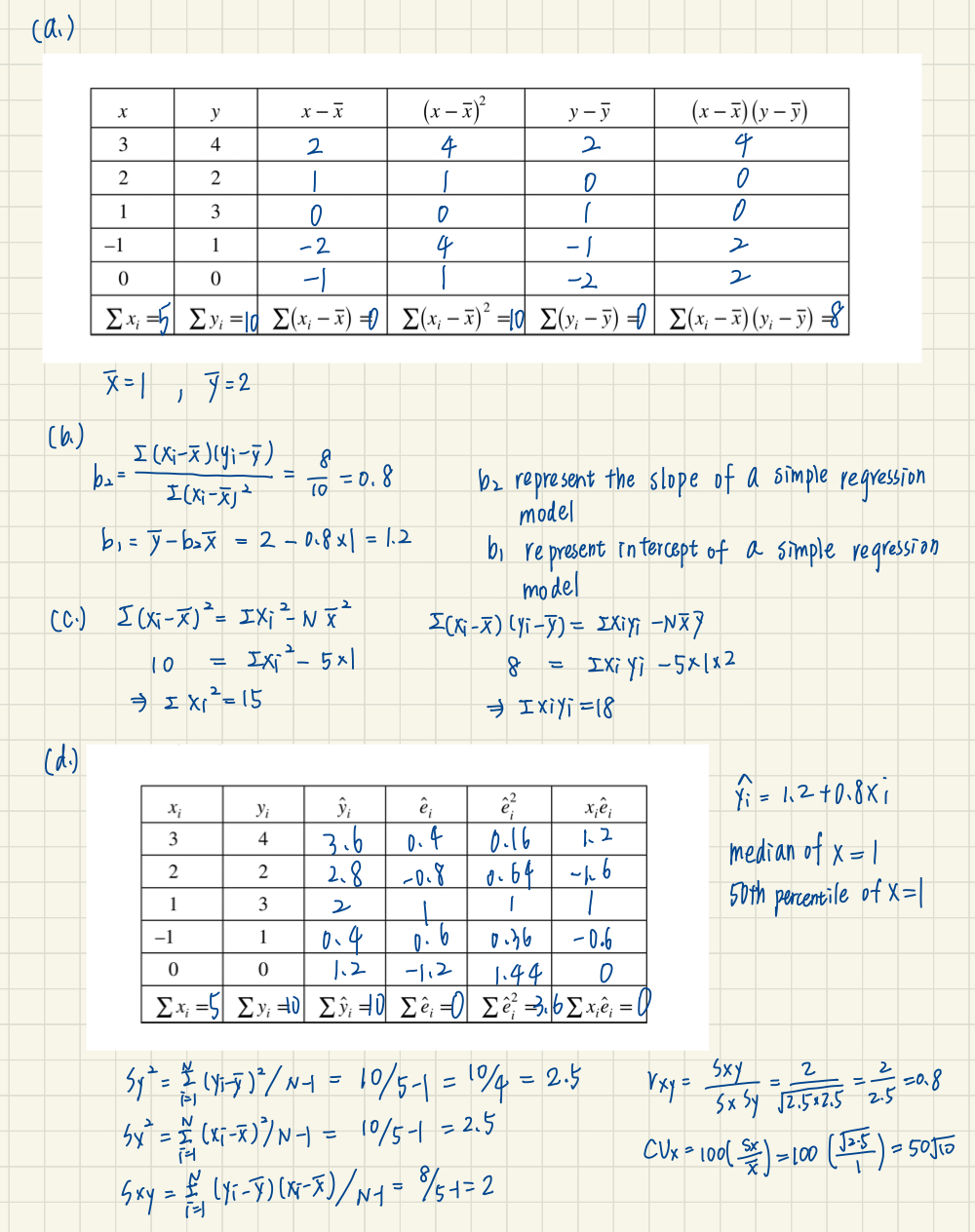
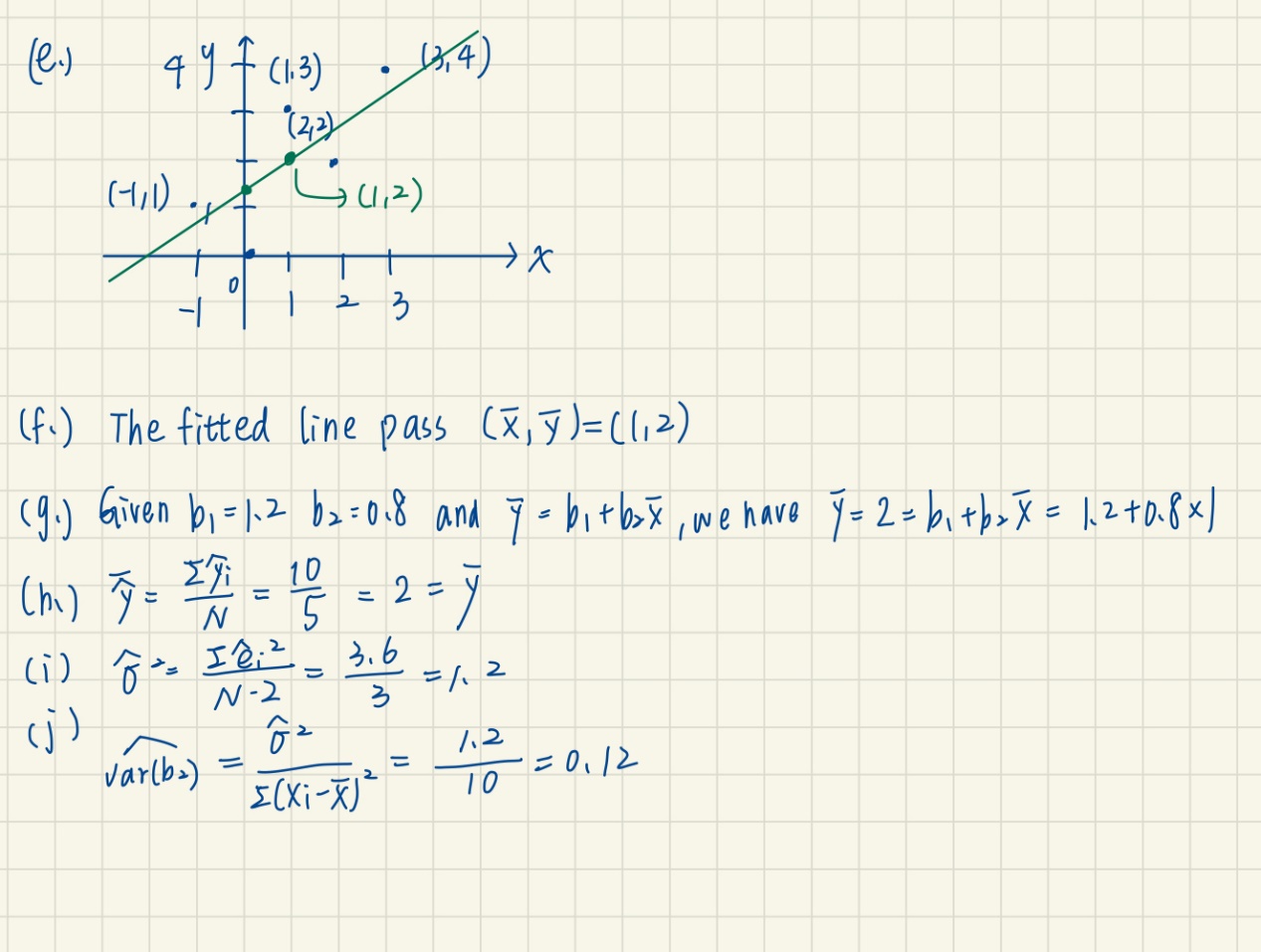
**Econometric Homework 2**

0613404 陸恭葦

**Part 1**

2-1

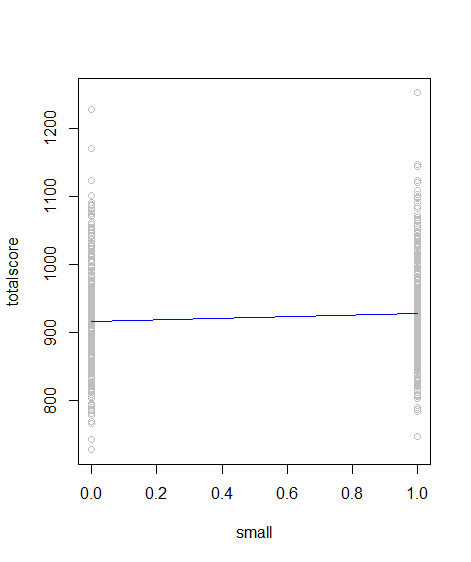




2-22

1. Because there are 425 students belong neither small nor regular class (small=0 & regular=0), we remove them then use a linear regression model to fit the data. The result shows that the R-squared is small and residual’s variance is large (from -189.44 to 324.38). The linear regression model can’t fit the data well. After plotting the scatter plot and fitted line. We find that the data only distribute in where small=0 (regular=1) or small=1 (regular=0), so it doesn’t make sense to use the linear fitted line in this data.

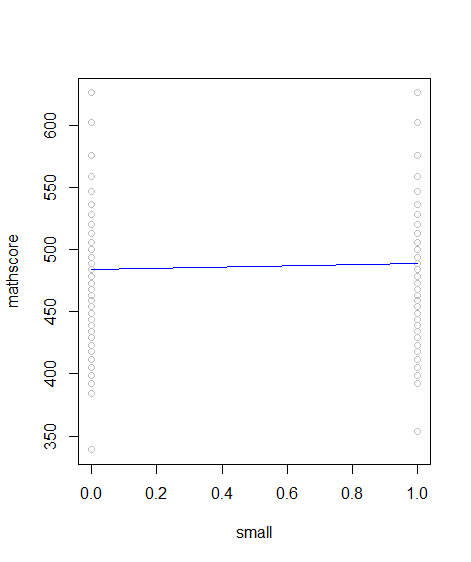
|  |  |
| --- | --- |
|  |  |



|  |
| --- |
| Rcode |
| total <-which(star5\_small$small==1|star5\_small$regular==1)  newstar<-star5\_small[total,]  newstar  mod1 <- lm(totalscore~small,data = newstar)  mod1  summary(mod1)  plot(newstar$small, newstar$totalscore,xlab = 'small',ylab = 'totalscore', col="grey")  lines(fitted(mod1)~newstar$small, col="blue") |

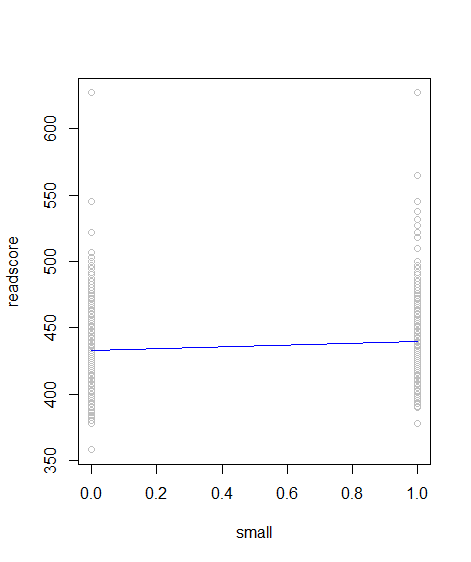
1. The result of the linear regression model shows the similar result. The R-squared is small and residual’s variance is large. Because the distribution of SMALL is the same, the fitted line can’t fit the data well.

|  |  |
| --- | --- |
|  |  |



The result of the linear regression model shows the similar result. The R-squared is small and residual’s variance is large. Because the distribution of SMALL is the same, the fitted line can’t fit the data well.

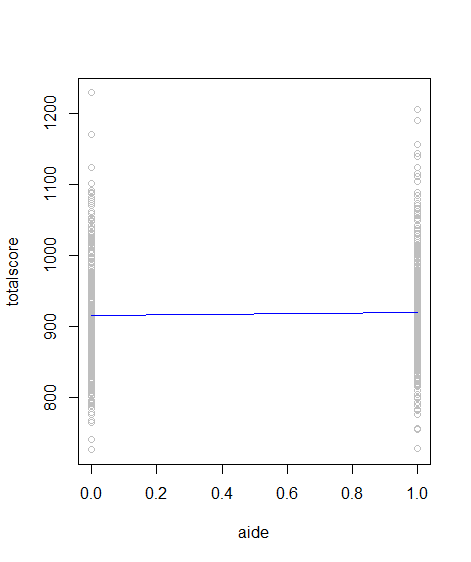
|  |  |
| --- | --- |
|  |  |



|  |
| --- |
| Rcode |
| mod2 <- lm(mathscore~small,data = newstar)  mod2  summary(mod2)  plot(newstar$small, newstar$mathscore,xlab = 'small',ylab = 'mathscore', col="grey")  lines(fitted(mod2)~newstar$small, col="blue")  mod3 <- lm(readscore~small,data = newstar)  mod3  summary(mod3)  plot(newstar$small, newstar$readscore,xlab = 'small',ylab = 'readscore', col="grey")  lines(fitted(mod3)~newstar$small, col="blue") |

1. We use a linear regression model to fit the data. The result shows that the R-squared is small and residual’s variance is large. The linear regression model can’t fit the data well. After plotting the scatter plot and fitted line. We find that the data only distribute in where aide=0 (regular=1) or aside=1 (regular=0), so it doesn’t make sense to use the linear fitted line in this data.

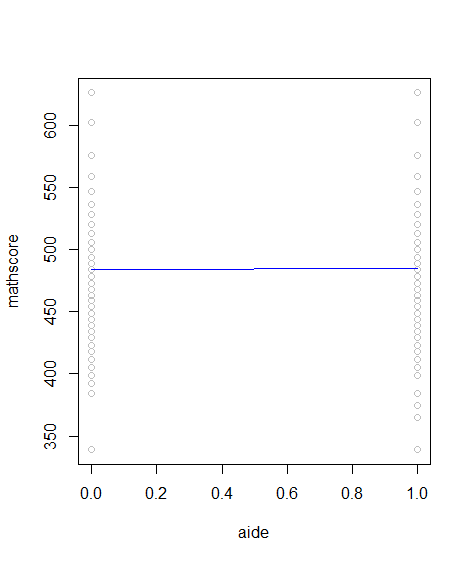
|  |  |
| --- | --- |
|  |  |



|  |
| --- |
| Rcode |
| total2 <-which(star5\_small$aide==1|star5\_small$regular==1)  newstar2<-star5\_small[total2,]  mod4 <- lm(totalscore~aide,data =newstar2)  mod4  summary(mod4)  plot(newstar2$aide, newstar2$totalscore,xlab = 'aide',ylab = 'totalscore', col="grey")  lines(fitted(mod4)~newstar2$aide, col="blue") |

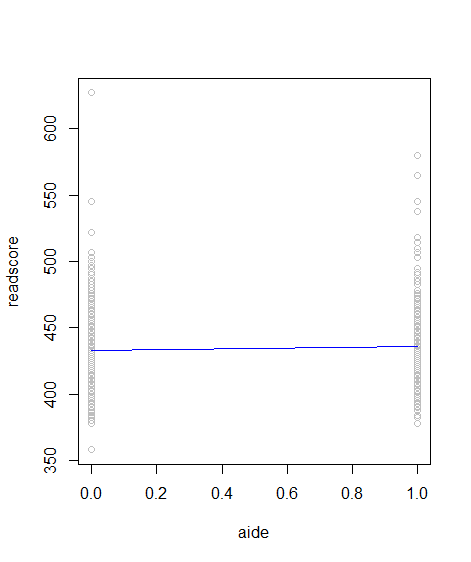
1. The result of the linear regression model shows the similar result. The R-squared is small and residual’s variance is large. Because the distribution of SMALL is the same, the fitted line can’t fit the data well.

|  |  |
| --- | --- |
|  |  |



The result of the linear regression model shows the similar result. The R-squared is small and residual’s variance is large. Because the distribution of SMALL is the same, the fitted line can’t fit the data well.

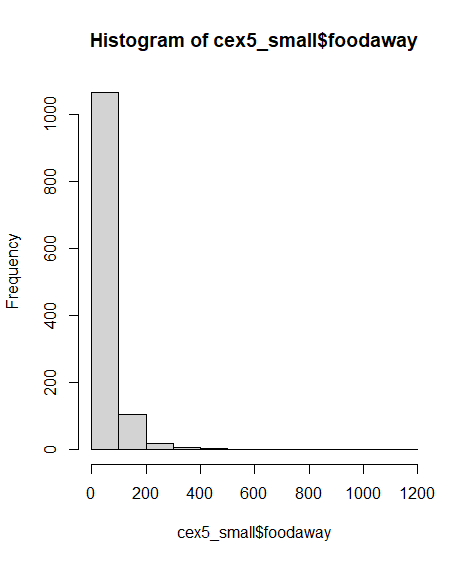
|  |  |
| --- | --- |
|  |  |



|  |
| --- |
| Rcode |
| mod5 <- lm(mathscore~aide,data =newstar2)  mod5  summary(mod5)  plot(newstar2$aide, newstar2$mathscore,xlab = 'aide',ylab = 'mathscore', col="grey")  lines(fitted(mod5)~newstar2$aide, col="blue")  mod6 <- lm(readscore~aide,data =newstar2)  mod6  summary(mod6)  plot(newstar2$aide, newstar2$readscore,xlab = 'aide',ylab = 'readscore', col="grey")  lines(fitted(mod6)~newstar2$aide, col="blue") |

2-25

1. The histogram of foodaway in data called cex5\_small



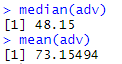
And statistic summary:

The mean is 49.27 and median is 32.55. The 25th percentiles is 12.04 and 75th percentiles is 67.50



|  |
| --- |
| Rcode |
| hist(cex5\_small$foodaway)  summary(cex5\_small$foodaway) |

1. The mean and median values of *FOODAWAY* for households including a member with an advanced degree is 48.15 and 73.15494



The mean and median values of *FOODAWAY* for households including a member with an college degree is 36.11 and 48.59718

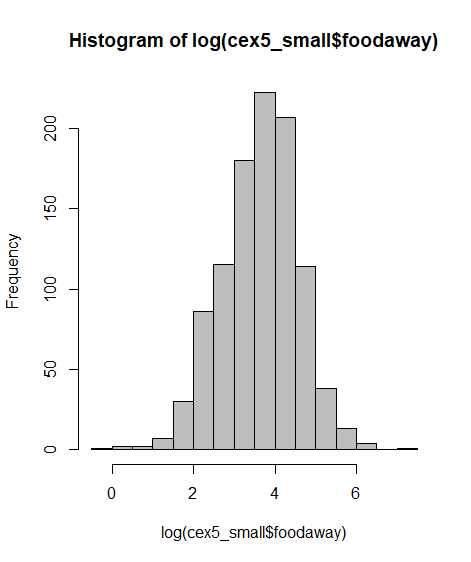


The mean and median values of *FOODAWAY* for households including a member with an advanced or college degree is 38.52 and 58.6792



|  |
| --- |
| Rcode |
| adv = cex5\_small$foodaway[cex5\_small$advanced==1]  median(adv)  mean(adv)  col = cex5\_small$foodaway[cex5\_small$college==1]  median(col)  mean(col)  adco=cex5\_small$foodaway[cex5\_small$advanced==1|cex5\_small$college==1]  median(adco)  mean(adco) |

1. The histogram of ln(*FOODAWAY*) is:



And the statistic summary is :



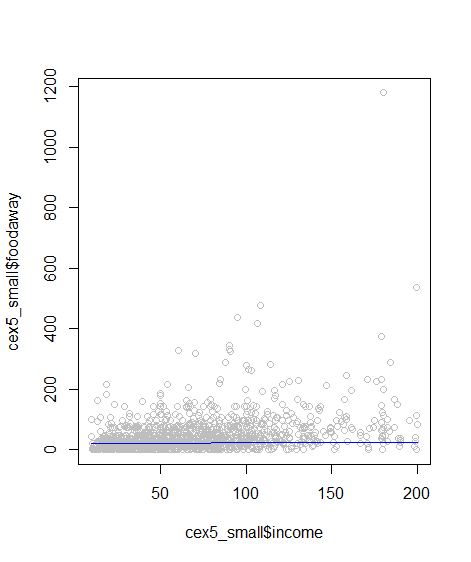
The difference between *FOODAWAY* and ln(*FOODAWAY*) is the distribution of the former is skew to right and the letter is more like a normal and have some negative values.

|  |
| --- |
| Rcode |
| hist(log(cex5\_small$foodaway), col='grey')  summary(log(cex5\_small$foodaway)) |

1. The result of the linear regression model shows that the is 0.0005375. We estimate that if the income goes up by $100, expected ln(*FOODAWAY*) (food away from home expenditure per month per person past quarter) will increase approximately by $0.0005375.

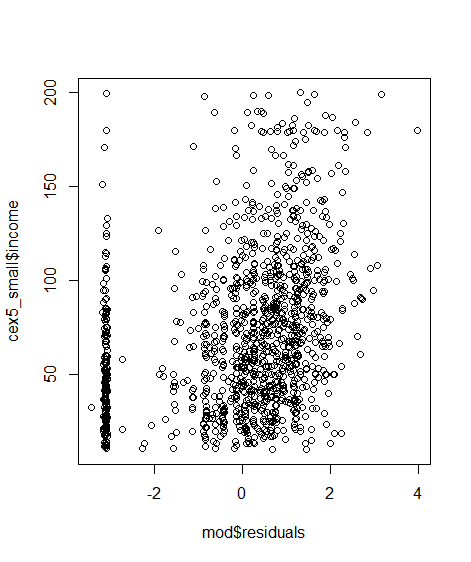


|  |
| --- |
| Rcode |
| ordat <- cex5\_small[order(cex5\_small$income), ]  log\_foodaway<-log(cex5\_small$foodaway)  log\_foodaway  sum(is.infinite(log\_foodaway))  log\_foodaway[is.infinite(log\_foodaway)]<-0  sum(is.infinite(log\_foodaway))  mod <- lm(log\_foodaway~income,data = ordat)  mod |

1. 

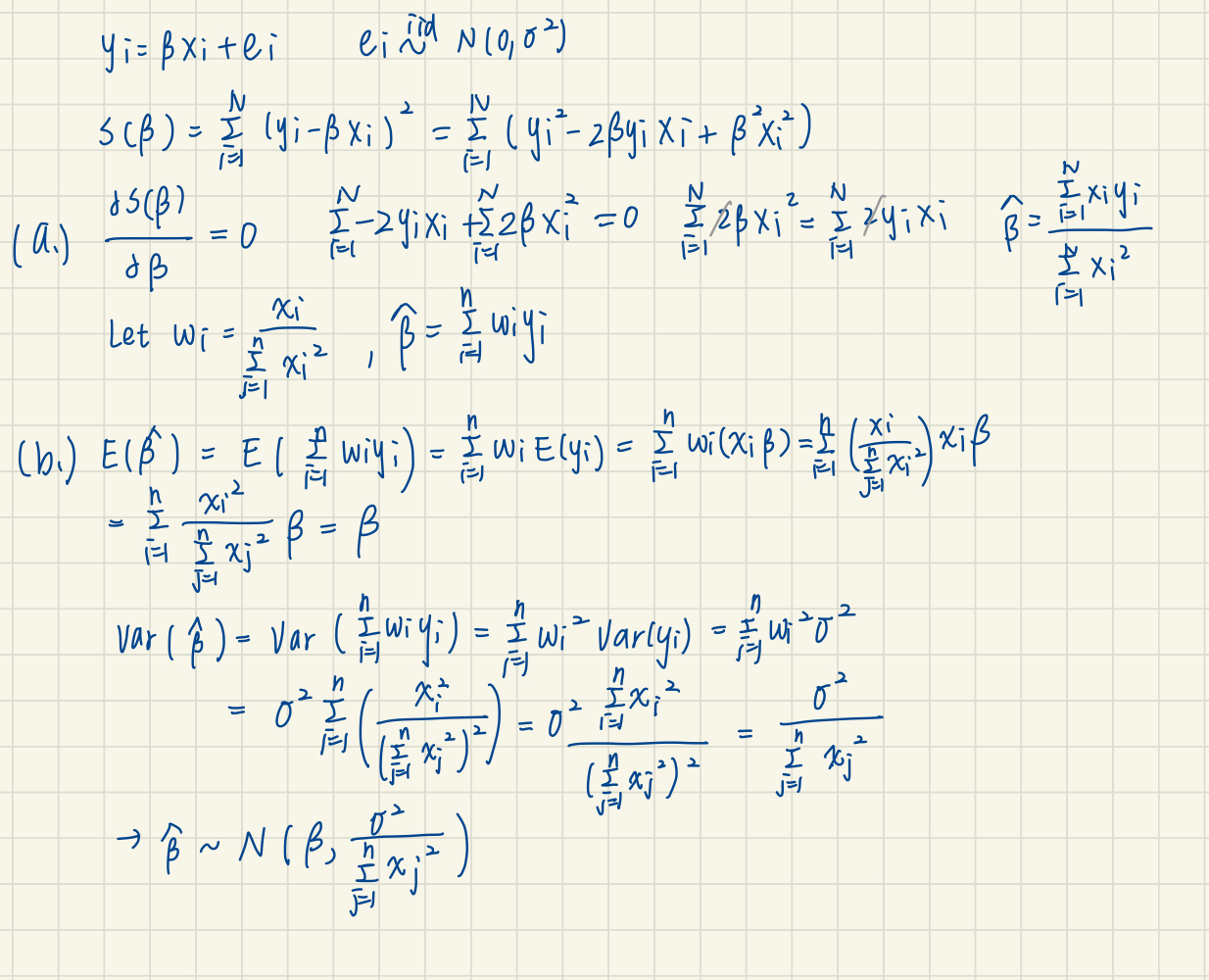
|  |
| --- |
| Rcode |
| plot(cex5\_small$income, log(cex5\_small$foodaway), col="grey")  plot(cex5\_small$income, cex5\_small$foodaway, col="grey")  lines(exp(fitted(mod))~ordat$income, col="blue", main="Log-linear Model") |

1. After calculating the least squares residuals from the estimation. We plot of residuals v.s income. And find that there are some residual is negative (about -3). We think these residual happened because we replace the negative ln(*FOODAWAY*) with 0, so their residual value is not like others that distribution is like normal.



|  |
| --- |
| Rcode |
| attributes(mod)  mod$residuals  plot(mod$residuals,cex5\_small$income) |

**Part 2**

****