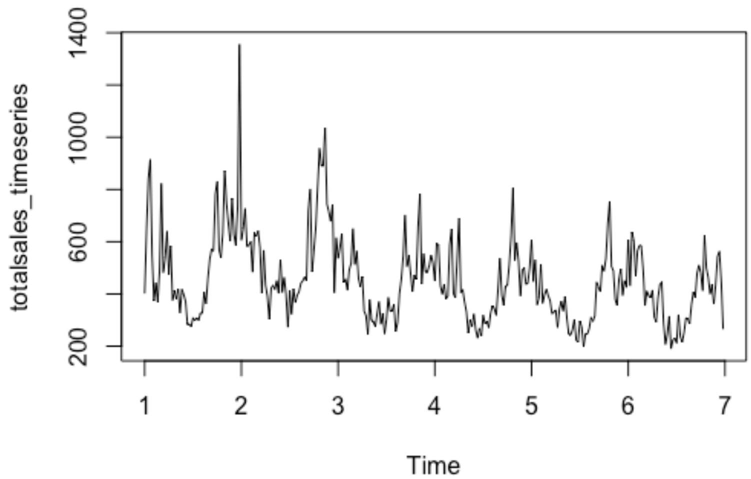
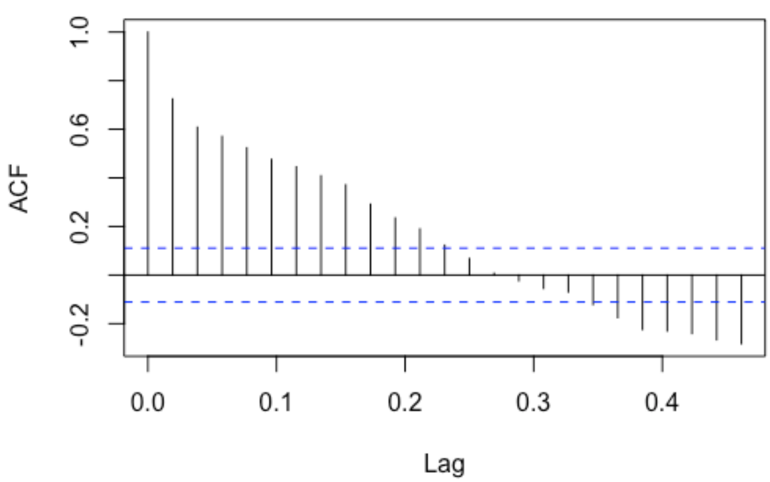
Discussion HW3
Dawei Jia

Part1

g)



Plot1: the time series



Plot2: the autocorrelations

The plot1 means total cost is changed overtime. The plot2 shows most of lines are outside the boundary, which means most of weeks are significantly correlated to itself previously. Thus, we can use forecasting model to forecast future price.

h)

auto.arima select the best model based on the information criteria. Selected model, generally speaking, has lower AIC or BIC. We can add an option in auto.arima (ic="aic") to control which method it use.

i)

autoARMA non-seasonal

```
Series: totalcost_timeseries
ARIMA(0,1,2)

Coefficients:
         ma1      ma2
     -0.4063  -0.1582
s.e.   0.0571   0.0560

sigma^2 estimated as 12160:  log likelihood=-1903.07
AIC=3812.14   AICc=3812.22   BIC=3823.36
```

autoARMA seasonal

```
Series: totalcost _timeseries
ARIMA(0,1,2)(2,0,0)[52]

Coefficients:
         ma1      ma2    sar1    sar2
     -0.4699  -0.1828  0.2490  0.1204
s.e.   0.0598   0.0589  0.0695  0.0762

sigma^2 estimated as 11366:  log likelihood=-1894.56
AIC=3799.12   AICc=3799.31   BIC=3817.82
```

AIC:

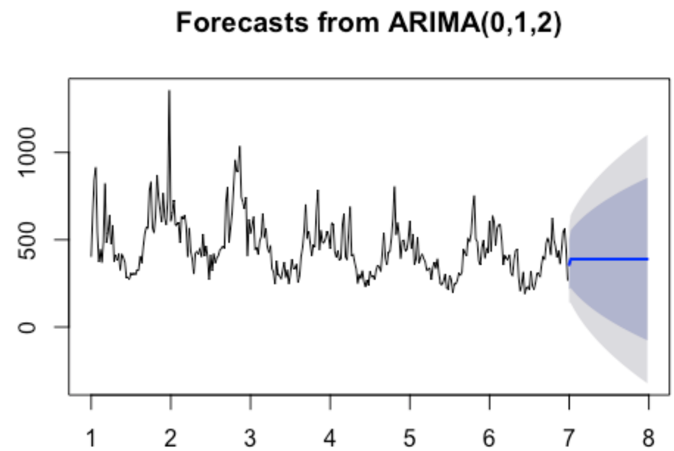
Non-seasonal model AIC =3812.139
Seasonal model AIC =3799.118

MSE:

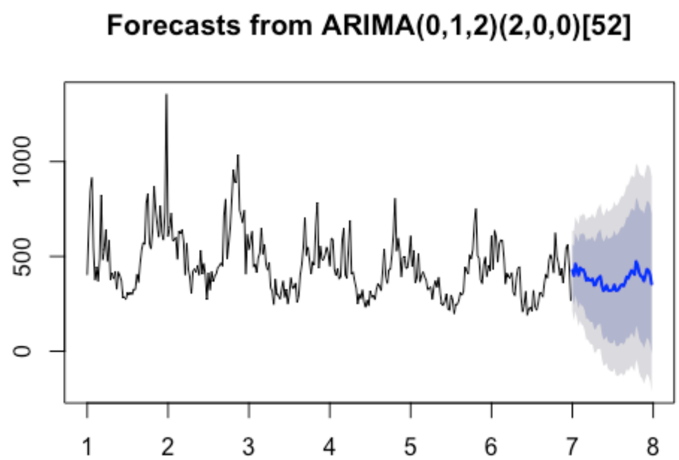
Non-seasonal model MSE=12766.32
Seasonal model MSE=9226.053

Plots:

Non-seasonal model



Seasonal model:



Seasonal model has better AIC.
Seasonal model (MSE= 9226.053) has better MSE than non-seasonal model (MSE= 12766.32).
Seasonal model (CI= (35.88312,728.18413)) has smaller confident interval than non-seasonal model (CI=(-34.43923, 811.43584)).
Overall, I would choose seasonal model.

Part2

e)

2b) Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept)   0.701441   0.004459   157.3   <2e-16 ***
price_per_can -0.329357   0.004829   -68.2   <2e-16 ***
CI: (-0.3388218, -0.3198922) (95% confidence interval)

2c) Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.605494   0.008100   74.75   <2e-16 ***
price_per_can -0.221610   0.007578 -29.24   <2e-16 ***
CI: (-0.2364629, -0.2067571) (95% confidence interval)

2d) Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.67779   0.04263 15.900 < 2e-16 ***
price_per_can -0.16848   0.04961 -3.396 0.000759 ***
CI: (-0.2657156, -0.0712444) (95% confidence interval)

Results above shows coefficients of three different models. The coefficients get bigger after aggregating by different criteria. According to calculating of the confident intervals, the confident intervals are overlapped in 2c and 2d. Thus, the difference of coefficients between 2d and 2c is not statistically significant. On the contrary, the difference of coefficients between 2b and 2c is statistically significant.

f)
When you aggregate the weekly data in part 2d to the monthly level, the change of original data is bigger than aggregating by week. Thus, we expect that coefficient should be bigger than 2b (original data) and 2c, 2d (aggregating by week). Because we compress the data more and makes data higher attenuation bias. The coefficient will tend to zero with more attenuation bias.

Part3

h)
Incorporating isFeature makes price coefficient bigger. The change is statistically significant (confidence intervals do not overlap). The sign of feature (in a mail out) and price is positive. The sign of feature and units is also negative. For example, mailed products will have higher price so that they sell less. Thus, the overall direction is negative, then we underestimate the effect of price if we omit feature.

i)
Incorporating storenum makes price coefficient smaller. The change is statistically significant (confidence intervals do not overlap). The sign of storenum and units or price is always at the same direction. For example, big store tends to sell more at higher price). Thus, the overall direction is positive, then we overestimate the effect of price if we omit feature.

j)

Incorporating productnum makes price coefficient bigger. The change is statistically significant (confidence intervals do not overlap). The sign of productnum and units or price is always at different direction. For example, products with high price sell less. Thus, the overall direction is negative, then we underestimate the effect of price if we omit feature.

k)

The standard error of the price coefficient gets bigger as I move from the regression in part 3b to the regression in 3g. When we incorporate more variables into our model, the model could overfitting. Thus, the standard error will go up because the variation goes up.