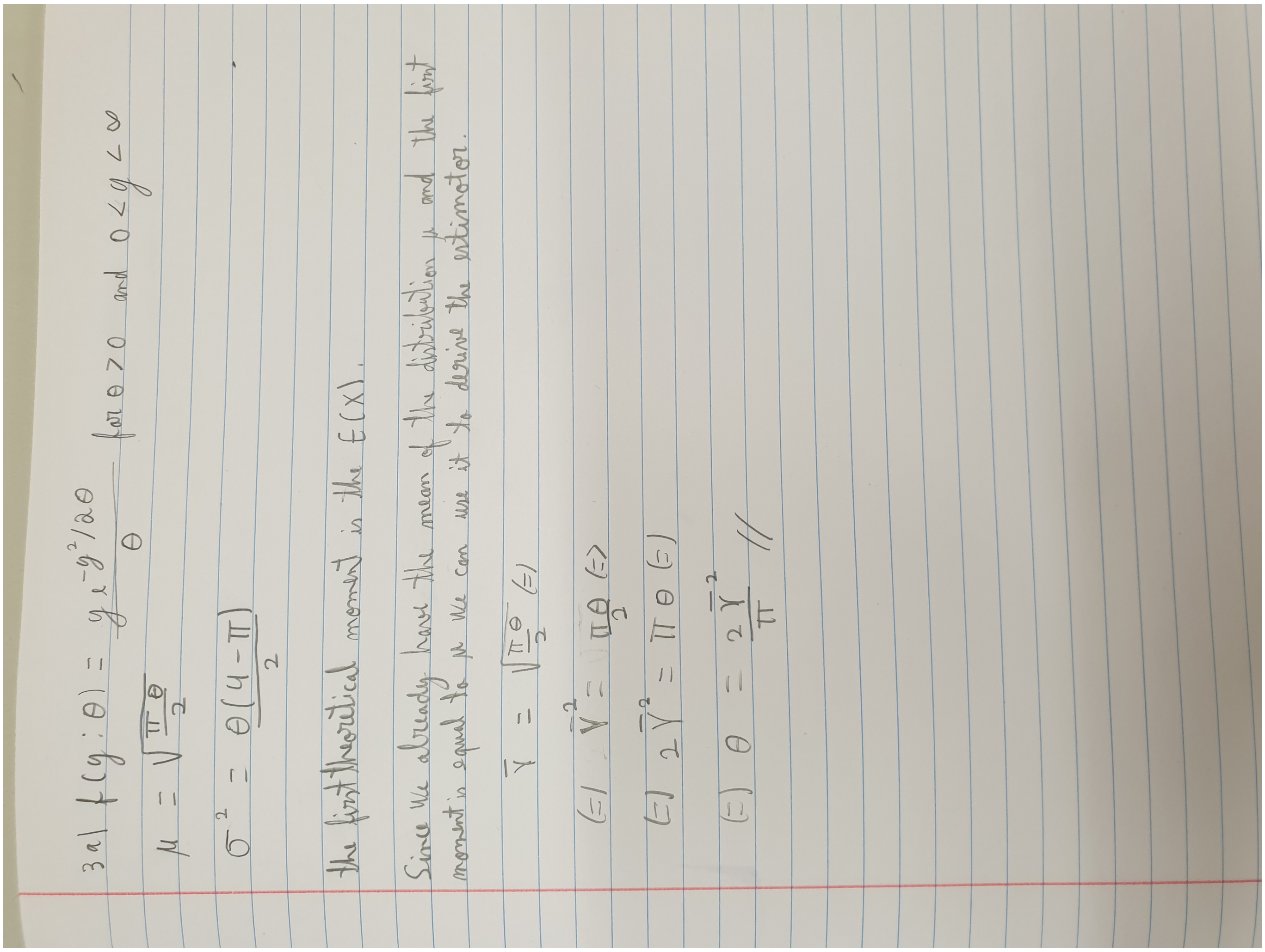
0.6 Question 2 d)

```
# set.seed(26041999)

candidateBWins = 0

## Here we set the number of times we will estimate
largeNumberOfExtimations = 1e+06

for (i in 1:largeNumberOfExtimations) {

    candidateB = 175  ## candidate B initial votes/support
    candidateM = 184  ## candidate M initial votes/support

    ## increment for each day from the initial day to the final day,
    ## the 60th
    for (i in 1:60) {

        ## will be used to hold the new value of votes that continue to
        ## support the same candidate
        candidateBAuxCalc = 0
        candidateMAuxCalc = 0

        ## here we will estimate how many candidates still support the
        ## same candidate by generating a vector where 1 means they
        ## still support the same candidate after the end of each day
        ## the 0 represent them changing the support and vote from one
```

```r
        ## candidate to another
        candidateBAuxCalc = sum(sample(c(1, 0), candidateB, replace = TRUE,
            prob = c(0.996, 0.004)))
        candidateMAuxCalc = sum(sample(c(1, 0), candidateM, replace = TRUE,
            prob = c(0.995, 0.005)))

        ## here we calculate the number of votes that will be exchanged
        ## by the candidates, so the votes that are subtracted from
        ## what they had previously
        votesMovedFromCandidaBToCandidateM = candidateB - candidateBAuxCalc
        votesMovedFromCandidaMToCandidateB = candidateM - candidateMAuxCalc

        ## here calculate that the new current votes for each candidate
        ## by adding the current votes from the mps that still support
        ## the same candidate plus the number of votes of the mps that
        ## exchanged support for their candidate
        candidateB = candidateBAuxCalc + votesMovedFromCandidaMToCandidateB
        candidateM = candidateMAuxCalc + votesMovedFromCandidaBToCandidateM

    }
    ## check if the candidate B did win the election by holding the
    ## majority of the votes
    if (candidateB > candidateM) {
        candidateBWins = candidateBWins + 1

    }

}

## probability estimation of candidate B winning after 14 days of
## campaign by holding the majority of the votes

probabilityB = candidateBWins/largeNumberOfExtimations

probabilityB
```

```
[1] 0.771314
```

As we can see the probability of candidate B winning after 60 days is approximately 0.77


3a) $f(y;\theta) = \frac{y e^{-y^2/2\theta}}{\theta}$ for $\theta > 0$ and $0 < y < \infty$

$\mu = \sqrt{\frac{\pi\theta}{2}}$

$\sigma^2 = \frac{\theta(4-\pi)}{2}$

The first theoretical moment is the $E(X)$.

Since we already have the mean of the distribution μ and the first moment is equal to μ we can use it to derive the estimator.

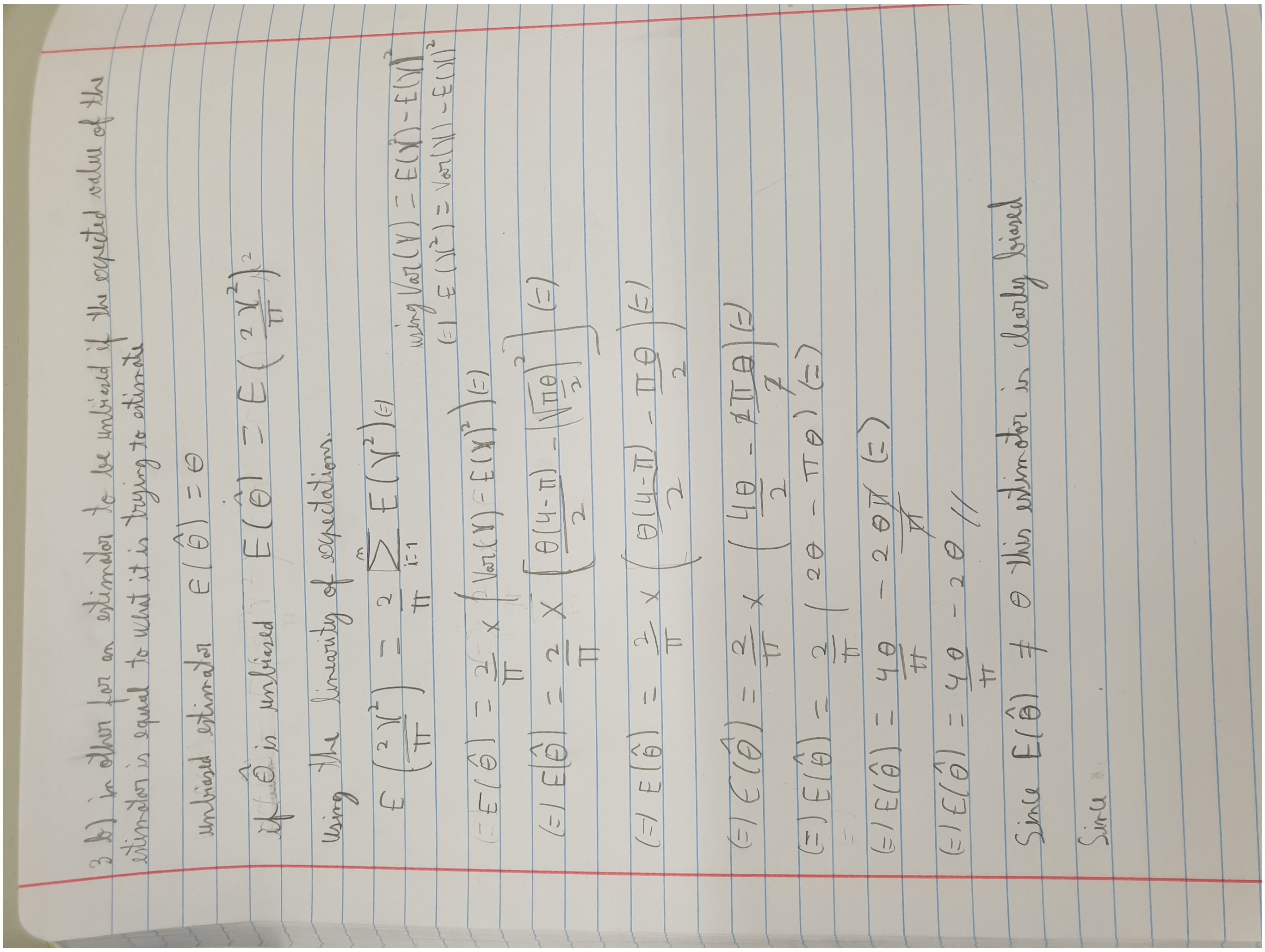
$\bar{Y} = \sqrt{\frac{\pi\theta}{2}} \Leftrightarrow$

$\Leftrightarrow \bar{Y}^2 = \frac{\pi\theta}{2} \Leftrightarrow$

$\Leftrightarrow 2\bar{Y}^2 = \pi\theta \Leftrightarrow$

$\Leftrightarrow \theta = \frac{2\bar{Y}^2}{\pi}$ //

Figure 7: Exercise 3 a)



3 b) in other for an estimator to be unbiased if the expected value of the estimator is equal to what it is trying to estimate

unbiased estimator $E(\hat{\theta}) = \theta$

if $\hat{\theta}$ is unbiased $E(\hat{\theta}) = E\left(\frac{2Y^2}{\pi}\right)^2$

Using the linearity of expectations.

$E\left(\frac{2Y^2}{\pi}\right) = \frac{2}{\pi}\sum_{i=1}^{n} E(Y^2) \Leftrightarrow$

using $Var(Y) = E(Y^2) - E(Y)]^2$

$\Leftrightarrow E(Y^2) = Var(Y) - E(Y)]^2$

$\Leftrightarrow E(\hat{\theta}) = \frac{2}{\pi} \times \left(Var(Y) - E(Y)^2\right) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{2}{\pi} \times \left[\frac{\theta(4-\pi)}{2} - \left(\sqrt{\frac{\pi\theta}{2}}\right)^2\right] \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{2}{\pi} \times \left(\frac{\theta(4-\pi)}{2} - \frac{\pi\theta}{2}\right) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{2}{\pi} \times \left(\frac{4\theta}{2} - \frac{\cancel{2}\pi\theta}{\cancel{2}}\right) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{2}{\pi}\left(2\theta - \pi\theta\right) \Leftrightarrow$

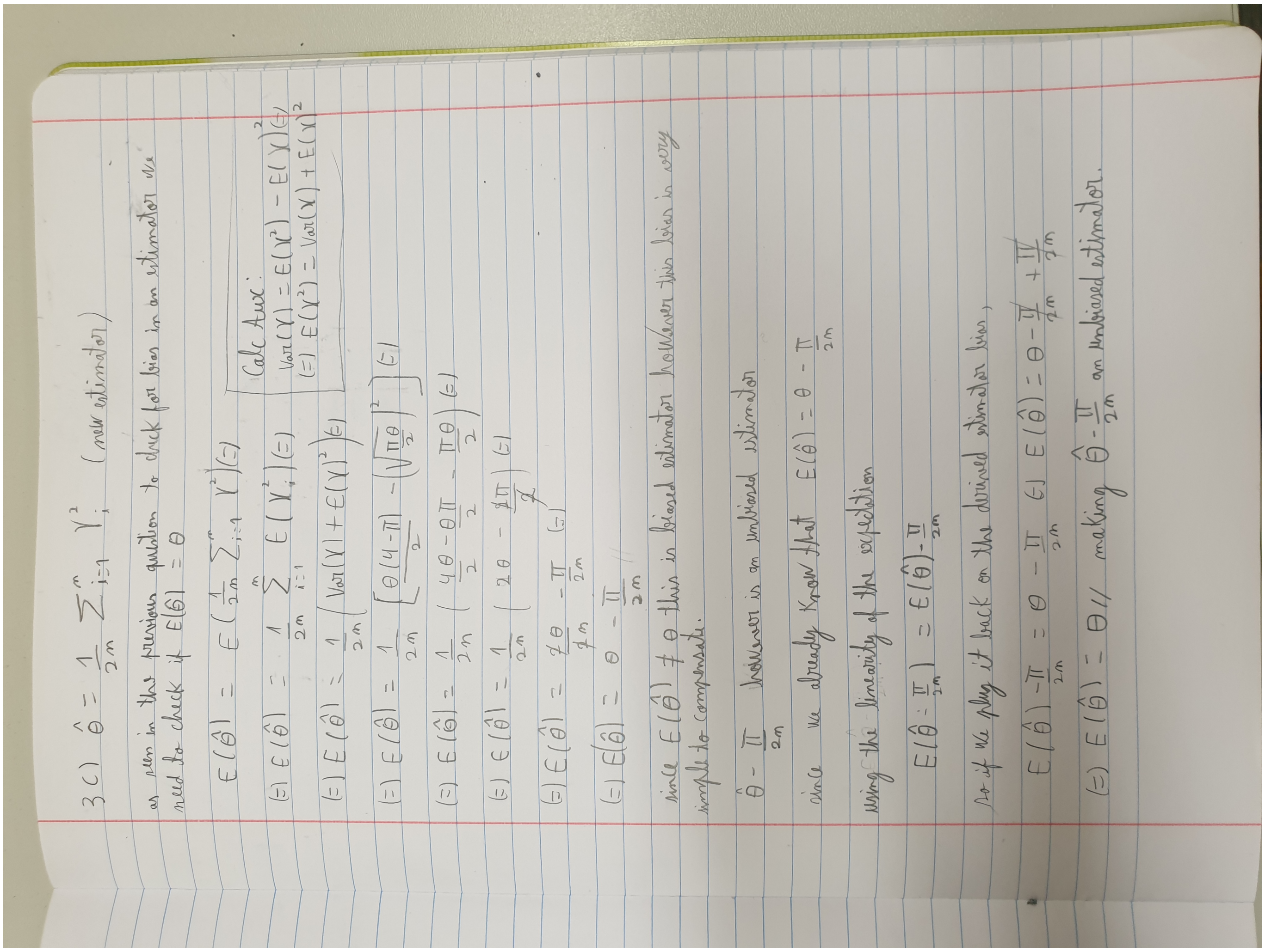
$\Leftrightarrow E(\hat{\theta}) = \frac{4\theta}{\pi} - \frac{2\theta\cancel{\pi}}{\cancel{\pi}} \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{4\theta}{\pi} - 2\theta$ //

Since $E(\hat{\theta}) \neq \theta$ this estimator is clearly biased

Since

Figure 8: Exercise 3 b)



3 c) $\hat{\theta} = \frac{1}{2n} \sum_{i=1}^{n} Y_i^2$ (new estimator)

as seen in the previous question to check for bias in an estimator we need to check if $E(\hat{\theta}) = \theta$

$E(\hat{\theta}) = E\left(\frac{1}{2n} \sum_{i=1}^{n} Y^2\right) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{1}{2n} \sum_{i=1}^{n} E(Y_i^2) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{1}{2n}\left(Var(Y) + E(Y)^2\right) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{1}{2n}\left[\frac{\theta(4-\pi)}{2} - \left(\sqrt{\frac{\pi\theta}{2}}\right)^2\right] \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{1}{2n}\left(\frac{4\theta}{2} - \frac{\theta\pi}{2} - \frac{\pi\theta}{2}\right) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{1}{2n}\left(2\theta - \frac{\cancel{2}\pi}{\cancel{2}}\right) \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \frac{\cancel{2}\theta}{\cancel{2}n} - \frac{\pi}{2n} \Leftrightarrow$

$\Leftrightarrow E(\hat{\theta}) = \theta - \frac{\pi}{2n}$ //

Calc Aux:
$Var(Y) = E(Y^2) - E(Y)^2 \Leftrightarrow$
$\Leftrightarrow E(Y^2) = Var(Y) + E(Y)^2$

since $E(\hat{\theta}) \neq \theta$ this is biased estimator however this bias is very simple to compensate.

$\hat{\theta} - \frac{\pi}{2n}$ however is an unbiased estimator

since we already know that $E(\hat{\theta}) = \theta - \frac{\pi}{2n}$

using the linearity of the expectation

$E(\hat{\theta} - \frac{\pi}{2n}) = E(\hat{\theta}) - \frac{\pi}{2n}$

so if we plug it back on the derived estimator bias,

$E(\hat{\theta}) - \frac{\pi}{2n} = \theta - \frac{\pi}{2n} \Leftrightarrow E(\hat{\theta}) = \theta - \cancel{\frac{\pi}{2n}} + \cancel{\frac{\pi}{2n}}$

$\Leftrightarrow E(\hat{\theta}) = \theta$ // making $\hat{\theta} - \frac{\pi}{2n}$ an unbiased estimator.

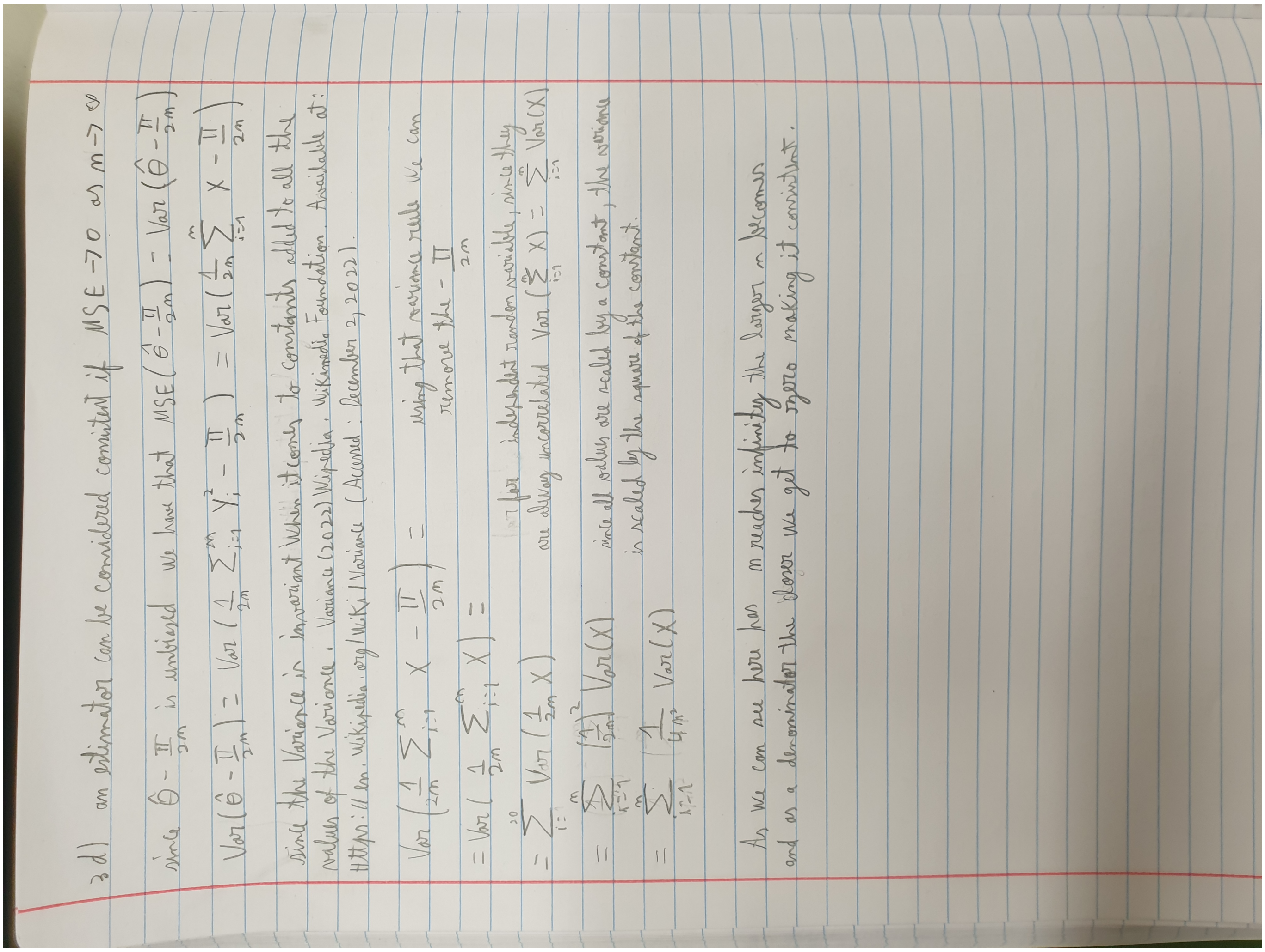
Figure 9: Exercise 3 c)

0.7 Question 3 a)

0.8 Question 3 b)

0.9 Question 3 c)

0.10 Question 3 d)



3 d) an estimator can be considered consistent if $MSE \to 0$ as $m \to \infty$

since $\hat{\theta} - \frac{\pi}{2m}$ is unbiased we have that $MSE(\hat{\theta} - \frac{\pi}{2m}) = Var(\hat{\theta} - \frac{\pi}{2m})$

$$Var(\hat{\theta} - \frac{\pi}{2m}) = Var\left(\frac{1}{2m}\sum_{i=1}^{m} Y_i^2 - \frac{\pi}{2m}\right) = Var\left(\frac{1}{2m}\sum_{i=1}^{m} X - \frac{\pi}{2m}\right)$$

since the Variance is invariant when it comes to constants added to all the values of the Variance. Variance (2022) Wikipedia. Wikimedia Foundation. Available at: Https://en.wikipedia.org/wiki/Variance (Accessed: December 2, 2022).

$$Var\left(\frac{1}{2m}\sum_{i=1}^{m} X - \frac{\pi}{2m}\right) =$$ using that variance rule we can remove the $-\frac{\pi}{2m}$

$$= Var\left(\frac{1}{2m}\sum_{i=1}^{m} X\right) =$$

$$= \sum_{i=1}^{\infty} Var\left(\frac{1}{2m} X\right)$$ for independent random variable, since they are always uncorrelated $Var(\sum_{i=1}^{m} X) = \sum_{i=1}^{m} Var(X)$

$$= \sum_{i=1}^{m} \left(\frac{1}{2m}\right)^2 Var(X)$$ since all values are scaled by a constant, the variance is scaled by the square of the constant.

$$= \sum_{i=1}^{m} \frac{1}{4m^2} Var(X)$$

As we can see here has m reaches infinity the larger m becomes and as a denominator the closer we get to zero making it consistent.

Figure 10: Exercise 3 d)

0.11 Question 4 c)

```
vectorY = c(0.573, 0.77, 0.652, 0.827, 0.821, 0.789, 0.898, 0.718, 0.382,
    0.668, 0.647, 0.477, 0.661, 0.38, 0.87, 0.794, 0.783, 0.732, 0.629, 0.777,
    0.6, 0.724, 0.553, 0.693, 0.687, 0.935, 0.494, 0.411, 0.53, 0.478)

# rt <- polyroot(t)
```

0.12 Question 5 a)

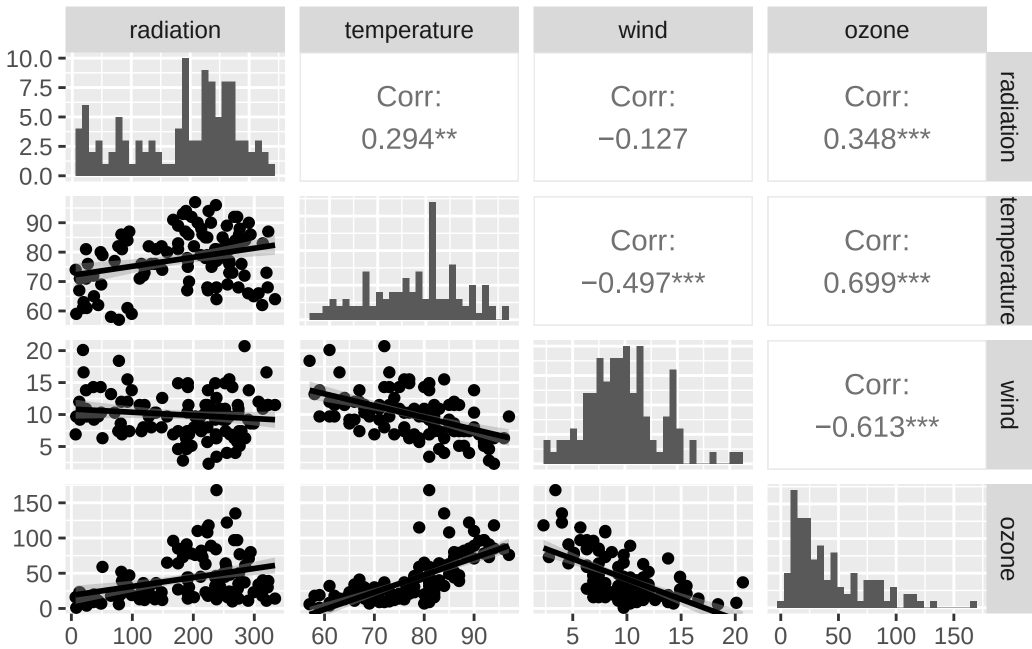
```r
summary(ozone)
```

```
   radiation       temperature          wind             ozone      
 Min.   :  7.0   Min.   :57.00   Min.   : 2.300   Min.   :  1.0  
 1st Qu.:113.5   1st Qu.:71.00   1st Qu.: 7.400   1st Qu.: 18.0  
 Median :207.0   Median :79.00   Median : 9.700   Median : 31.0  
 Mean   :184.8   Mean   :77.79   Mean   : 9.939   Mean   : 42.1  
 3rd Qu.:255.5   3rd Qu.:84.50   3rd Qu.:11.500   3rd Qu.: 62.0  
 Max.   :334.0   Max.   :97.00   Max.   :20.700   Max.   :168.0  
```

Here we can see that wind and ozone have some pretty extremely high max values compared to both the median

```r
ggpairs(ozone, lower = list(continuous = "smooth"), diag = list(continuous = "barDiag"),
    axisLabels = "show")
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Firstly, as it can be observed in the graph that ozone has a very significant positive skewness and is possibly normally distributed. It also noticeable from the ozone histogram that it resembles a normal distribution with positive skewness.

We can also observe that temperature have a positive and strong linear correlation with ozone with only a small amount of variance overall with the exception of a few points between the 3rd quartile and the maximum, we can also see that there is a a positive slope meaning that as temperature increases the amount of ozone detected increases as well.

Furthermore, radiation has a positive correlation with ozone so radiation has a positive effect on ozone. The variance is more extreme between the 2nd quartile and maximum but maintaining a relatively low variance between the minimum and the 2nd quartile.

Lastly, Wind's has a negative correlation with ozone, meaning that has wind increases the less ozone is detected. Most of the variance below the line of best fit, is between the first and third quartile while the values that are more on the extreme, between the the minimum and 1st quartile and the 3rd quartile and the maximum are almost all above the line of best fit. The wind histogram also displays what looks to be a normal distribution with close to zero skewness with some irregularities near the 15 bin.

##question 5 b)

```r
model = lm(ozone ~ radiation + temperature + wind, data = ozone)

summary(model)
```

```
Call:
lm(formula = ozone ~ radiation + temperature + wind, data = ozone)

Residuals:
    Min      1Q  Median      3Q     Max 
-40.485 -14.210  -3.556  10.124  95.600 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -64.23208   23.04204  -2.788  0.00628 ** 
radiation     0.05980    0.02318   2.580  0.01124 *  
temperature   1.65121    0.25341   6.516 2.43e-09 ***
wind         -3.33760    0.65384  -5.105 1.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.17 on 107 degrees of freedom
Multiple R-squared:  0.6062,    Adjusted R-squared:  0.5952 
F-statistic: 54.91 on 3 and 107 DF,  p-value: < 2.2e-16
```

```r
## we can observe that the intercept so when \t

## residuals are the values of the differences between the line we made
## and the observations

## the coefficients are the point estimations intercept is the beta0
## and the wt is the beta1
```

First thing we can observe is the confidence intervals of the 3 variables, both the temperature and wind have confidence intervals of 99,9% as it can be see by the 3 stars next to their respective p-values, radiation is in the 95% confidence interval but is close to the 99% confidence interval, meaning all 3 variables have a significant association and and are a meaningful addiction to our model. Significance being p-values < 0.05.

We can also see by the value of the R-squared and adjusted R-squared that this model around 60% of the variation in ozone levels.

From the estimates we can than wind is the variable with the biggest impact per unit however when comparing it is also important to compare using the minimum and maximum values so we can determine how much each of the independent variables have been recorded to affect the ozone readings so we will be using the minimum and maximum to determine the maximum and minimum variance.

About the Coefficients, we can observe that radiation is the least impactful of the 3 independent variables, the changes in ozone radiation detected vary between [0.4186,19.9732] from the minimum and maximum values, which compared to the [94.11897,160.16737] minimum and maximum variance from the temperature readings which is by far the most impactful variable or the [-7.67648,-66.752] variance from the minimum and maximum values from the wind readings.

Are these findings consistent with your earlier descriptive plots? Also include suitable residual plots, commenting as appropriate.

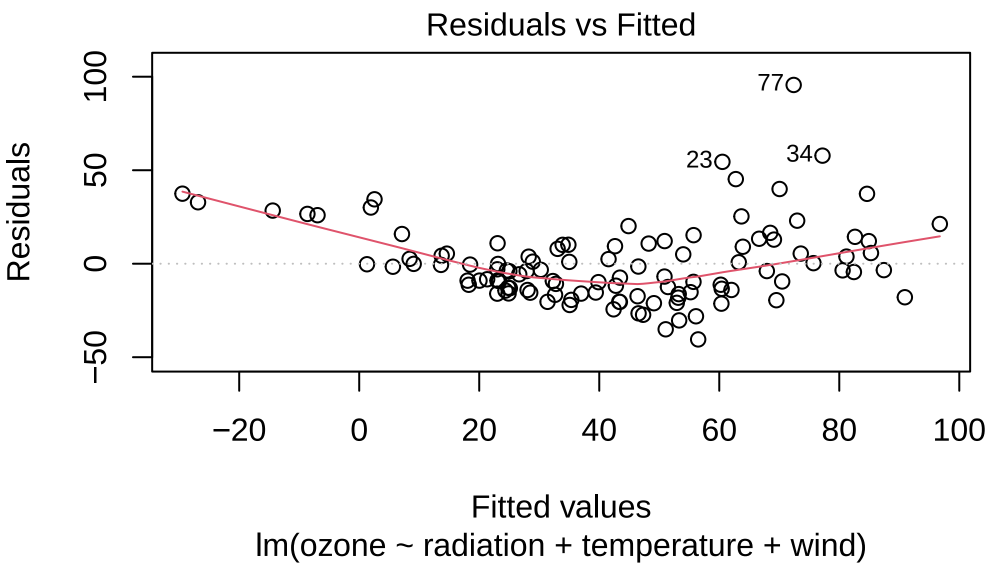
0.13 the intercept value is impossible so lets grpah what we have and have a look

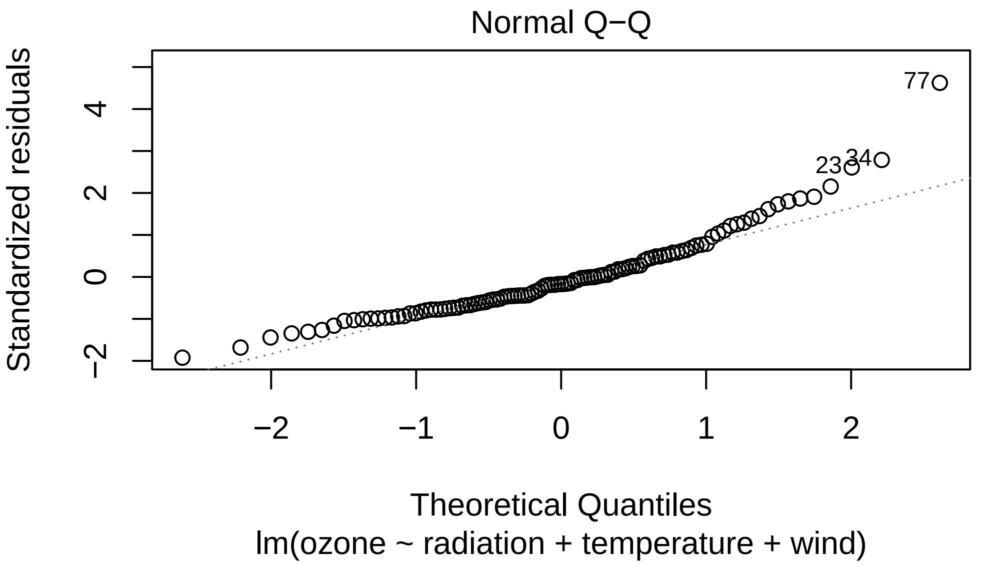
Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) -64.23208 23.04204 -2.788 0.00628 ** radiation 0.05980 0.02318 2.580 0.01124 *
temperature 1.65121 0.25341 6.516 2.43e-09 ***wind -3.33760 0.65384 -5.105 1.45e-06***

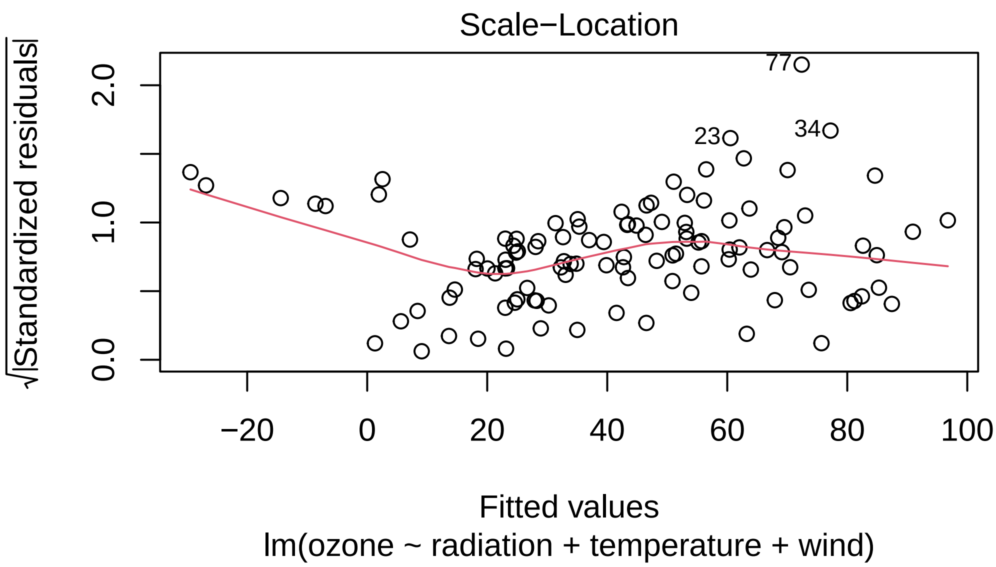
radiation temperature wind ozone
Min. : 7.0 Min. :57.00 Min. : 2.300 Min. : 1.0
1st Qu.:113.5 1st Qu.:71.00 1st Qu.: 7.400 1st Qu.: 18.0
Median :207.0 Median :79.00 Median : 9.700 Median : 31.0
Mean :184.8 Mean :77.79 Mean : 9.939 Mean : 42.1
3rd Qu.:255.5 3rd Qu.:84.50 3rd Qu.:11.500 3rd Qu.: 62.0
Max. :334.0 Max. :97.00 Max. :20.700 Max. :168.0

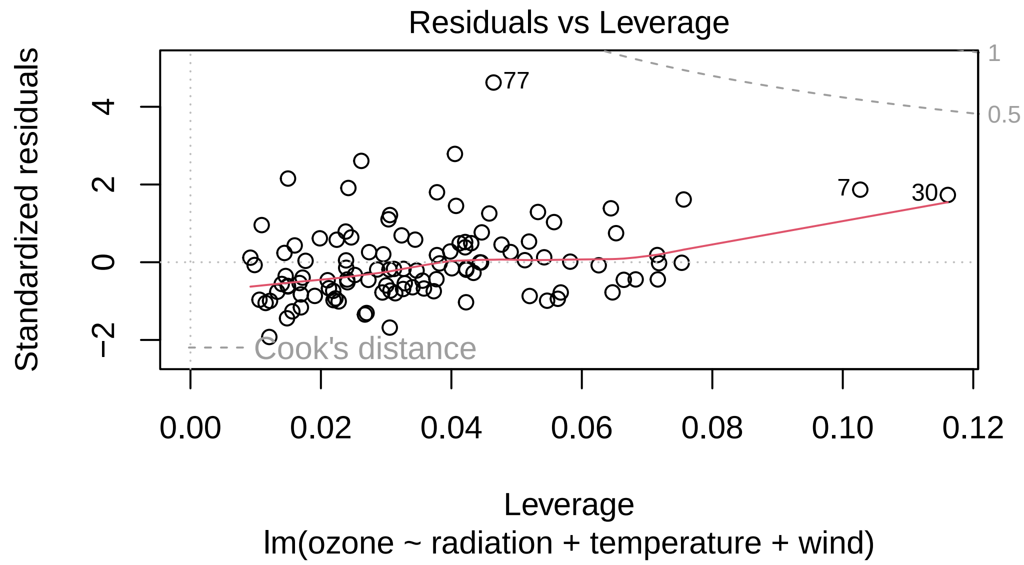
$\beta 0 = \beta 1$

```
plot(model)
```


Residuals vs Fitted
Residuals
Fitted values
lm(ozone ~ radiation + temperature + wind)
77
23
34
100
50
0
−50
−20
0
20
40
60
80
100


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(ozone ~ radiation + temperature + wind)
77
23
34
4
2
0
−2
−2
−1
0
1
2






```
## a residual is the distance between our data points and our
## regression line
```

As observed here on Residuals vs Fitted graph

##question 5 c)

```r
## log(ozone) =  0+ 1 log(radiation)+ 2 log(temperature)+ 3
## log(wind)+ i where  i   N (0,  2)  0 = intercept

modelLogFriend = lm(log(ozone) ~ log(radiation) + log(temperature) + log(wind),
    data = ozone)


summary(modelLogFriend)
```

```

Call:
lm(formula = log(ozone) ~ log(radiation) + log(temperature) +
    log(wind), data = ozone)

Residuals:
     Min       1Q   Median       3Q      Max
-1.63961 -0.30073 -0.00097  0.34414  1.11545

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -10.55570    2.08818  -5.055 1.79e-06 ***
log(radiation)     0.30500    0.05868   5.198 9.73e-07 ***
log(temperature)   3.20478    0.46019   6.964 2.79e-10 ***
log(wind)         -0.66305    0.13751  -4.822 4.74e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4907 on 107 degrees of freedom
Multiple R-squared:  0.6876,	Adjusted R-squared:  0.6788
F-statistic: 78.49 on 3 and 107 DF,  p-value: < 2.2e-16
```

```r
summary(model)
```

```

Call:
lm(formula = ozone ~ radiation + temperature + wind, data = ozone)

Residuals:
    Min      1Q  Median      3Q     Max
-40.485 -14.210  -3.556  10.124  95.600

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -64.23208   23.04204  -2.788  0.00628 **
radiation     0.05980    0.02318   2.580  0.01124 *
temperature   1.65121    0.25341   6.516 2.43e-09 ***
wind         -3.33760    0.65384  -5.105 1.45e-06 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.17 on 107 degrees of freedom
Multiple R-squared:  0.6062,    Adjusted R-squared:  0.5952
F-statistic: 54.91 on 3 and 107 DF,  p-value: < 2.2e-16
```