

Homework 3

Rolando Santos

2023-10-13

Question 1: In this question, we will use the prostate dataset. Import this dataset and answer following questions.

```
prostate <- read.csv("prostate.csv")
head(prostate)
```

```
##      lcavol lweight age      lbph svi      lcp gleason pgg45      lpsa
## 1 -0.5798185 2.7695 50 -1.386294 0 -1.38629      6      0 -0.43078
## 2 -0.9942523 3.3196 58 -1.386294 0 -1.38629      6      0 -0.16252
## 3 -0.5108256 2.6912 74 -1.386294 0 -1.38629      7     20 -0.16252
## 4 -1.2039728 3.2828 58 -1.386294 0 -1.38629      6      0 -0.16252
## 5  0.7514161 3.4324 62 -1.386294 0 -1.38629      6      0  0.37156
## 6 -1.0498221 3.2288 50 -1.386294 0 -1.38629      6      0  0.76547
```

```
lm.model <- lm(lpsa ~ ., data = prostate)
summary(lm.model)
```

```
##
## Call:
## lm(formula = lpsa ~ ., data = prostate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7331 -0.3713 -0.0170  0.4141  1.6381
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.669337   1.296387   0.516  0.60693
## lcavol       0.587022   0.087920   6.677 2.11e-09 ***
## lweight      0.454467   0.170012   2.673  0.00896 **
## age         -0.019637   0.011173  -1.758  0.08229 .
## lbph         0.107054   0.058449   1.832  0.07040 .
## svi          0.766157   0.244309   3.136  0.00233 **
## lcp         -0.105474   0.091013  -1.159  0.24964
## gleason      0.045142   0.157465   0.287  0.77503
## pgg45        0.004525   0.004421   1.024  0.30886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.7084 on 88 degrees of freedom
## Multiple R-squared: 0.6548, Adjusted R-squared: 0.6234
## F-statistic: 20.86 on 8 and 88 DF, p-value: < 2.2e-16
```

a. Compute a 95% CI for the parameter associated with age. Use the manual method.

```
#95% CI for age
-0.019637 + c(-1,1) * qnorm(0.975) * 0.011173
```

```
## [1] -0.041535678 0.002261678
```

```
confint(lm.model, level = 0.95)
```

```
##              2.5 %      97.5 %
## (Intercept) -1.906960983 3.245634379
## lcavol      0.412298699 0.761744954
## lweight     0.116603435 0.792331414
## age         -0.041840618 0.002566267
## lbph        -0.009101499 0.223209561
## svi         0.280644232 1.251670420
## lcp         -0.286344443 0.075395916
## gleason     -0.267786053 0.358069248
## pgg45       -0.004260932 0.013311395
```

b. Compute a 90% CI for the parameter associated with age. Use the manual method.

```
#90% CI for age
-0.019637 + c(-1,1) * qnorm(0.95) * 0.011173
```

```
## [1] -0.03801495 -0.00125905
```

```
confint(lm.model, level = 0.90)
```

```
##              5 %      95 %
## (Intercept) -1.485718237 2.824391633
## lcavol      0.440867156 0.733176497
## lweight     0.171846568 0.737088281
## age         -0.038210200 -0.001064151
## lbph        0.009890745 0.204217317
## svi         0.360029029 1.172285623
## lcp         -0.256770899 0.045822373
## gleason     -0.216620186 0.306903382
## pgg45       -0.002824333 0.011874796
```

c. Based on these two CIs, what can we expect the p-value of this parameter in t-test? Compare your conclusion with the p-value output by summary function.

```
se <- (0.002261678 - (-0.041535678))/(2 * 1.96)
z <- -0.019637/se
2*pnorm(-abs(z))
```

```
## [1] 0.07882023
```

```
se <- (-0.00125905 - (-0.03801495))/(2 * 1.645)
z <- -0.019637/se
2*pnorm(-abs(z))
```

```
## [1] 0.0787991
```

The p-value we get from either CI is ~0.0788. This is close to the p-value from the summary function 0.08229. This indicates that age is not significant.

d. Conduct a permutation t-test for predictor age in this model.

```
set.seed(408)
T.original <- abs(summary(lm.model)$coef[4,3])

Ts <- c()
for (i in 1:10000){
  lm.model.sample <- lm(lpsa ~ lcavol + lweight + sample(age) +
                        lbph + svi + lcp + gleason + pgg45, data = prostate)
  Ts[i] <- abs(summary(lm.model.sample)$coef[4,3])
}

p.value <- mean(Ts >= T.original)
p.value
```

```
## [1] 0.084
```

Our p-value from our permutation t-test is 0.084 which is close to the one from our summary function 0.08229.

e. Remove all the predictors not significant at the 5% level. Use anova function to conduct an F test to test this model against the original full model. Which model is preferred? Give your reason.

```
lm.model.sig <- lm(lpsa ~ lcavol + lweight + svi, data = prostate)
summary(lm.model.sig)
```

```
##
## Call:
## lm(formula = lpsa ~ lcavol + lweight + svi, data = prostate)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -1.72964 -0.45764  0.02812  0.46403  1.57013
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.26809    0.54350  -0.493  0.62298
## lcavol      0.55164    0.07467   7.388  6.3e-11 ***
## lweight     0.50854    0.15017   3.386  0.00104 **
## svi         0.66616    0.20978   3.176  0.00203 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7168 on 93 degrees of freedom
## Multiple R-squared:  0.6264, Adjusted R-squared:  0.6144
## F-statistic: 51.99 on 3 and 93 DF,  p-value: < 2.2e-16
```

```
anova(lm.model.sig, lm.model)
```

```
## Analysis of Variance Table
##
## Model 1: lpsa ~ lcavol + lweight + svi
## Model 2: lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason +
##          pgg45
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      93 47.785
## 2      88 44.163  5    3.6218 1.4434 0.2167
```

We can see that our p-value is not significant. This means that the more complex model is not any better than our small model, thus the smaller model should be used instead.

Question 2: In this question, we will use the cheddar dataset. Import this dataset and answer following questions.

```
cheddar <- read.csv("cheddar.csv")
head(cheddar)
```

```
##   taste Acetic  H2S Lactic
## 1  12.3  4.543 3.135  0.86
## 2  20.9  5.159 5.043  1.53
## 3  39.0  5.366 5.438  1.57
## 4  47.9  5.759 7.496  1.81
## 5   5.6  4.663 3.807  0.99
## 6  25.9  5.697 7.601  1.09
```

a. Fit a regression model with taste as the response and the three chemical contents as predictors. Identify the predictors that are statistically significant at the 5% level.

```
lm.model <- lm(taste ~ Acetic + H2S + Lactic, data = cheddar)
summary(lm.model)
```

```
##
## Call:
## lm(formula = taste ~ Acetic + H2S + Lactic, data = cheddar)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.390  -6.612  -1.009   4.908  25.449
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -28.8768    19.7354  -1.463  0.15540
## Acetic       0.3277     4.4598   0.073  0.94198
## H2S          3.9118     1.2484   3.133  0.00425 **
## Lactic       19.6705     8.6291   2.280  0.03108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.13 on 26 degrees of freedom
## Multiple R-squared:  0.6518, Adjusted R-squared:  0.6116
## F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

In our model based on the p-values, H2S and Lactic predictors are statistically significant at the 5% level, where H2S ($0.00425 < 0.05$) and Lactic ($0.03108 < 0.05$).

b. Acetic and H2S are measured on a log scale. Fit a linear model where all three predictors are measured on their original scale. Identify the predictors that are statistically significant at the 5% level for this model. Hint: exponential function is `exp()`.

```
exp.model <- lm(taste ~ exp(Acetic) + exp(H2S) + Lactic, data = cheddar)
summary(exp.model)
```

```
##
## Call:
## lm(formula = taste ~ exp(Acetic) + exp(H2S) + Lactic, data = cheddar)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.209  -7.266  -1.651   7.385  26.335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.897e+01  1.127e+01  -1.684  0.1042
## exp(Acetic)  1.891e-02  1.562e-02   1.210  0.2371
## exp(H2S)     7.668e-04  4.188e-04   1.831  0.0786 .
## Lactic       2.501e+01  9.062e+00   2.760  0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.19 on 26 degrees of freedom
## Multiple R-squared:  0.5754, Adjusted R-squared:  0.5264
## F-statistic: 11.75 on 3 and 26 DF,  p-value: 4.746e-05
```

We see now that only Lactic is significant at the 5% level ($0.0105 < 0.05$).

c. Can we use an F-test to compare these two models? Which model provides a better fit to the data? Explain your reasoning for these two questions.

An F-test cannot be used to compare the two models since neither models are a subset of the other. The original linear model is better based on the r-squared value (0.6518) compared to the model with exponents (0.5754).

d. If H2S is increased 0.01 for the model used in (a), what change in the taste would be expected?

```
set.seed(408)
adjust.H2S <- function(sample.index, data = cheddar, lm.model){
  line.break <- paste("\n")
  sample.row <- cheddar[sample.index, ]
  ret <- paste("Acetic value:", sample.row[, 2],
              "H2S Value:", sample.row[, 3],
              "Lactic value:", sample.row[, 4],
              "Taste Value", sample.row[, 1], sep = "\n")

  cat(ret)
  sample.row$taste <- NULL
  sample.row <- as.matrix(cbind(intercept = 1, sample.row))

  beta.hat <- as.matrix(lm.model$coefficients)
  y.original <- (sample.row) %*% beta.hat

  cat(line.break)
  cat(line.break)
  ret <- paste("Response:", y.original, sep = "\n")
  cat(ret)

  sample.row.updated <- cheddar[sample.index, ]
  sample.row.updated$H2S <- sample.row.updated$H2S + 0.01
  sample.row.updated$taste <- NULL
  sample.row.updated <- as.matrix(cbind(intercept = 1, sample.row.updated))

  y.new <- (sample.row.updated) %*% beta.hat

  cat(line.break)
  cat(line.break)
  ret <- paste("New Response:", y.new, sep = "\n")
  cat(ret)

  cat(line.break)
  cat(line.break)
  ret <- paste("Difference:", y.new - y.original, sep = "\n")
  cat(ret)
}
```

```
adjust.H2S(sample.index = sample(nrow(cheddar), 1), lm.model = lm.model)
```

```
## Acetic value:
## 6.043
```

```
## H2S Value:
## 3.219
## Lactic value:
## 1.46
## Taste Value
## 11.6
##
## Response:
## 14.4149807103273
##
## New Response:
## 14.4540991210639
##
## Difference:
## 0.0391184107366485
```

```
adjust.H2S(sample.index = sample(nrow(cheddar), 1), lm.model = lm.model)
```

```
## Acetic value:
## 5.74
## H2S Value:
## 6.142
## Lactic value:
## 1.68
## Taste Value
## 34.9
##
## Response:
## 30.0775060962636
##
## New Response:
## 30.1166245070002
##
## Difference:
## 0.0391184107366485
```

We can see that by changing the H2S value by 0.01, we get a ~0.0391 difference between the responses.

Question 3: In this question, we will use the teengamb dataset. Import this dataset and answer following questions.

```
teengamb <- read.csv("teengamb.csv")
head(teengamb)
```

```
##   sex status income verbal gamble
## 1   1     51   2.00      8    0.0
## 2   1     28   2.50      8    0.0
## 3   1     37   2.00      6    0.0
## 4   1     28   7.00      4    7.3
## 5   1     65   2.00      8   19.6
## 6   1     61   3.47      6    0.1
```

a. Fit a model with gamble as the response and the other variables as predictors. Which variables are statistically significant at the 5% level?

```
lm.model <- lm(gamble ~ ., data = teengamb)
summary(lm.model)

##
## Call:
## lm(formula = gamble ~ ., data = teengamb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -51.082 -11.320  -1.451   9.452  94.252
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.55565   17.19680   1.312   0.1968
## sex          -22.11833    8.21111  -2.694   0.0101 *
## status         0.05223    0.28111   0.186   0.8535
## income         4.96198    1.02539   4.839 1.79e-05 ***
## verbal        -2.95949    2.17215  -1.362   0.1803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.69 on 42 degrees of freedom
## Multiple R-squared:  0.5267, Adjusted R-squared:  0.4816
## F-statistic: 11.69 on 4 and 42 DF,  p-value: 1.815e-06
```

In our model sex is significant at the 5% level and income is significant at the 0.1% level.

b. Check the meaning of each variable. Does the variable significance in (a) make sense? Give your reasoning.

For sex the variables are either 0 (for males) and 1 (for females), this means that the gambling habits differ between both genders. The income level makes sense as well, we could deduce that the gambling habits may differ amongst teenagers that have a flow of income vs. those that do not.

c. Fit a model with just income as a predictor and use an F-test to compare it to the full model.

```
lm.model.income <- lm(gamble ~ income, data = teengamb)
summary(lm.model.income)
```

```
##
## Call:
## lm(formula = gamble ~ income, data = teengamb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```



```
## -46.020 -11.874 -3.757 11.934 107.120
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.325      6.030  -1.049    0.3
## income        5.520      1.036   5.330 3.05e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.95 on 45 degrees of freedom
## Multiple R-squared:  0.387, Adjusted R-squared:  0.3734
## F-statistic: 28.41 on 1 and 45 DF, p-value: 3.045e-06
```

```
anova(lm.model.income, lm.model)
```

```
## Analysis of Variance Table
##
## Model 1: gamble ~ income
## Model 2: gamble ~ sex + status + income + verbal
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      45 28009
## 2      42 21624  3    6384.8 4.1338 0.01177 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see by the significance of the p-value (0.01177) that we do have enough evidence to reject the null hypothesis and that the complex model is a better fit than the model with just income.

Question 4: In this question, we will use the sat dataset. It was collected to study the relationship between expenditures on public education and test results.

Using the sat dataset, fit a linear model with the total SAT score as the response and expend, salary, ratio, and takers as predictors. Perform regression diagnostics on this model to answer the following questions. Display any plots that are relevant. Some questions may be subjective. Show the most valid judgment and give your seasons.

```
sat <- read.csv("sat.csv")
head(sat)
```

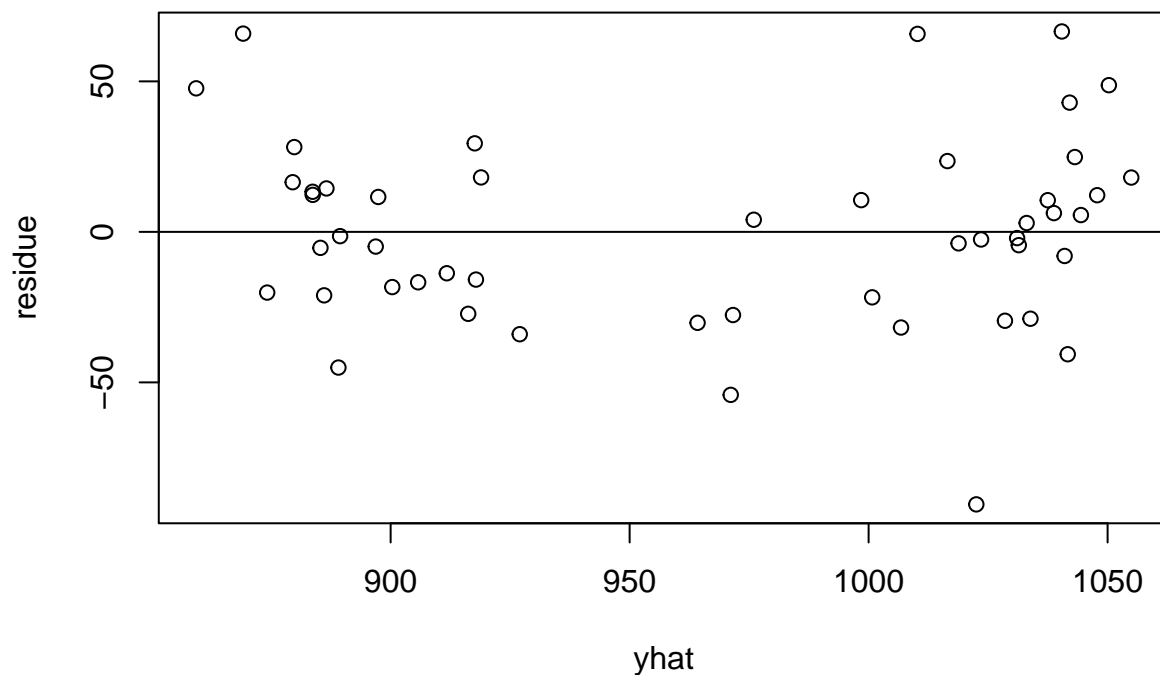
```
##   expend ratio salary takers verbal math total
## 1  4.405  17.2 31.144      8    491  538 1029
## 2  8.963  17.6 47.951     47    445  489  934
## 3  4.778  19.3 32.175     27    448  496  944
## 4  4.459  17.1 28.934      6    482  523 1005
## 5  4.992  24.0 41.078     45    417  485  902
## 6  5.443  18.4 34.571     29    462  518  980
```

```
lm.model <- lm(total ~ expend + salary + ratio + takers, data = sat)
summary(lm.model)
```

```
##
## Call:
## lm(formula = total ~ expend + salary + ratio + takers, data = sat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -90.531 -20.855  -1.746  15.979  66.571
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1045.9715    52.8698   19.784 < 2e-16 ***
## expend         4.4626    10.5465    0.423  0.674
## salary         1.6379     2.3872    0.686  0.496
## ratio        -3.6242     3.2154   -1.127  0.266
## takers        -2.9045     0.2313  -12.559 2.61e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.7 on 45 degrees of freedom
## Multiple R-squared:  0.8246, Adjusted R-squared:  0.809
## F-statistic: 52.88 on 4 and 45 DF,  p-value: < 2.2e-16
```

a. Plot residual vs. fitted response to check the constant variance assumption for the errors.

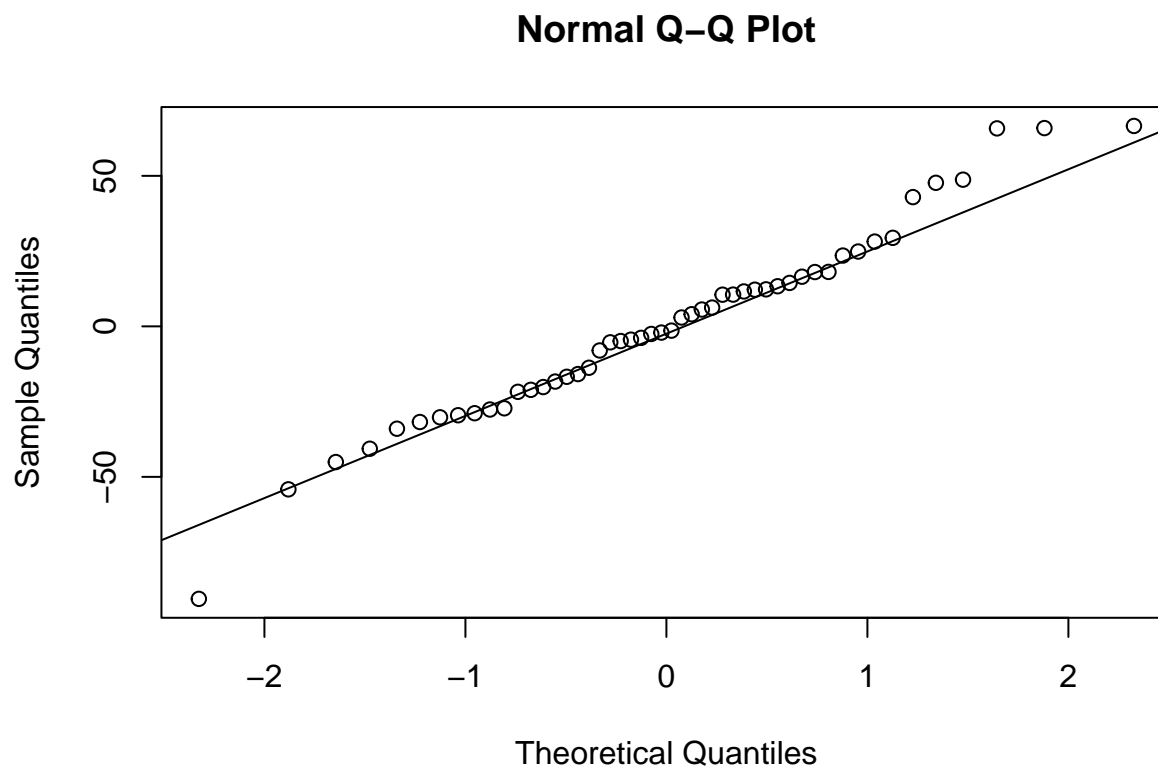
```
plot(lm.model$residuals ~ lm.model$fitted.values, xlab='yhat', ylab='residue')
abline(h=0)
```



The data is scattered and shows non-linearity with the residuals, the data doesn't necessarily show any strong outliers.

b. Use Q-Q plot to check the normality assumption. What is the shape of the error distribution?

```
qqnorm(lm.model$residuals)
qqline(lm.model$residuals)
```



From the Q-Q plot we can see that most of the values lie on the line, we can deduce that the data is normally distributed.

c. Use studentized residuals to check the outliers. Set the cutoff of being “large” as the 5% critical value in t distribution. Note that we need to consider the two sides.

```
n <- dim(sat)[1]
df <- n - dim(sat)[2] - 1
sr <- rstudent(lm.model)
sr
```

```
##          1          2          3          4          5          6
## -0.06574555 -1.02091600 -0.86394293 -0.90546917 -0.56764583  0.12375609
##          7          8          9         10         11         12
##  0.97811087  0.41584816 -0.89619150 -0.64380944 -0.52185887 -0.68138366
##         13         14         15         16         17         18
##  0.34240342 -0.57108388  1.56037417  0.38967283 -0.92044224 -0.08013841
##         19         20         21         22         23         24
##  0.53951189  0.36099300  1.55127336 -0.26522896  1.36779678  0.09173934
##         25         26         27         28         29         30
```

```
## 0.19418490 0.34070604 0.17614652 -1.73200396 2.19000605 -0.47250752
##          31          32          33          34          35          36
## -0.11823952 -0.16415440 -0.67193751 2.21368580 -0.99430251 -0.13785064
##          37          38          39          40          41          42
## 0.94590973 -0.17154457 -0.04457306 -1.46883156 0.79638212 0.75084252
##          43          44          45          46          47          48
## -1.07088136 2.52958734 0.45748671 0.40448538 0.58126580 -3.12442832
##          49          50
## 0.57665563 -1.31188995
```

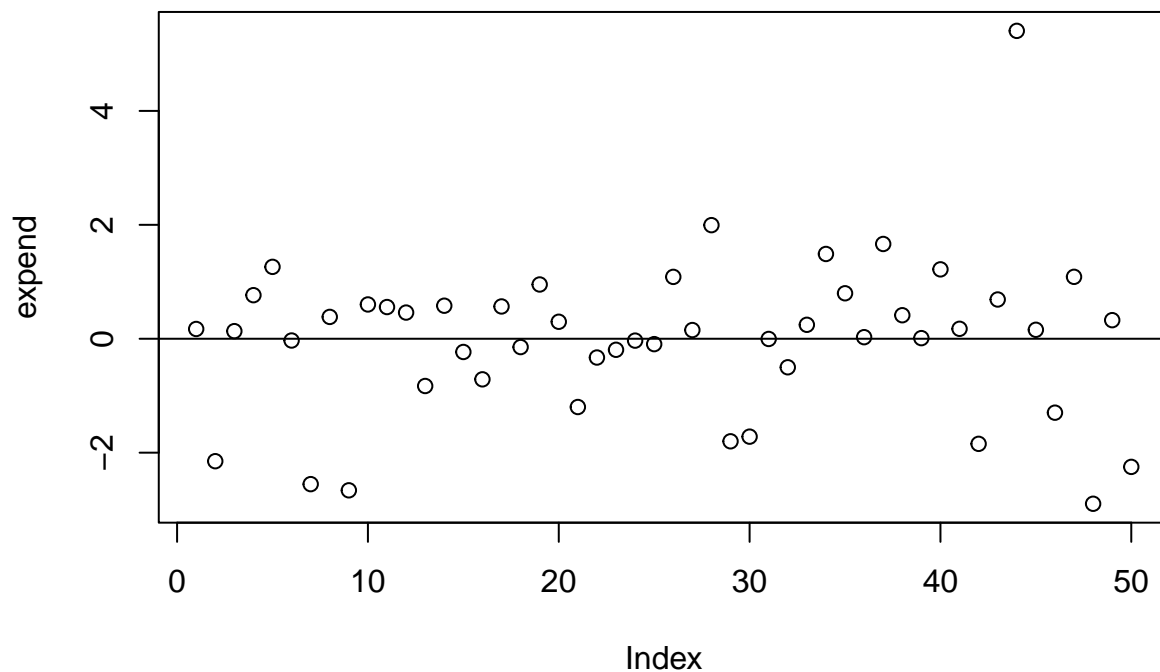
```
sr[which(abs(sr) > qt(0.975, df))]
```

```
##          29          34          44          48
## 2.190006 2.213686 2.529587 -3.124428
```

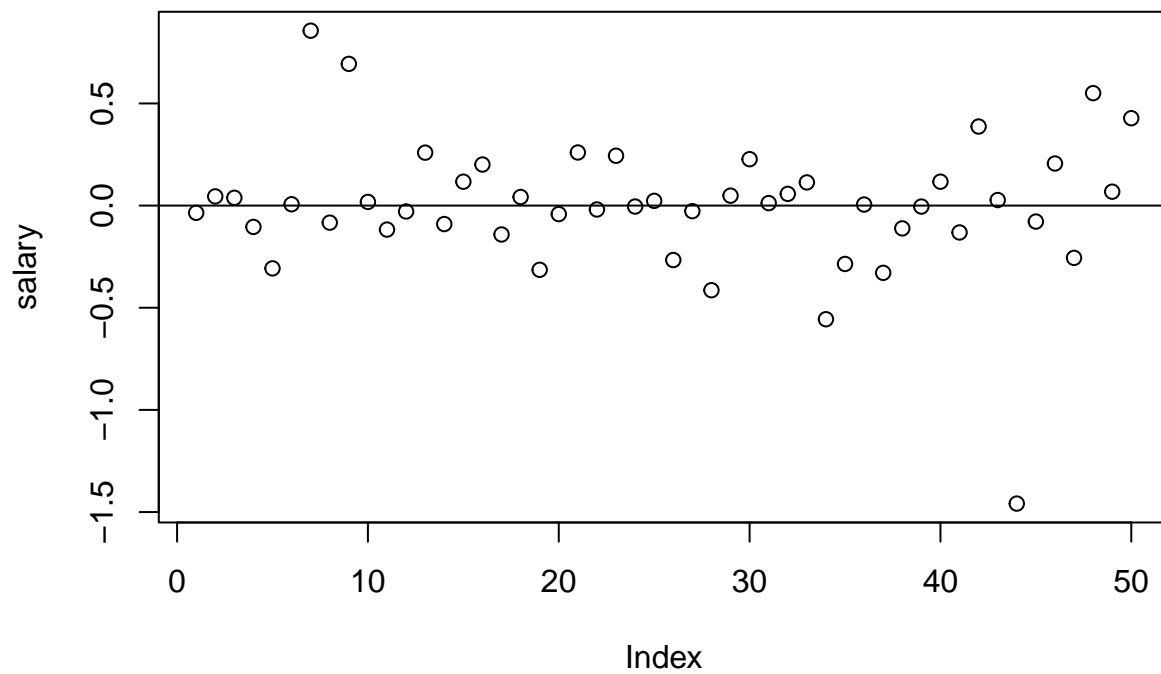
We can see that we have 4 outliers, if we take a glance at all the studentized residuals we can see that it makes sense that these 4 outliers are farthest apart from the average data point.

d. Using `dfbeta` function to plot the change of parameter estimation and check influential points. Check influential points in terms of each parameter except the intercept.

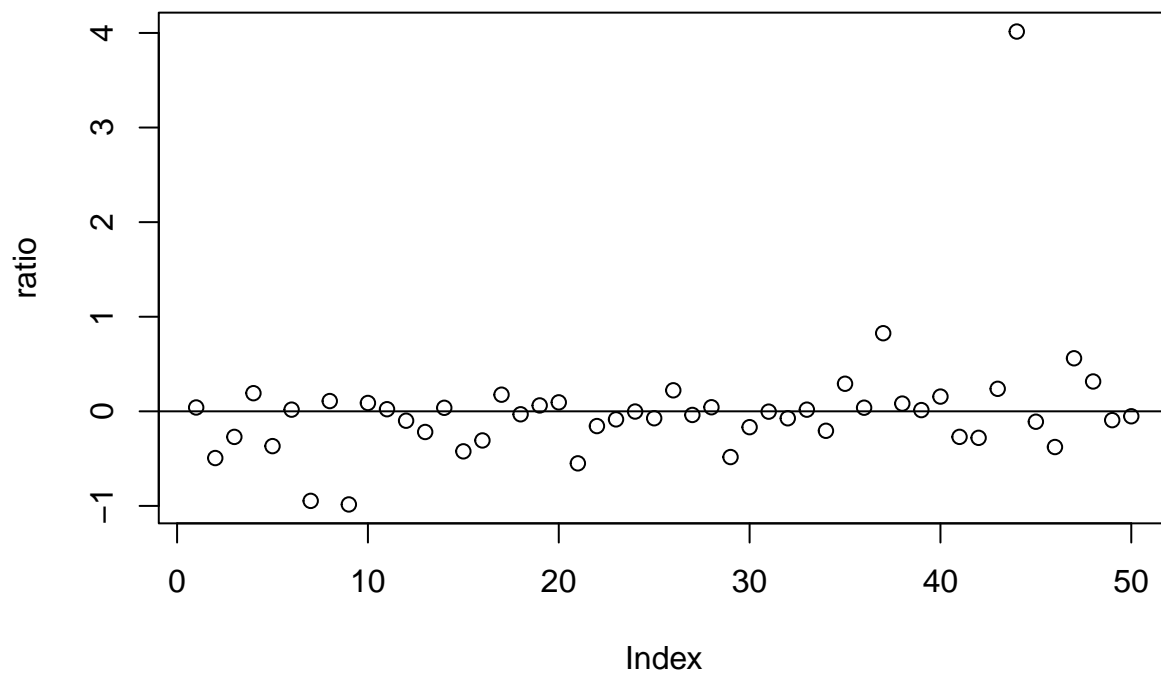
```
beta.change <- dfbeta(lm.model)
plot(beta.change[,2], ylab = "expend")
abline(h=0)
```



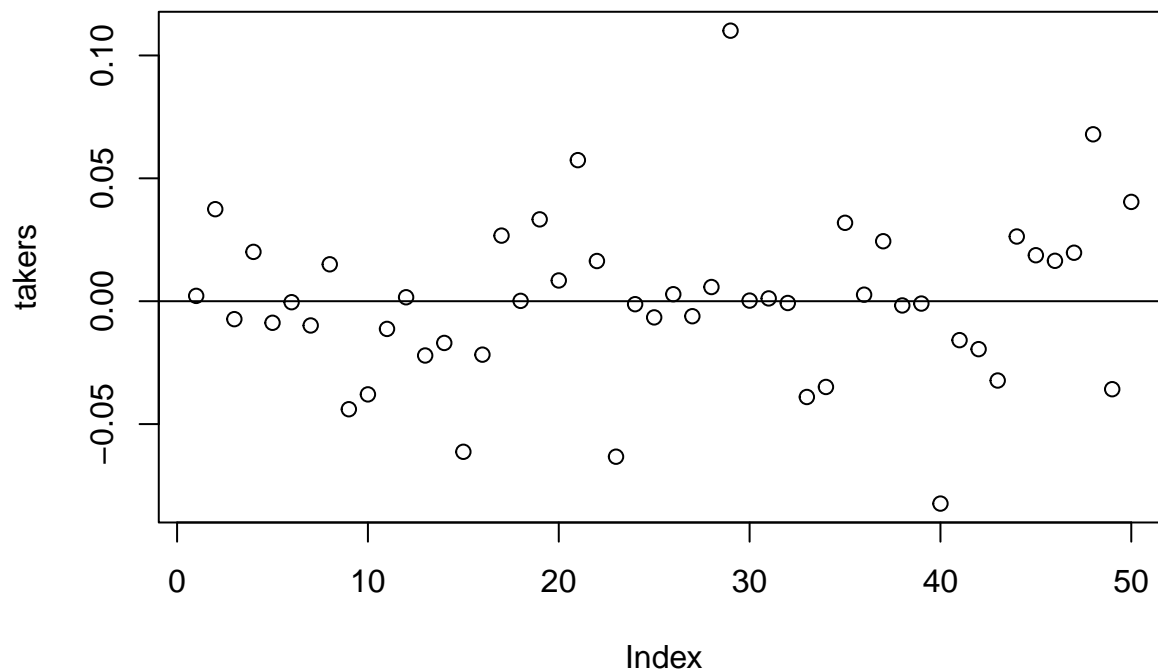
```
plot(beta.change[,3], ylab = "salary")
abline(h=0)
```



```
plot(beta.change[,4], ylab = "ratio")
abline(h=0)
```



```
plot(beta.change[,5], ylab = "takers")
abline(h=0)
```



We can see in the ratio, salary, and expend plots we have a single point that has a large amount of influence, in the takers plot we have multiple but they are not too far off from the line to be as influential as the ones in the other plots.

Question 5: The divusa dataset records the US divorce and social-economic variables in 77 years.

Fit a model with divorce as the response and the other variables, except year, as predictors. Check for error correlation using three methods: plot residuals vs. years; plot residual (t+1) vs. residual (t); fit a linear model of residual (t+1) ~ residual (t). Give your insight and conclusion.

```
divusa <- read.csv("divusa.csv")
head(divusa)
```

```
##   year divorce unemployed femlab marriage birth military
## 1 1920      8.0         5.2  22.70     92.0 117.9   3.2247
## 2 1921      7.2        11.7  22.79     83.0 119.8   3.5614
## 3 1922      6.6         6.7  22.88     79.7 111.2   2.4553
## 4 1923      7.1         2.4  22.97     85.2 110.5   2.2065
## 5 1924      7.2         5.0  23.06     80.3 110.9   2.2889
## 6 1925      7.2         3.2  23.15     79.2 106.6   2.1735
```

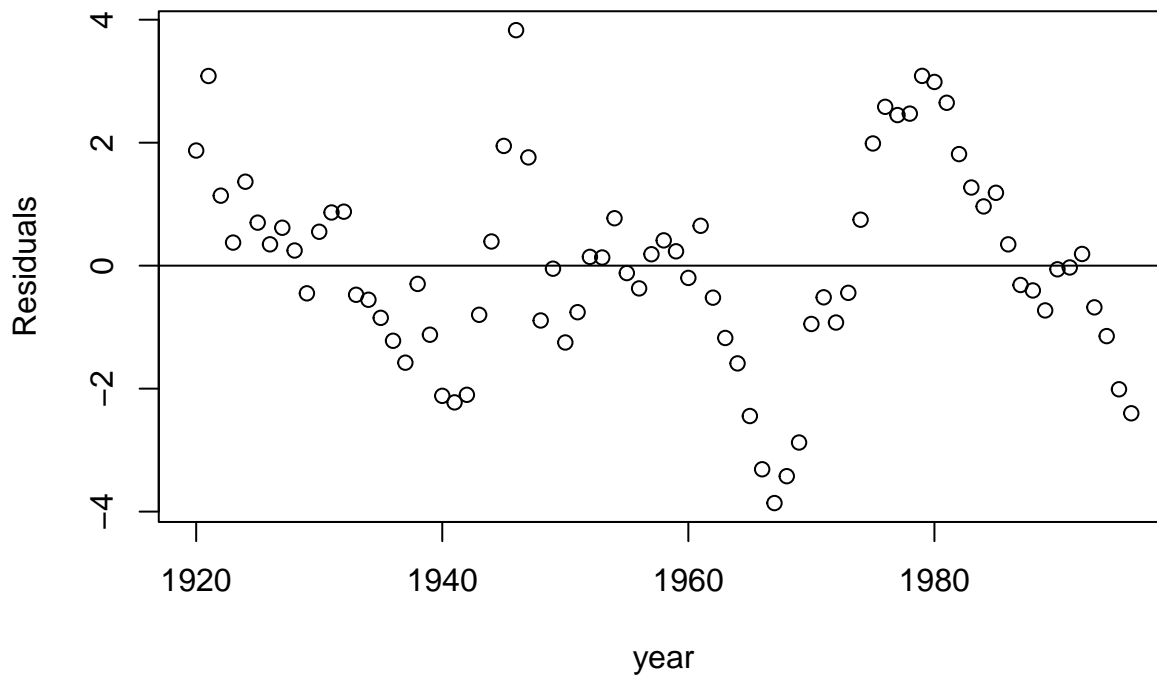
```
lm.model <- lm(divorce ~ unemployed + femlab + marriage + birth +
               military, data = divusa)
summary(lm.model)
```

```
##
## Call:
## lm(formula = divorce ~ unemployed + femlab + marriage + birth +
```

```
##      military, data = divusa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.8611 -0.8916 -0.0496  0.8650  3.8300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.48784    3.39378   0.733  0.4659
## unemployed  -0.11125    0.05592  -1.989  0.0505 .
## femlab       0.38365    0.03059  12.543 < 2e-16 ***
## marriage     0.11867    0.02441   4.861 6.77e-06 ***
## birth       -0.12996    0.01560  -8.333 4.03e-12 ***
## military    -0.02673    0.01425  -1.876  0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.65 on 71 degrees of freedom
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9152
## F-statistic: 165.1 on 5 and 71 DF,  p-value: < 2.2e-16
```

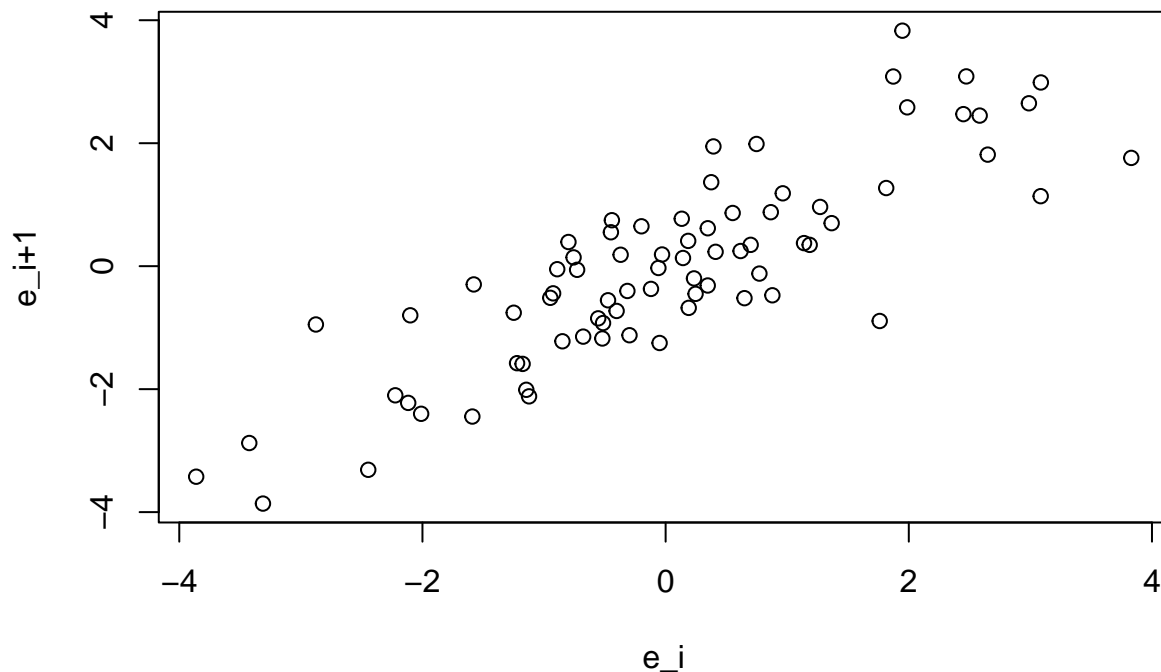
#Plot of Residuals vs. Years

```
plot(lm.model$residuals ~ year, na.omit(divusa), ylab="Residuals")
abline(h=0)
```



#Plot of Residuals (t+1) vs. Residuals (t)

```
n <- dim(divusa)[1]
plot(tail(lm.model$residuals, n-1) ~ head(lm.model$residuals, n-1),
     xlab = "e_i", ylab = "e_i+1")
```



```
cor(tail(lm.model$residuals, n-1), head(lm.model$residuals, n-1))
```

```
## [1] 0.8469792
```

```
#Linear Model of Residuals (t+1) ~ Residuals (t)
```

```
lm.residuals <- lm(tail(lm.model$residuals, n-1) ~ head(lm.model$residuals, n-1) )
summary(lm.residuals)
```

```
##
## Call:
## lm(formula = tail(lm.model$residuals, n - 1) ~ head(lm.model$residuals,
##      n - 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.34046 -0.54703 -0.08307  0.47315  2.22203
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.05155    0.09769  -0.528   0.599
## head(lm.model$residuals, n - 1)  0.85213    0.06218  13.705  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8515 on 74 degrees of freedom
## Multiple R-squared:  0.7174, Adjusted R-squared:  0.7136
## F-statistic: 187.8 on 1 and 74 DF,  p-value: < 2.2e-16
```

From the residuals vs. years plot we can see that there is a positive serial correlation (residuals are followed, in time, by other residuals of the same sign and about the same magnitude) between the residuals.

The plot of residuals (t+1) vs. residuals (t) indicates a positive correlation between residuals at this time point.

The model shows a 0.1% significance between of predictor residuals ϵ_i for response ϵ_{i+1} .