[1]
```
library(MASS)
library(lmtest)
library(gridExtra)
library(zoo)
library(ggcorrplot)
library(tidyverse)
library(readxl)
```

```
Loading required package: zoo

Attaching package: ‘zoo’

The following objects are masked from ‘package:base’:

    as.Date, as.Date.numeric

Loading required package: ggplot2
── Attaching packages ─────────────────────────────────────── tidyverse
1.2.1 ──
✔ tibble  2.1.3     ✔ purrr   0.3.2
✔ tidyr   1.0.0     ✔ dplyr   0.8.3
✔ readr   1.3.1     ✔ stringr 1.4.0
✔ tibble  2.1.3     ✔ forcats 0.4.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::combine() masks gridExtra::combine()
✖ dplyr::filter()  masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
✖ dplyr::select()  masks MASS::select()
```

Sections 1 and 2

[2]
```
original_df <- readxl::read_xlsx("S14 UDJ Harmon.xlsx")
```

[3]
```
original_df
```

A tibble: 48 × 9

TIME	Sales	CP	CP(t-1)	CP(t-2)	DA	DA(t-1)	DA(t
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<db

TIME	Sales	CP	CP(t-1)	CP(t-2)	DA	DA(t-1)	DA(t
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl
1	425075	75253	0	0	457732	352872	3040
2	315305	15036	75253	0	254396	457732	3528
3	367286	134440	15036	75253	259952	254396	4577
4	429432	119740	134440	15036	267368	259952	2543
5	347874	135590	119740	134440	158504	267368	2599
6	435529	189636	135590	119740	430012	158504	2673
7	299403	9308	189636	135590	388516	430012	1585
8	296505	41099	9308	189636	225616	388516	4300
9	426701	9391	41099	9308	1042304	225616	3885
10	329722	942	9391	41099	974092	1042304	2256
11	281783	1818	942	9391	301892	974092	1042
12	166391	672	1818	942	76148	301892	9740
13	629404	548704	672	1818	0	76148	3018
14	263467	52819	548704	672	315196	0	7614
15	398320	2793	52819	548704	703624	315196	0
16	376569	27749	2793	52819	198464	703624	3151
17	444404	21887	27749	2793	478880	198464	7036
18	386986	1110	21887	27749	457172	478880	1984
19	414314	436	1110	21887	709480	457172	4788
20	253493	1407	436	1110	45380	709480	4571
21	484365	376650	1407	436	28080	45380	7094
22	305989	122906	376650	1407	111520	28080	4538
23	315407	15138	122906	376650	267200	111520	2808
24	182784	5532	15138	122906	354304	267200	1115
25	655748	544807	5532	15138	664712	354304	2672

TIME	Sales	CP	CP(t-1)	CP(t-2)	DA	DA(t-1)	DA(t
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl
26	270483	43704	544807	5532	536824	664712	3543
27	365058	5740	43704	544807	551560	536824	6647
28	313135	9614	5740	43704	150080	551560	5368
29	528210	1507	9614	5740	580800	150080	5515
30	379856	13620	1507	9614	435080	580800	1500
31	472058	101179	13620	1507	361144	435080	5808
32	254516	80309	101179	13620	97844	361144	4350
33	551354	335768	80309	101179	30372	97844	3611
34	335826	91710	335768	80309	150324	30372	9784
35	320408	9856	91710	335768	293044	150324	3037
36	276901	107172	9856	91710	162788	293044	1503
37	455136	299781	107172	9856	32532	162788	2930
38	247570	21218	299781	107172	23468	32532	1627
39	622204	157	21218	299781	4503456	23468	3253
40	429331	12961	157	21218	500904	4503456	2346
41	453156	333529	12961	157	0	500904	4503
42	320103	178105	333529	12961	0	0	5009
43	451779	315564	178105	333529	46104	0	0
44	249482	80206	315564	178105	92252	46104	0
45	744583	5940	80206	315564	4869952	92252	4610
46	421186	36819	5940	80206	376556	4869952	9225
47	397367	234562	36819	5940	376556	376556	4869
48	269096	71881	234562	36819	552536	376556	3765

[4]
```
summary(original_df)
```

```
      TIME           Sales              CP            CP(t-1)      
 Min.   : 1.00   Min.   :166391   Min.   :   157   Min.   :     0  
 1st Qu.:12.75   1st Qu.:298678   1st Qu.:  8466   1st Qu.:  5890  
 Median :24.50   Median :371928   Median : 38959   Median : 32284  
 Mean   :24.50   Mean   :382522   Mean   :100953   Mean   : 99456  
 3rd Qu.:36.25   3rd Qu.:437748   3rd Qu.:125790   3rd Qu.:125790  
 Max.   :48.00   Max.   :744583   Max.   :548704   Max.   :548704  
    CP(t-2)             DA             DA(t-1)           DA(t-2)      
 Min.   :     0   Min.   :      0   Min.   :      0   Min.   :      0  
 1st Qu.:  5688   1st Qu.: 108101   1st Qu.: 108101   1st Qu.: 108101  
 Median : 24818   Median : 297468   Median : 297468   Median : 297468  
 Mean   : 94569   Mean   : 497807   Mean   : 493647   Mean   : 492136  
 3rd Qu.:120532   3rd Qu.: 484386   3rd Qu.: 463019   3rd Qu.: 463019  
 Max.   :548704   Max.   :4869952   Max.   :4869952   Max.   :4869952  
    SeasIndx    
 Min.   : 65.0  
 1st Qu.: 96.5  
 Median :103.0  
 Mean   :100.1  
 3rd Qu.:108.5  
 Max.   :119.0  
```

[5]
```
str(original_df)
```

```
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':	48 obs. of  9 variables:
 $ TIME    : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Sales   : num  425075 315305 367286 429432 347874 ...
 $ CP      : num  75253 15036 134440 119740 135590 ...
 $ CP(t-1) : num  0 75253 15036 134440 119740 ...
 $ CP(t-2) : num  0 0 75253 15036 134440 ...
 $ DA      : num  457732 254396 259952 267368 158504 ...
 $ DA(t-1) : num  352872 457732 254396 259952 267368 ...
 $ DA(t-2) : num  304004 352872 457732 254396 259952 ...
 $ SeasIndx: num  113 98 102 107 119 104 107 81 113 97 ...
```

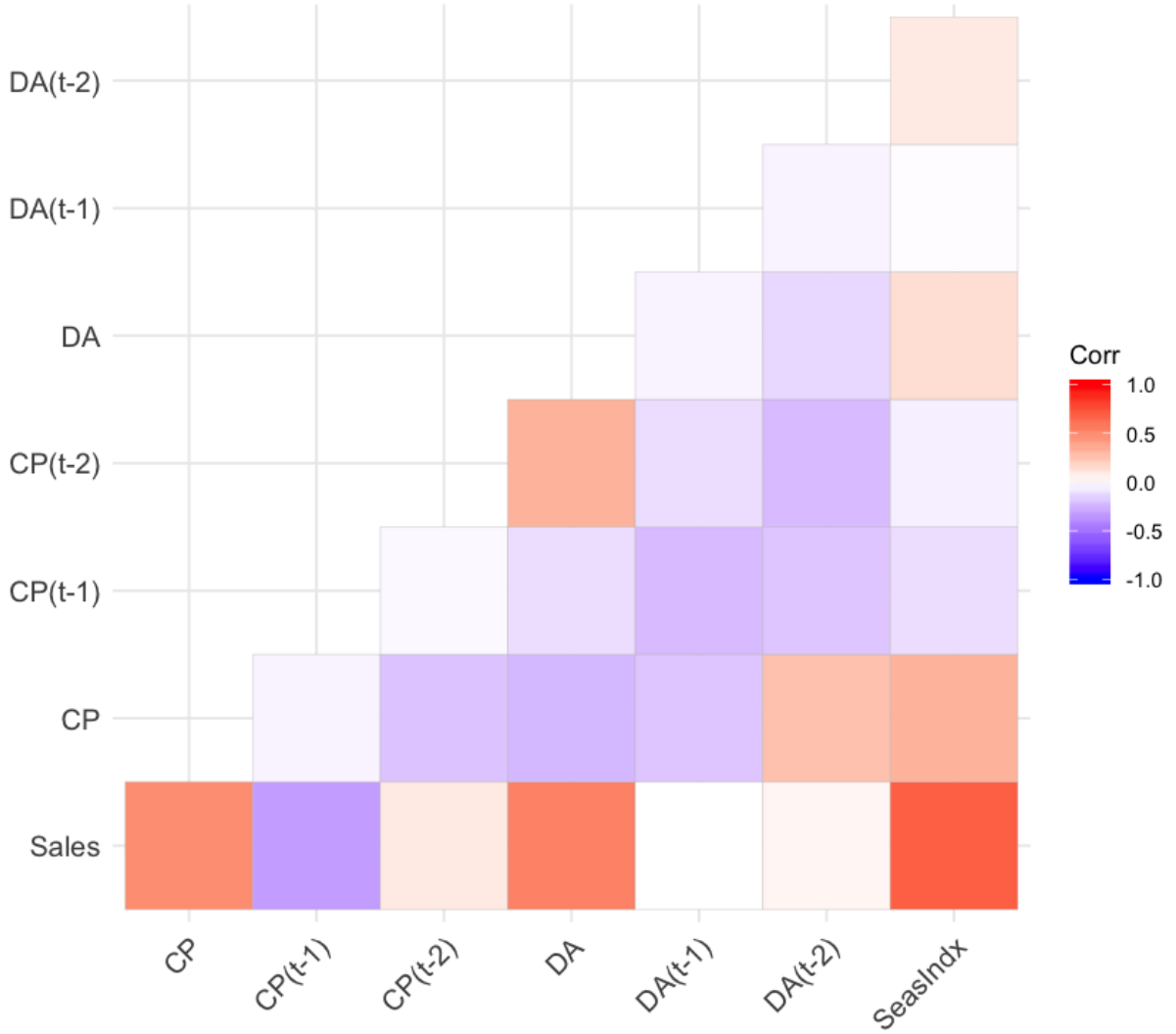
[6]
```
cor(select(original_df, -TIME))
```

A matrix: 8 × 8 of type dbl

	Sales	CP	CP(t-1)	CP(t-2)	DA
Sales	1.000000000	0.48849998	-0.33516467	0.09349610	0.5419
CP	0.488499982	1.00000000	-0.03569904	-0.19752915	-0.235
CP(t-1)	-0.335164666	-0.03569904	1.00000000	-0.02092965	-0.114
CP(t-2)	0.093496103	-0.19752915	-0.02092965	1.00000000	0.3318
DA	0.541992314	-0.23542103	-0.11458073	0.33180797	1.0000

	Sales	CP	CP(t-1)	CP(t-2)	DA
DA(t-1)	0.002475396	-0.19187045	-0.23156354	-0.11414486	-0.042
DA(t-2)	0.036732681	0.26297438	-0.18827697	-0.22699929	-0.128
SeasIndx	0.689601674	0.32575859	-0.11326489	-0.04593521	0.1404

[7]
```
ggcorrplot(cor(select(original_df, -TIME)), type="lower")
```



[8]
```
sum(original_df$Sales[1:12])/12
```

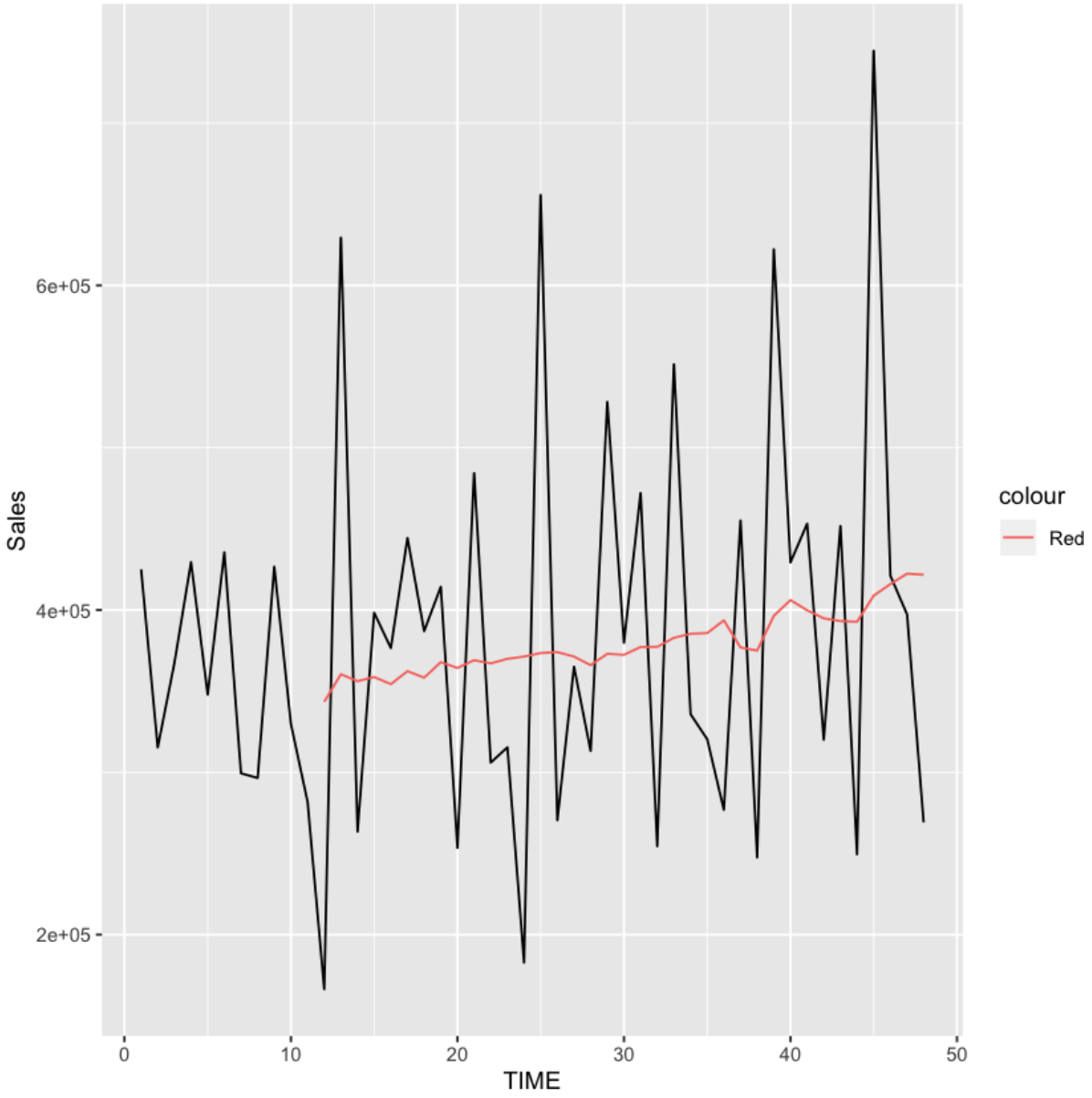
343417.166666667

[9]
```
zoo::rollmeanr(original_df$Sales, 12)
```

```
343417.166666667    360444.583333333    356124.75    358710.916666667
354305.666666667    362349.833333333    358304.583333333    367880.5
364296.166666667    369101.5    367123.75    369925.75    371291.833333333
373487.166666667    374071.833333333    371300    366013.833333333
372997.666666667    372403.5    377215.5    377300.75    382883.166666667
385369.583333333    385786.333333333    393629.416666667    376911.75
375002.333333333    396431.166666667    406114.166666667
399859.666666667    394880.25    393190.333333333    392770.833333333
408873.25    415986.583333333    422399.833333333    421749.416666667
```

[10]
```
ggplot(data=original_df, aes(x=TIME, y=Sales)) + geom_line() +
    geom_line(aes(y=zoo::rollmeanr(Sales, 12, na.pad=TRUE),
colour="Red"))
```

```
Warning message:
“Removed 11 rows containing missing values (geom_path).”
```

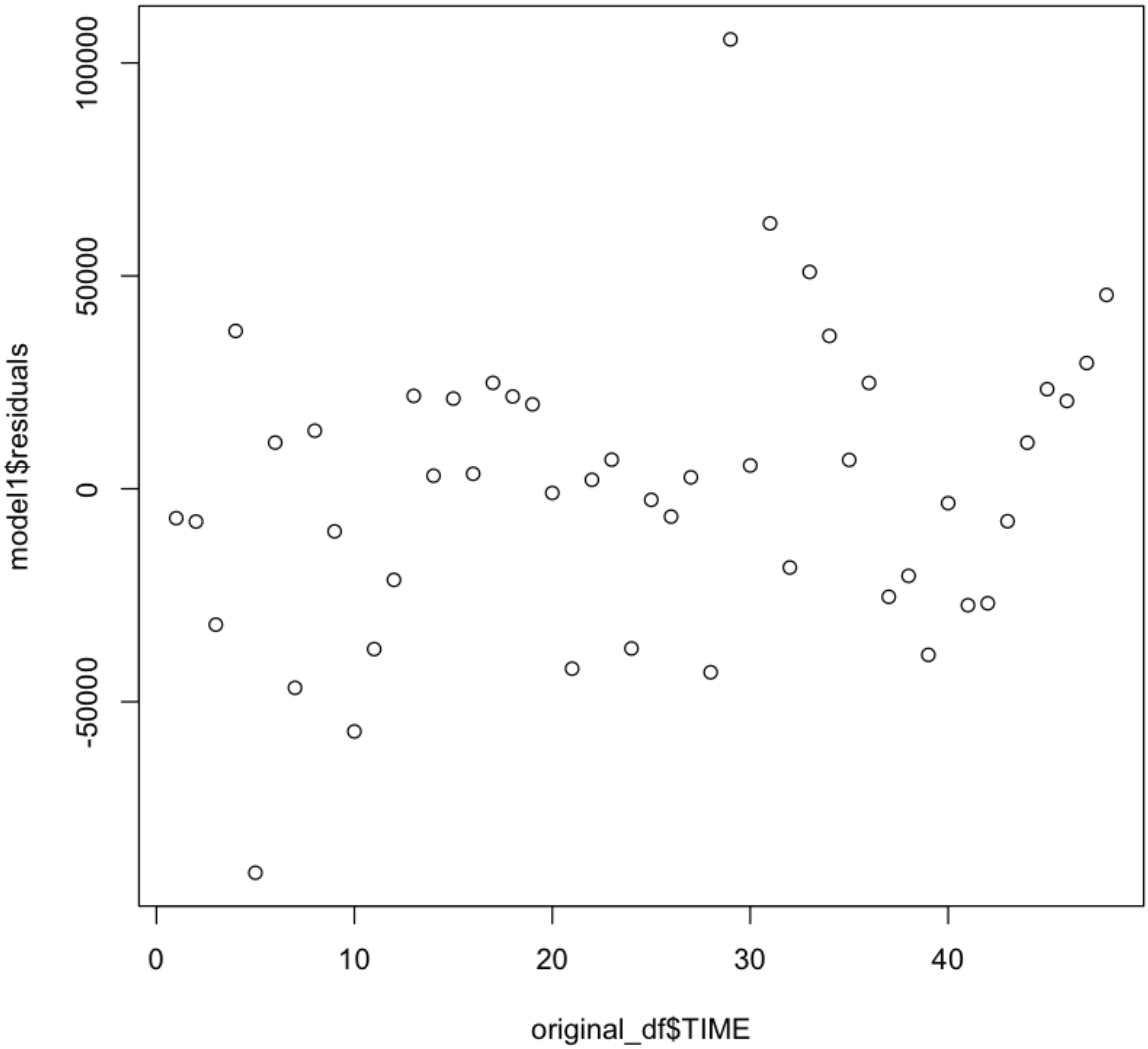


[11]
```
model1 <- lm(Sales ~ . , data=select(original_df, -TIME))
```

[12]
```
summary(model1)
```

```
Call:
lm(formula = Sales ~ ., data = select(original_df, -TIME))

Residuals:
   Min     1Q Median     3Q    Max
-90159 -22383   2405  21306 105537

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.216e+04  4.045e+04  -1.042 0.303460
CP           4.497e-01  4.492e-02  10.012 1.87e-12 ***
`CP(t-1)`   -1.752e-01  4.087e-02  -4.287 0.000111 ***
`CP(t-2)`    8.305e-03  4.219e-02   0.197 0.844940
DA           7.507e-02  6.562e-03  11.439 3.48e-14 ***
`DA(t-1)`    1.043e-02  6.277e-03   1.662 0.104327
`DA(t-2)`   -1.289e-02  6.303e-03  -2.045 0.047486 *
SeasIndx     3.594e+03  4.075e+02   8.822 6.31e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36750 on 40 degrees of freedom
Multiple R-squared:  0.9217,    Adjusted R-squared:  0.908
F-statistic: 67.29 on 7 and 40 DF,  p-value: < 2.2e-16
```

[13]
```
plot(original_df$TIME, model1$residuals)
```



[14]
```
dwtest(model1)
```

```
        Durbin-Watson test

data:  model1
DW = 2.0172, p-value = 0.5163
alternative hypothesis: true autocorrelation is greater than 0
```

[15]
```
model2 <- lm(Sales ~ CP + `CP(t-1)` + DA + `DA(t-2)` + SeasIndx,
data=select(original_df, -TIME))
```

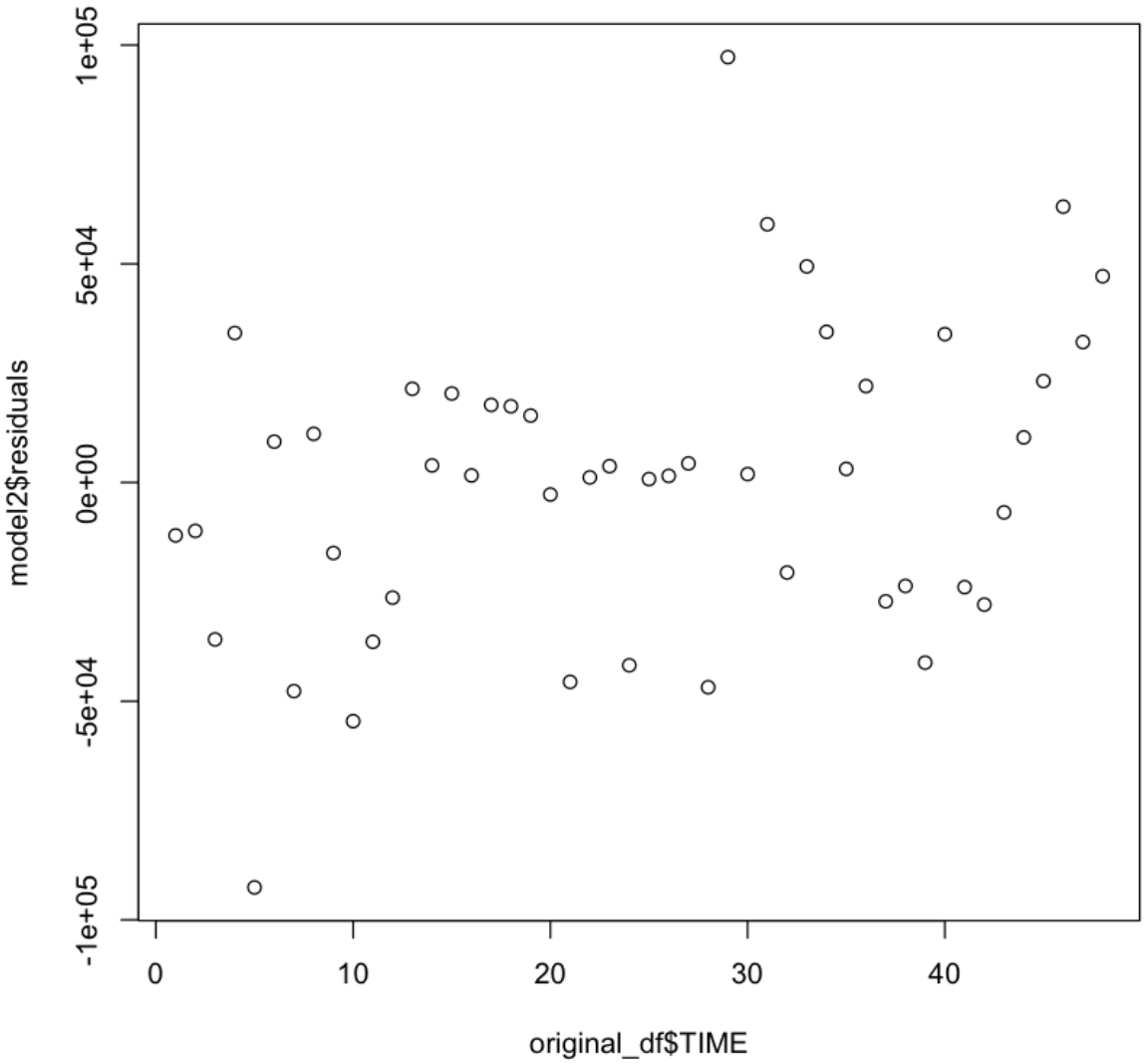
[16]
```
summary(model2)
```

```
Call:
lm(formula = Sales ~ CP + `CP(t-1)` + DA + `DA(t-2)` + SeasIndx,
    data = select(original_df, -TIME))

Residuals:
   Min     1Q Median     3Q    Max
-92569 -24519   1768  20618  97250

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.589e+04  4.014e+04  -0.894   0.3764
CP           4.325e-01  4.399e-02   9.830 1.87e-12 ***
`CP(t-1)`   -1.931e-01  3.976e-02  -4.856 1.70e-05 ***
DA           7.392e-02  6.306e-03  11.722 7.95e-15 ***
`DA(t-2)`   -1.364e-02  6.245e-03  -2.184   0.0346 *
SeasIndx     3.636e+03  4.098e+02   8.871 3.53e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37080 on 42 degrees of freedom
Multiple R-squared:  0.9163,    Adjusted R-squared:  0.9064
F-statistic: 91.98 on 5 and 42 DF,  p-value: < 2.2e-16
```

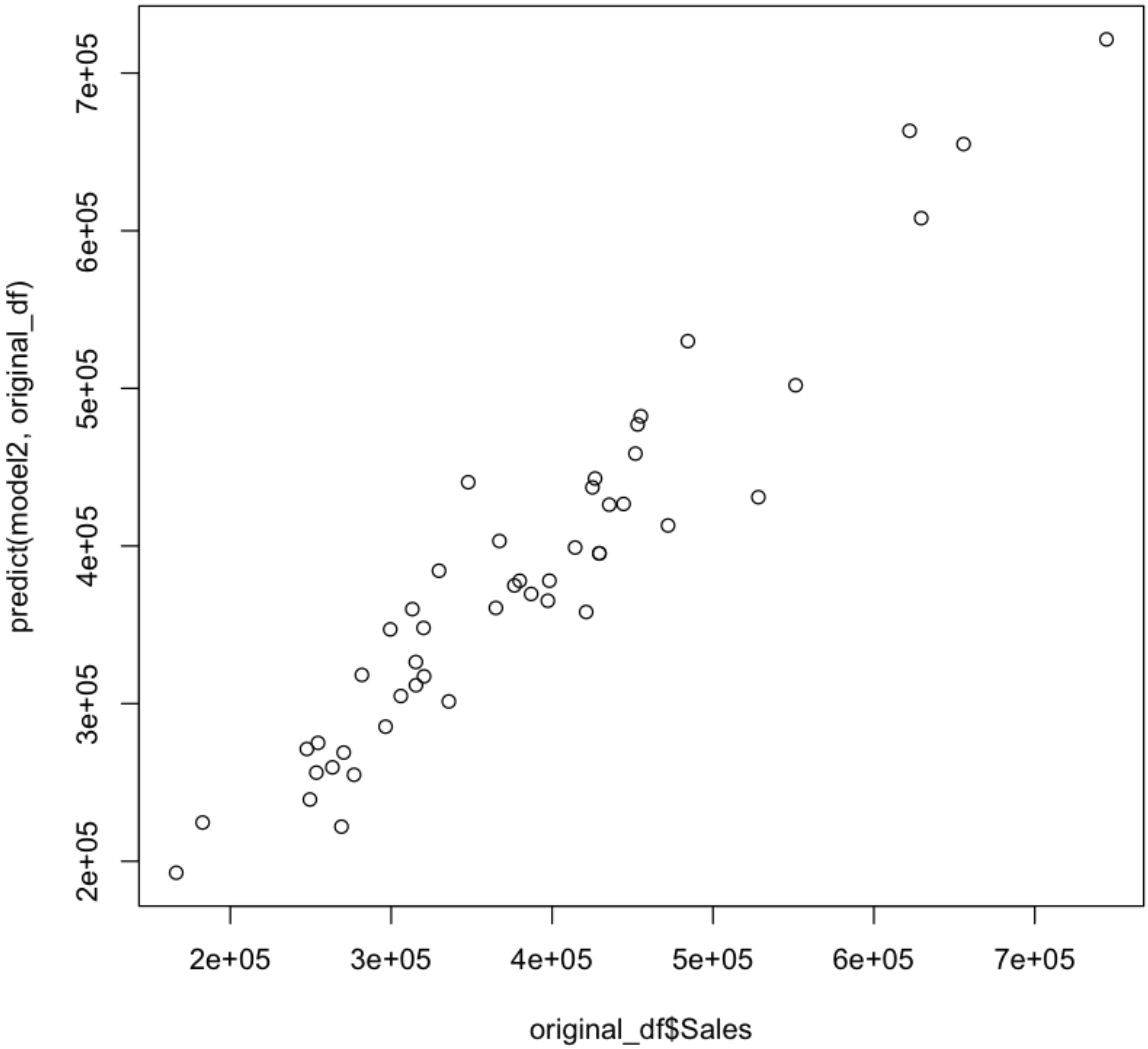
[17]
```
dwtest(model2)
```

```
        Durbin-Watson test

data:  model2
DW = 2.026, p-value = 0.5346
alternative hypothesis: true autocorrelation is greater than 0
```

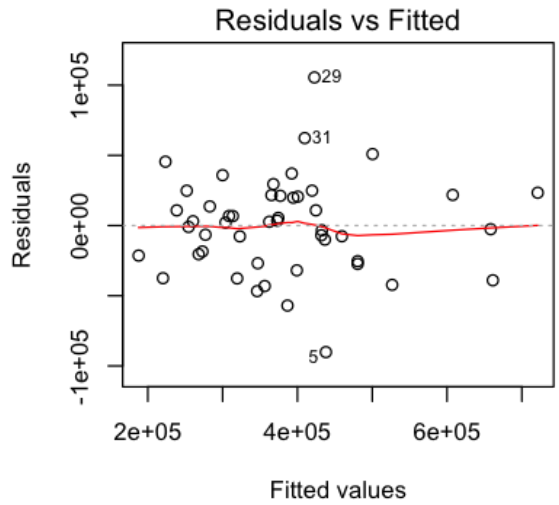
[18]
```
plot(original_df$TIME, model2$residuals)
```

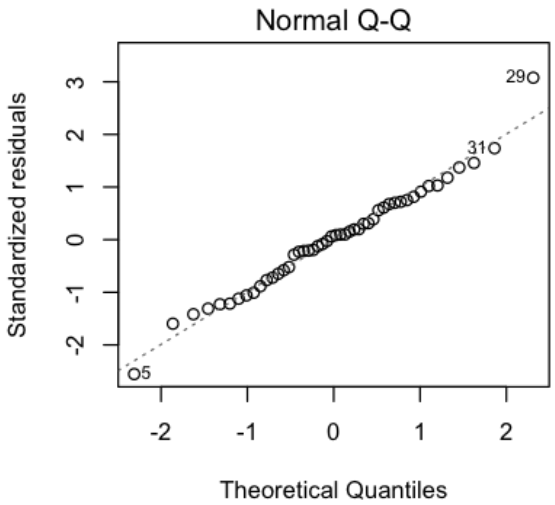


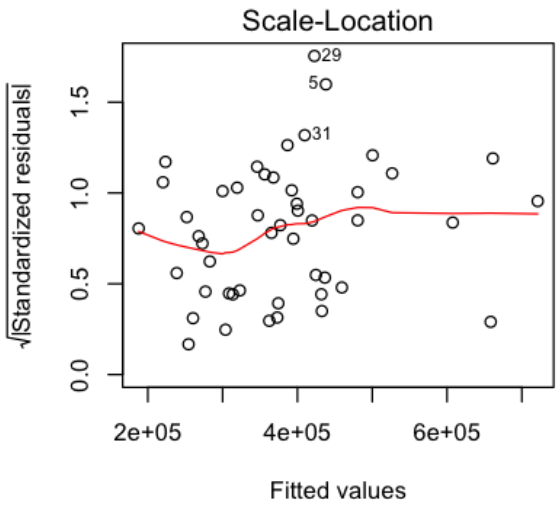

```
[19]  plot(original_df$Sales, predict(model2, original_df))
```

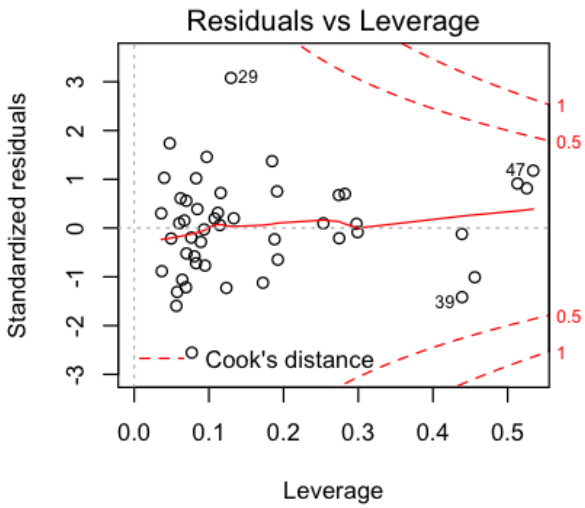



```
[20]  par(mfrow=c(2,2))
      plot(model1)
```

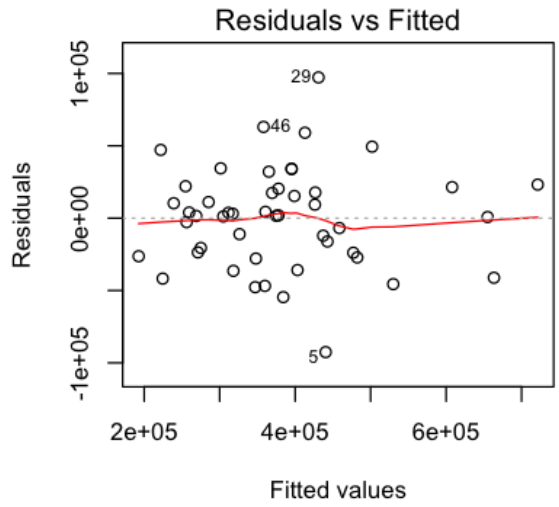


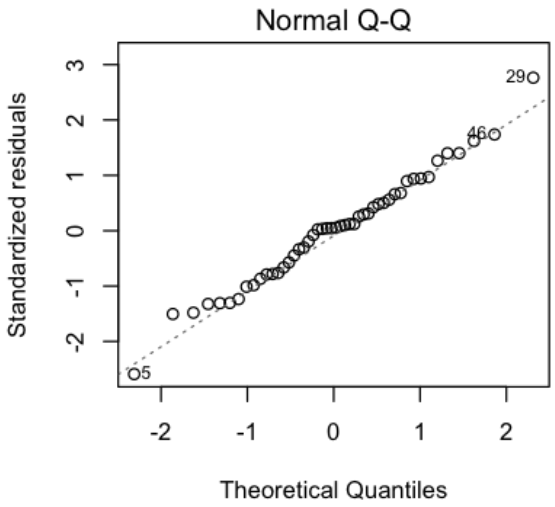


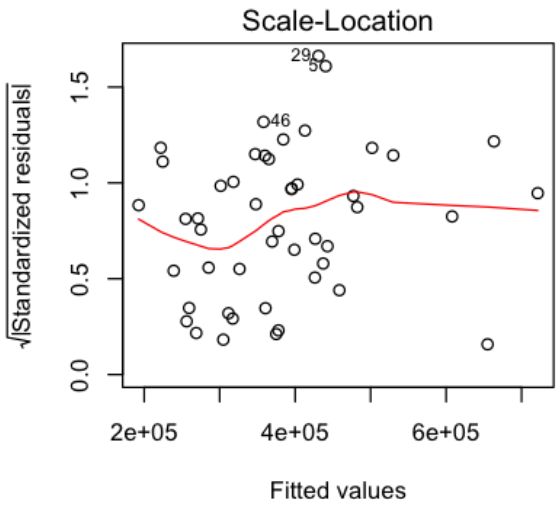


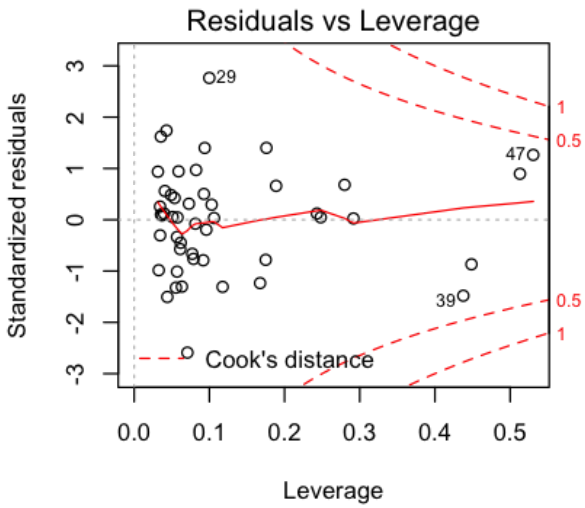



```
[21]  par(mfrow=c(2,2))
      plot(model2)
```

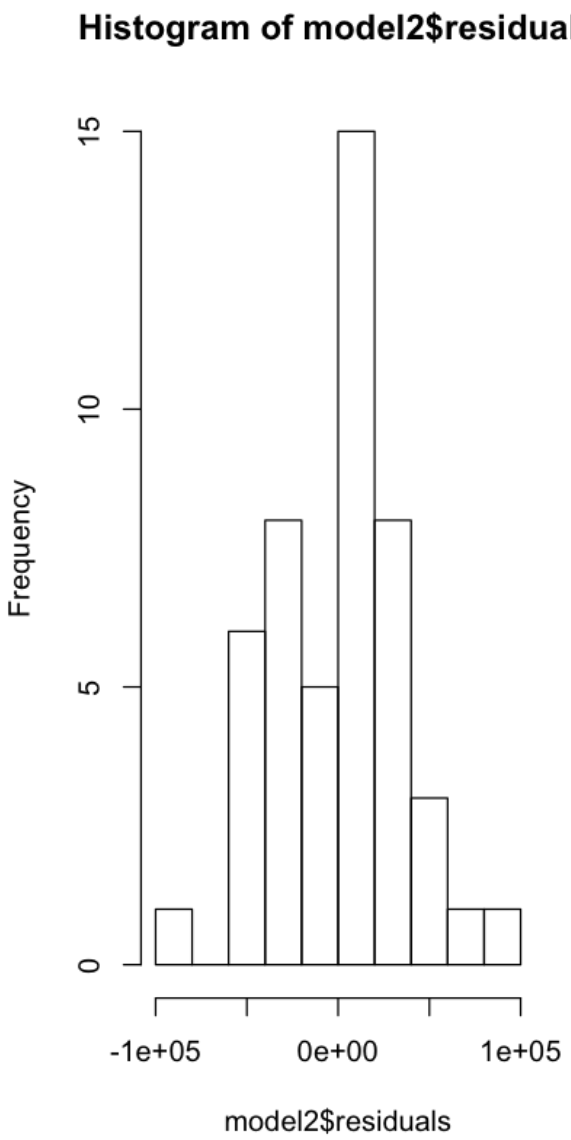


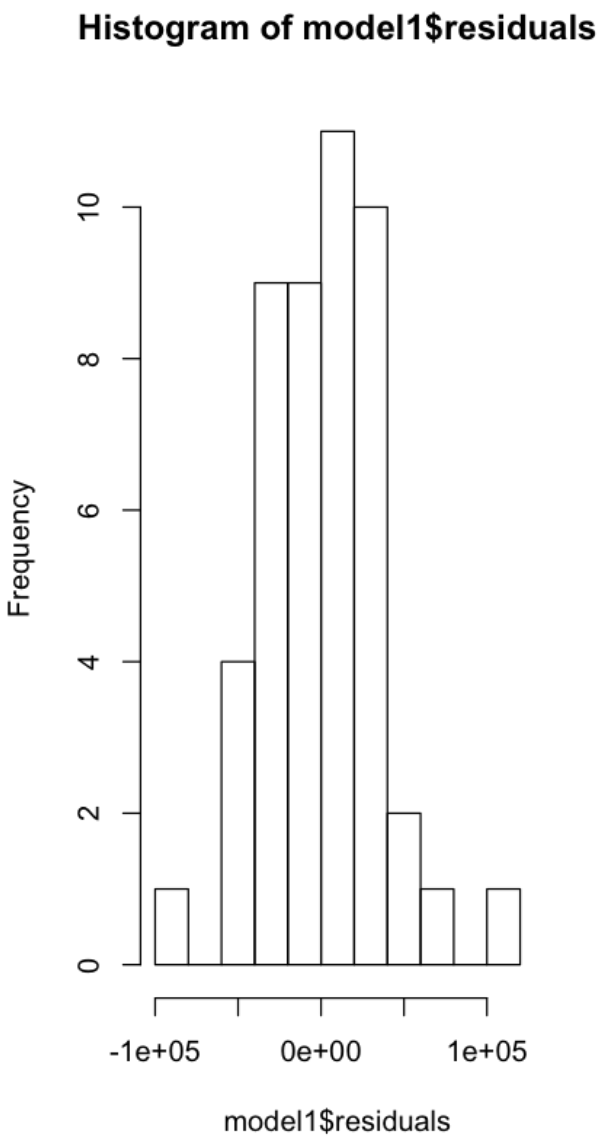







```
[22]  par(mfrow = c(1,2))
      hist(model2$residuals)
      hist(model1$residuals)
```





Section 3

For the next month, January 1988, the planned amount of Consumer Packs is 100 000 cases and Dealer Allowances are set at 500 000$. Use your model to predict sales in January 1988. Please report a point estimate and a 95% prediction interval.

[23]
```
tail(original_df)
```

A tibble: 6 × 9

TIME	Sales	CP	CP(t-1)	CP(t-2)	DA	DA(t-1)	DA(t
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl
43	451779	315564	178105	333529	46104	0	0

TIME	Sales	CP	CP(t-1)	CP(t-2)	DA	DA(t-1)	DA(t
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl
44	249482	80206	315564	178105	92252	46104	0
45	744583	5940	80206	315564	4869952	92252	461(
46	421186	36819	5940	80206	376556	4869952	922!
47	397367	234562	36819	5940	376556	376556	486!
48	269096	71881	234562	36819	552536	376556	376!

[24]
```
# ?predict.lm
```

[25]
```
summary(model2)
```

```
Call:
lm(formula = Sales ~ CP + `CP(t-1)` + DA + `DA(t-2)` + SeasIndx, 
    data = select(original_df, -TIME))

Residuals:
   Min     1Q Median     3Q    Max 
-92569 -24519   1768  20618  97250 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.589e+04  4.014e+04  -0.894   0.3764    
CP           4.325e-01  4.399e-02   9.830 1.87e-12 ***
`CP(t-1)`   -1.931e-01  3.976e-02  -4.856 1.70e-05 ***
DA           7.392e-02  6.306e-03  11.722 7.95e-15 ***
`DA(t-2)`   -1.364e-02  6.245e-03  -2.184   0.0346 *  
SeasIndx     3.636e+03  4.098e+02   8.871 3.53e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37080 on 42 degrees of freedom
Multiple R-squared:  0.9163,    Adjusted R-squared:  0.9064 
F-statistic: 91.98 on 5 and 42 DF,  p-value: < 2.2e-16
```

[26]
```
summary(model2)$sigma
```

37080.3429721249

[27]
```
cp = 100000
da = 500000
cp_lag1 = 71881
da_lag2 = 376556
seasindex = 113
```

[28]
```
 -3.589e+04 + (4.325e-01*cp)  + (-1.931e-01*cp_lag1) + (7.392e-
02*da) + (-1.364e-02*da_lag2) + (3.636e+03*seasindex)
```

436171.55506

[29]
```
1.96 * 37080.3429721249
```

72677.4722253648

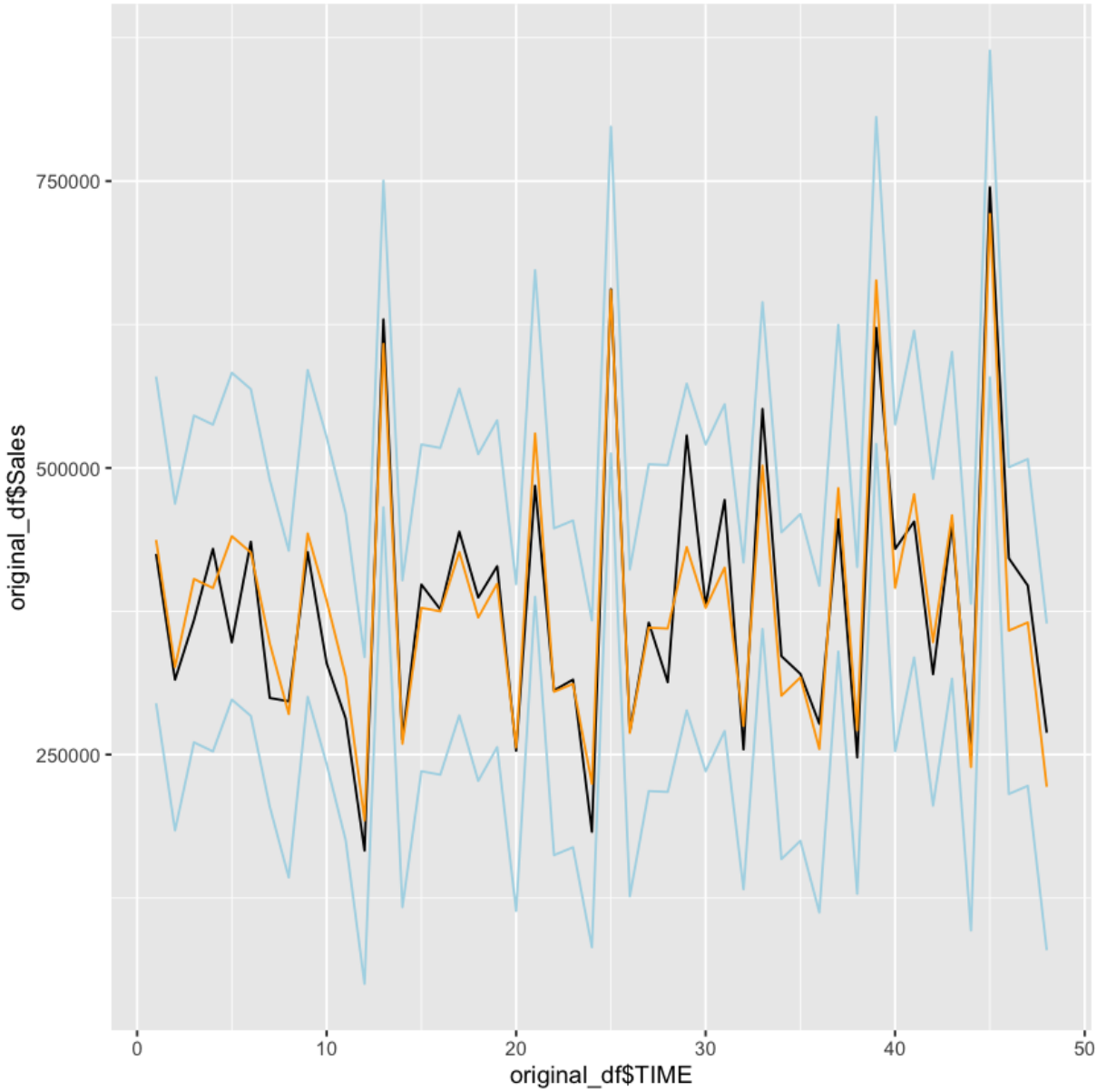
[30]
```
436171.55506 - 72677.4722253648
```

363494.082834635

[31]
```
436171.55506 + 72677.4722253648
```

508849.027285365

[32]
```
ggplot(data=original_df, aes(original_df$TIME,
original_df$Sales)) + geom_line() +
    geom_line(aes(y=predict(model2, original_df)),
color="orange") +
    geom_line(aes(y=predict(model2, original_df) + 1.96 *
72677.4722253648), color="lightblue") +
    geom_line(aes(y=predict(model2, original_df) - 1.96 *
72677.4722253648), color="lightblue")
```



Section 4

[33]
```
summary(model2)
```

```
Call:
lm(formula = Sales ~ CP + `CP(t-1)` + DA + `DA(t-2)` + SeasIndx,
    data = select(original_df, -TIME))

Residuals:
   Min     1Q Median     3Q    Max
-92569 -24519   1768  20618  97250

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -3.589e+04  4.014e+04  -0.894   0.3764
CP           4.325e-01  4.399e-02   9.830 1.87e-12 ***
`CP(t-1)`   -1.931e-01  3.976e-02  -4.856 1.70e-05 ***
DA           7.392e-02  6.306e-03  11.722 7.95e-15 ***
`DA(t-2)`   -1.364e-02  6.245e-03  -2.184   0.0346 *
SeasIndx     3.636e+03  4.098e+02   8.871 3.53e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37080 on 42 degrees of freedom
Multiple R-squared:  0.9163,    Adjusted R-squared:  0.9064
F-statistic: 91.98 on 5 and 42 DF,  p-value: < 2.2e-16
```

```
[ ]
```

Section 5

```
[34]  # dealer allowances
```

```
[ ]
```