

# ISOM5610 Homework 2 - Blue Stem

Team 1

30 November 2018

## 1. Data Manipulation

### 1.1 Check Missing Values

```
setwd("~/MSBA/ISOM5610/HW2")
library(readxl)
bluestem <- as.data.frame(read_excel("bluestem.xls", sheet=1))
sum(is.na(bluestem)) #check number of missing data
```

```
## [1] 0
```

```
bluestem[,2:9] <- lapply(bluestem[,2:9], factor) #convert to factors
colnames(bluestem)[1] <- "Sales"
colnames(bluestem)[10] <- "Index"
summary(bluestem)
```

```
##      Sales      Promotion Monday Tuesday Wednesday Thursday Friday
## Min.   : 30.0    0:244      0:210   0:208   0:209       0:208   0:208
## 1st Qu.: 456.5    1: 5      1: 39   1: 41   1: 40       1: 41   1: 41
## Median : 715.6
## Mean   : 944.3
## 3rd Qu.:1360.5
## Max.   :2994.3
## Saturday Sunday      Index
## 0:207      0:244  Min.    :0.920
## 1: 42      1: 5    1st Qu.:1.000
##                      Median :1.180
##                      Mean    :1.664
##                      3rd Qu.:2.530
##                      Max.    :2.870
```

```
str(bluestem)
```

```
## 'data.frame': 249 obs. of 10 variables:
## $ Sales : num 144 159 653 934 1854 ...
## $ Promotion: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 ...
## $ Monday : Factor w/ 2 levels "0","1": 2 1 1 1 1 2 1 1 1 ...
## $ Tuesday : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 2 1 1 ...
## $ Wednesday: Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 2 1 ...
## $ Thursday : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 2 ...
## $ Friday : Factor w/ 2 levels "0","1": 1 1 1 2 1 1 1 1 2 ...
## $ Saturday : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 ...
## $ Sunday : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 ...
## $ Index : num 1 1.18 1.42 2.53 2.87 1 1 1.18 1.42 2.53 ...
```

### 1.2 Deseasonalize Sales

```
## standardize index
bluestem$Weekday <- factor(names(bluestem[3:9])[max.col(bluestem[3:9])], levels = c("Sunday", "Monday",
index_table <- unique(bluestem[names(bluestem) %in% c("Weekday", "Index")])
bluestem$Index <- bluestem$Index*7/sum(index_table[,1])
```

```

index_table <- unique(bluestem[names(bluestem) %in% c("Weekday", "Index")])

# add Deseasonalize Sales
bluestem$desSales <- bluestem$Sales/bluestem$Index

```

### 1.3 Add Lagged Predictors for Promotion

```

library(Hmisc)

## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##     format.pval, units
bluestem$prolag1 <- Lag(bluestem$Promotion,1)
bluestem$prolag2 <- Lag(bluestem$Promotion,2)

```

## 2. Data Exploration - Explore Underlying Patterns

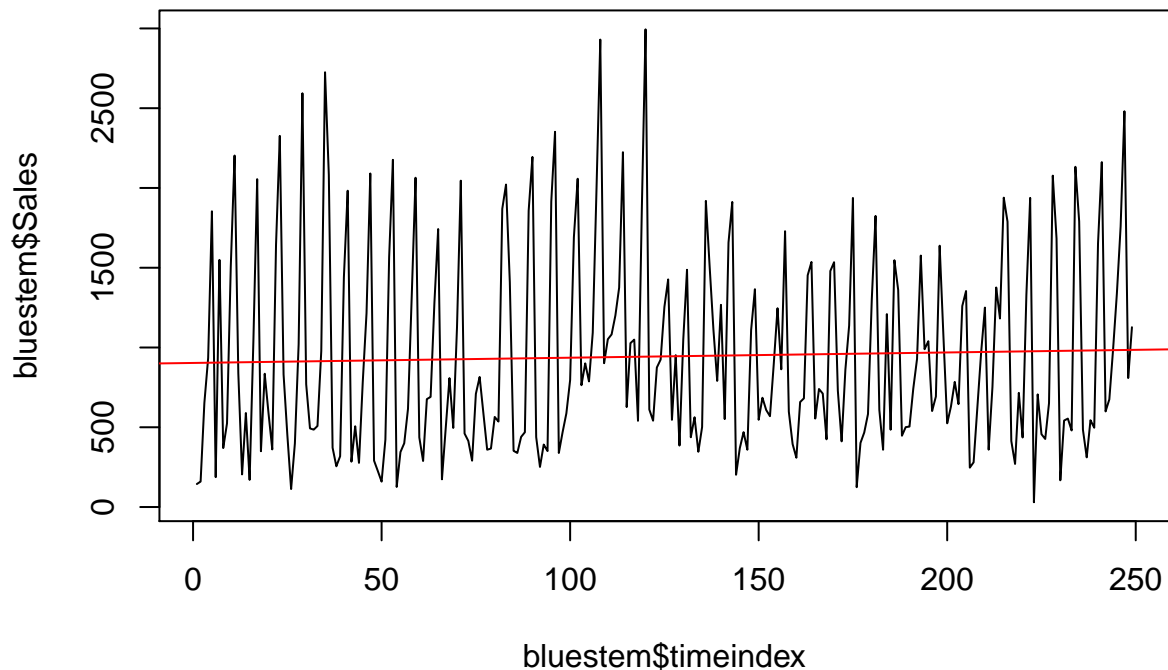
### 2.1 Sales Versus Time Index

```

# add the time index predictor
bluestem$timeindex <- c(1:249)

# see if there is linear trend along the time line
fit0.1 <- lm(Sales ~ timeindex,data=bluestem) # fit linear model
plot(bluestem$timeindex,bluestem$Sales, type = "l")
# the plot is same as the time series plot:
# plot(as.ts(bluestem$Sales))
abline(fit0.1,col=2)

```



```
summary(fit0.1)
```

```
##
## Call:
## lm(formula = Sales ~ timeindex, data = bluestem)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-946.5	-494.7	-229.5	403.0	2051.7

```
##
## Coefficients:
```

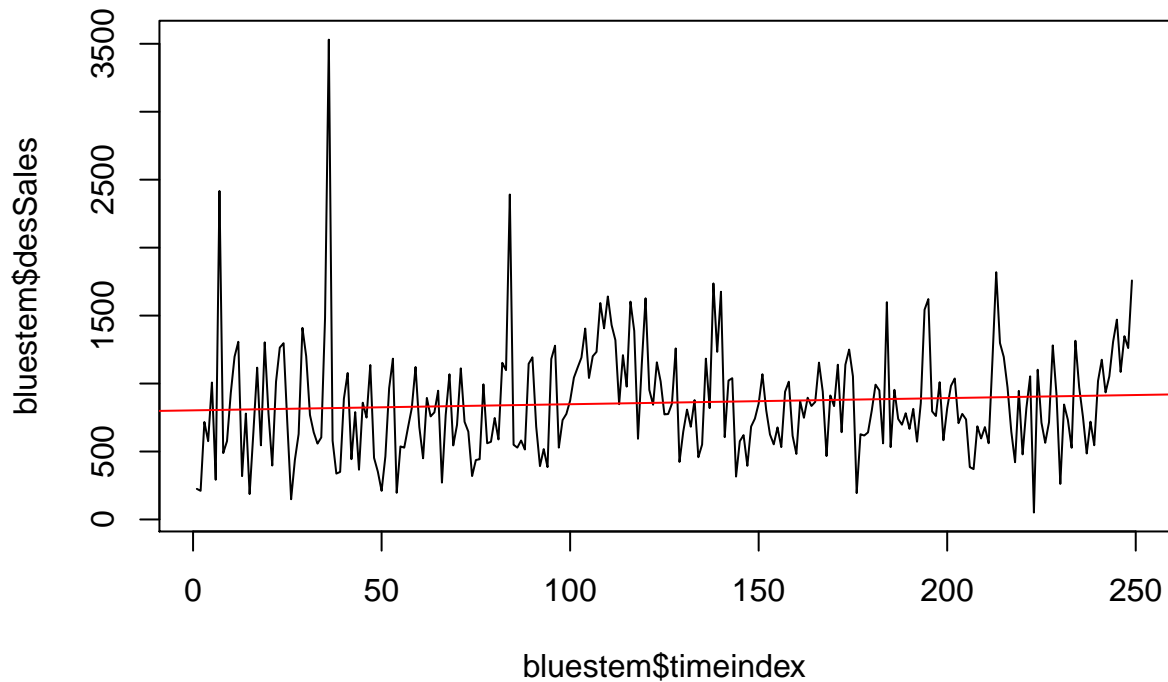
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	903.1718	80.7452	11.185	<2e-16 ***
timeindex	0.3287	0.5600	0.587	0.558

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 635.2 on 247 degrees of freedom
## Multiple R-squared:  0.001393, Adjusted R-squared: -0.00265
## F-statistic: 0.3445 on 1 and 247 DF, p-value: 0.5578
```

## 2.2 Deseasonalized Sales Versus Time Index

```
fit0.2 <- lm(desSales ~ timeindex,data=bluestem)
plot(bluestem$timeindex,bluestem$desSales, type = "l")
# the plot is same as the time series plot:
# plot(as.ts(bluestem$desSales))
```

```
abline(fit0.2,col=2)
```



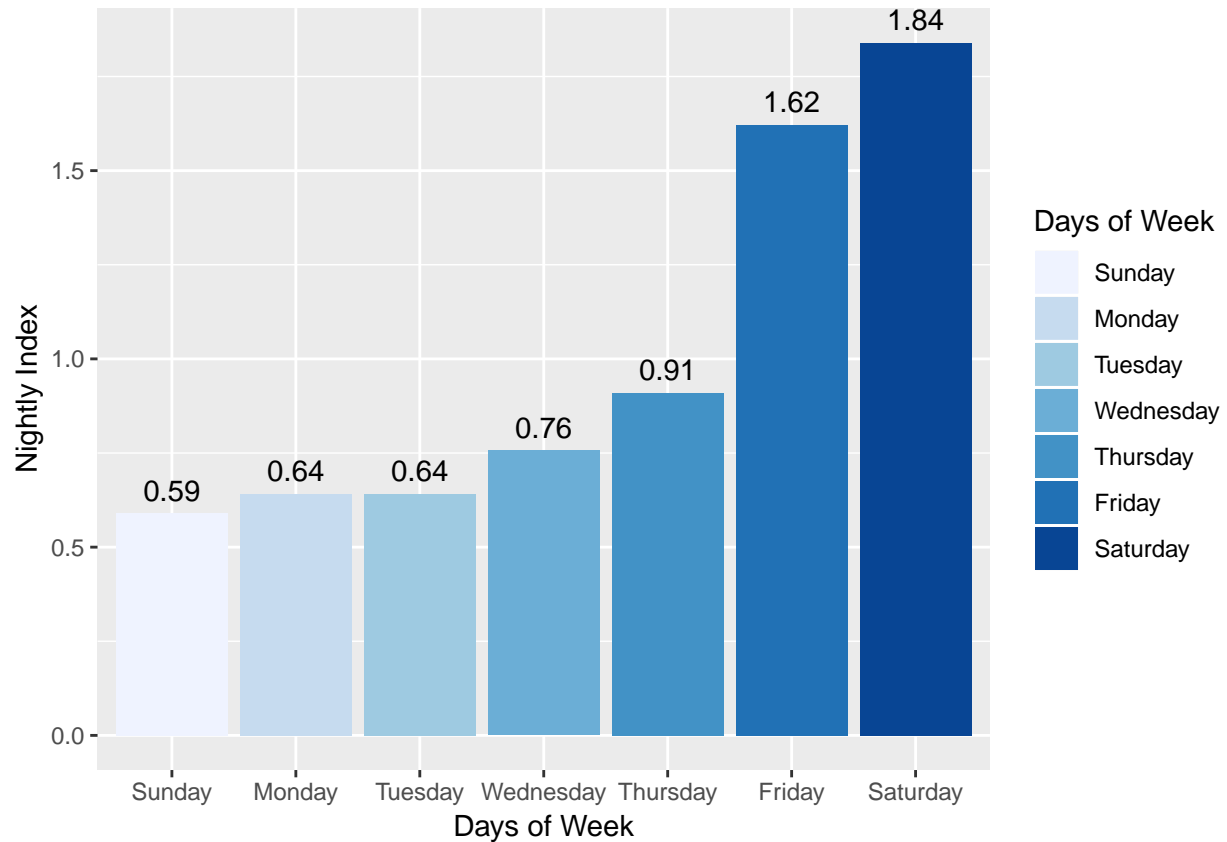
```
summary(fit0.2)
```

```
##
## Call:
## lm(formula = desSales ~ timeindex, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -852.50 -265.67  -67.46  256.53 2711.29
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  802.7256    52.3366   15.338  <2e-16 ***
## timeindex     0.4513     0.3630    1.243   0.215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 411.7 on 247 degrees of freedom
## Multiple R-squared:  0.006221, Adjusted R-squared:  0.002198
## F-statistic: 1.546 on 1 and 247 DF, p-value: 0.2149
```

### 2.3 Nightly Indices Versus Different Days in a Week

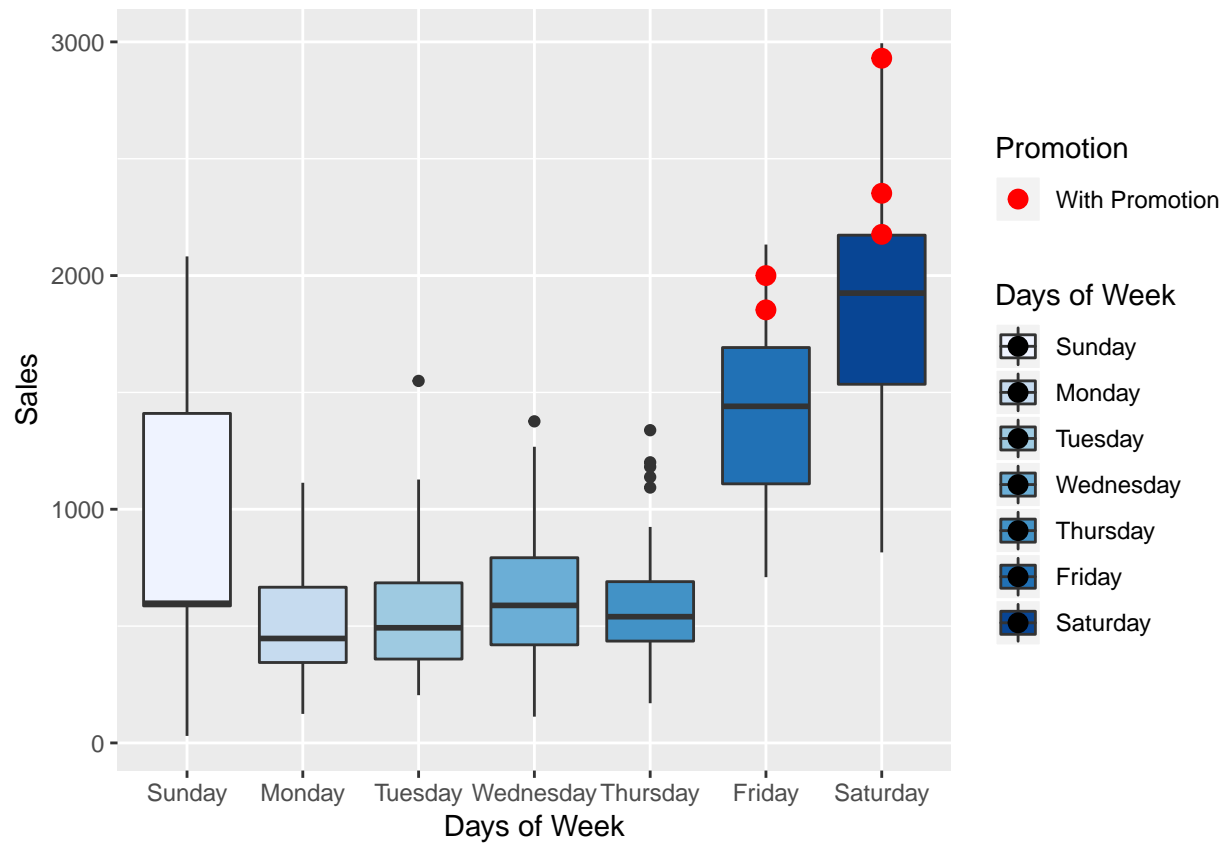
```
library(ggplot2)
library(RColorBrewer)
```

```
ggplot(index_table, aes(x = Weekday, y = Index, fill=Weekday)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = sprintf("%.2f", round(Index, digits = 2))), vjust=-0.6) +
  scale_fill_manual(name="Days of Week", values=brewer.pal(7, "Blues")[1:7]) +
  labs(y="Nightly Index", x="Days of Week")
```

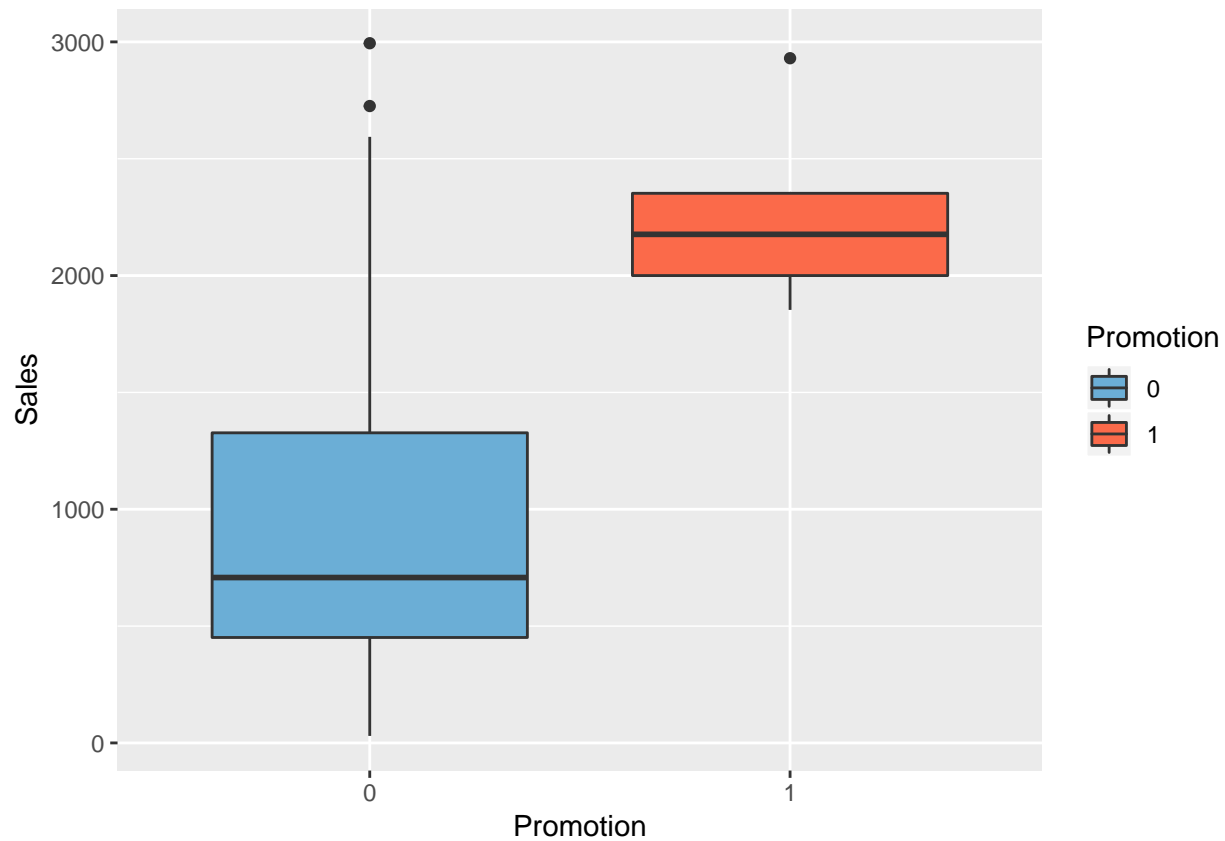


#### 2.4 Sales Versus Different Days in a Week / Promotion

```
# sales vs. different days in a week
ggplot(bluestem, aes(x = Weekday, y = Sales, fill=Weekday)) +
  geom_boxplot() +
  geom_point(data=subset(bluestem, Promotion==1), mapping=aes(color=Promotion), size=3) +
  scale_fill_manual(name="Days of Week", values=brewer.pal(7, "Blues")[1:7]) +
  scale_color_manual(labels = "With Promotion", values="red") +
  labs(x="Days of Week")
```

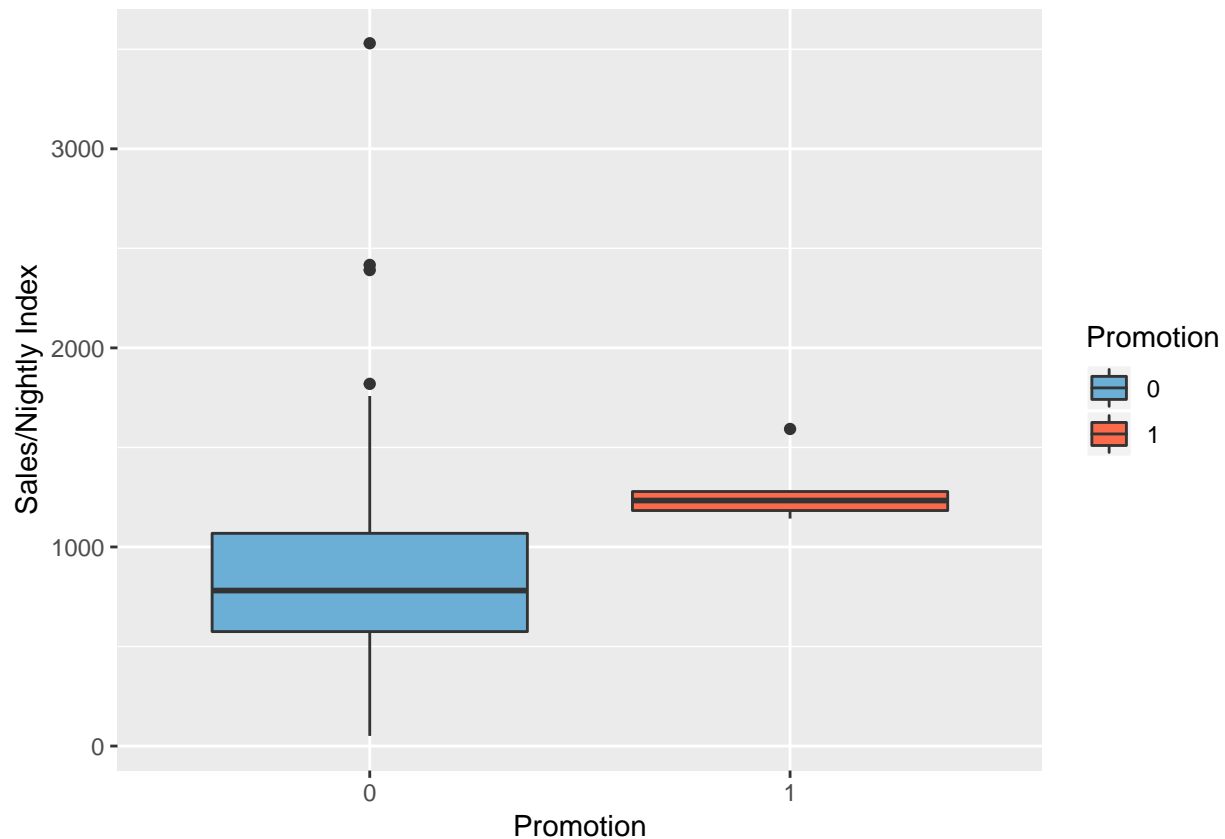


```
# sales with or without promotions
ggplot(bluestem, aes(x = Promotion, y = Sales, fill= Promotion)) +
  geom_boxplot()+
  scale_fill_manual(values=c(brewer.pal(7, "Blues")[4],brewer.pal(7, "Reds")[4]))
```



### 2.5 Deseasonalized Sales Versus Promotion

```
ggplot(bluestem, aes(x = Promotion, y = desSales, fill= Promotion)) +  
  geom_boxplot() +  
  scale_fill_manual(values=c(brewer.pal(7, "Blues")[4],brewer.pal(7, "Reds")[4])) +  
  labs(y="Sales/Nightly Index")
```



### 3. Model Building

```
library(car)
```

```
## Loading required package: carData
```

```
# This function is a combination of
# (1) Diagnostic plots;
# (2) Breusch-Pagan test ("BP test" for short);
# (3) Multicollinearity check for independent variables (by VIF); and
# (4) Cochrane-Orcutt test ("CO test" for short)
check_assumption <- function(testfit){
  #Diagnostic plots
  stdres=rstandard(testfit)
  par(mfrow=c(2,2))
  qqnorm(stdres,main='Normal Probability Plot',xlab='Normal Quantiles',ylab='Standardized Residual Quantiles')
  abline(0,1)
  plot(testfit$fitted.values,stdres,main='Versus Fits',xlab='Fitted Value',ylab='Standardized Residual')
  abline(0,0,lty=3)
  hist(stdres,main='Histogram',xlab='Standardized Residual')
  plot(as.ts(stdres),type="o",main='Versus Order',xlab='TimeIndex',ylab='Standardized Residual')
  abline(0,0,lty=3)
  #BP test
  cat("\n=====Breusch-Pagan test=====\\n\\n")
  print(ncvTest(testfit))
  #Checking multicollinearity for independent variables
  cat("\n=====Checking multicollinearity for independent variables=====\\n\\nVIF:\\n")
}
```



```

tryCatch({
  print(vif(testfit))
},
warning = function(msg) {
  message("Original warning message:")
  message(paste0(msg, "\n"))
  return(NULL)
},
error = function(msg) {
  message("Original error message:")
  message(paste0(msg, "\n"))
  return(NA)
}
)

#Cochrane-Orcutt test
cat("\n=====Cochrane-Orcutt test===== \n")
summary(lm(testfit$residuals~Lag(testfit$residuals, 1)))
}

```

### 3.1 Additive Model

#### 3.1.1 Model 0.1

```

# additive model 0.1
bluestem_1 <- bluestem[,c(1:8)]
names(bluestem_1) # a subset of bluestem with "Sales", "Promotion" dummy and seasonal dummies from "Mon

```

```

## [1] "Sales"      "Promotion" "Monday"     "Tuesday"    "Wednesday" "Thursday"
## [7] "Friday"     "Saturday"

```

```

add.fit.1 <- lm(Sales ~ ., data = bluestem_1)
summary(add.fit.1)

```

```

##
## Call:
## lm(formula = Sales ~ ., data = bluestem_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1033.06  -215.68   -57.86   197.35  1145.85
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    941.2      153.6   6.126 3.64e-09 ***
## Promotion1     593.5      158.7   3.741 0.000229 ***
## Monday1       -441.3      163.2  -2.704 0.007331 **
## Tuesday1      -383.9      162.7  -2.359 0.019138 *
## Wednesday1    -316.6      163.0  -1.943 0.053203 .
## Thursday1     -329.0      162.7  -2.021 0.044347 *
## Friday1        458.3      162.9   2.813 0.005314 **
## Saturday1      907.2      162.9   5.568 6.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 343.5 on 241 degrees of freedom

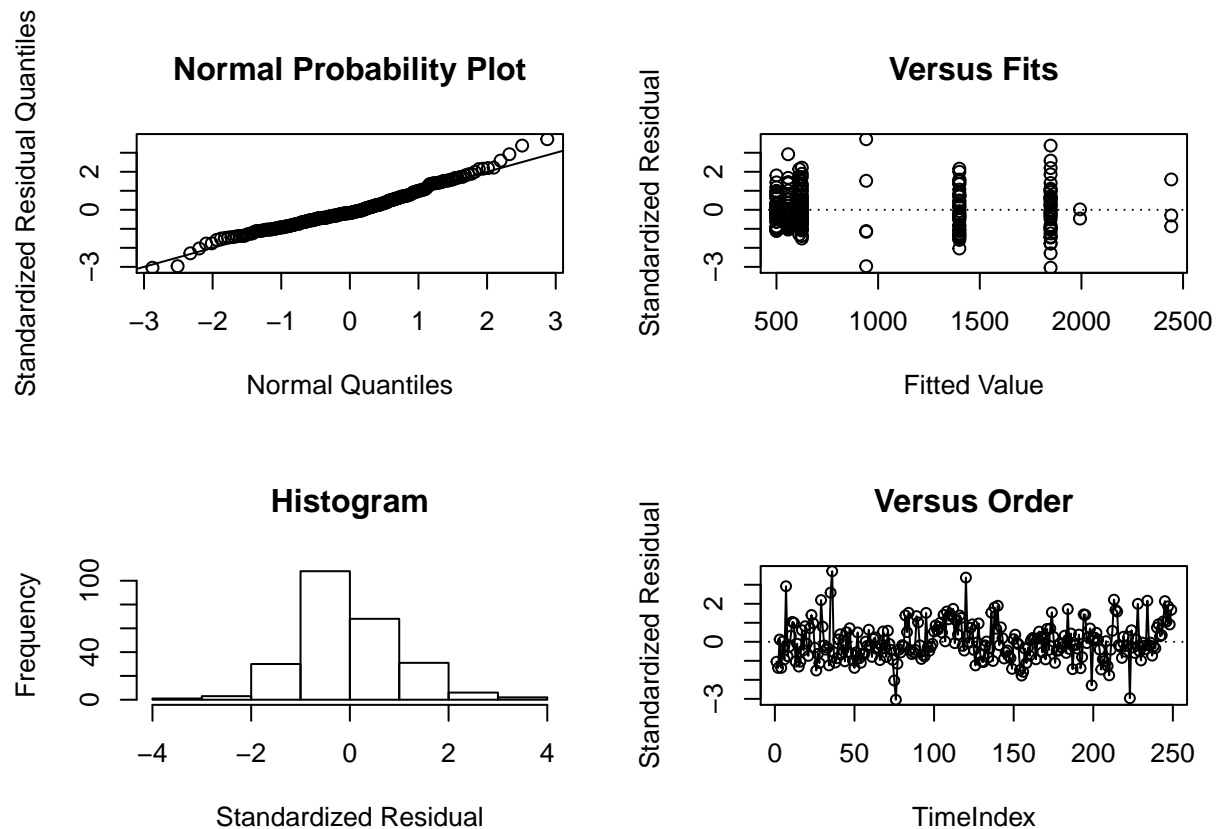
```

```
## Multiple R-squared:  0.7149, Adjusted R-squared:  0.7067
## F-statistic: 86.35 on 7 and 241 DF,  p-value: < 2.2e-16
```

```
add.fit.1$coefficients[2] # check promotion effect
```

```
## Promotion1
##      593.5352
```

```
# Checking model assumptions
check_assumption(add.fit.1)
```



```
##
## =====Breusch-Pagan test=====
##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 16.42694, Df = 1, p = 5.0561e-05
##
## =====Checking multicollinearity for independent variables=====
##
## VIF:
## Promotion    Monday    Tuesday Wednesday Thursday    Friday    Saturday
## 1.045102    7.421687    7.685141    7.554217    7.685141    7.702524    7.852456
##
## =====Cochrane-Orcutt test=====
##
## Call:
```

```
## lm(formula = testfit$residuals ~ Lag(testfit$residuals, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -936.58 -226.46  -42.65  199.86 1180.32
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.03367    20.80333   0.098   0.922
## Lag(testfit$residuals, 1) 0.26101     0.06178   4.225 3.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 327.6 on 246 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.06765, Adjusted R-squared:  0.06386
## F-statistic: 17.85 on 1 and 246 DF, p-value: 3.371e-05
```

### 3.1.2 Model 0.2 & Model 0.3 with Lagged Predictors

```
# additive model 0.2 with lag 1
bluestem_1 <- bluestem[,c(1:8,13)]
names(bluestem_1) # a subset of bluestem with "Sales", "Promotion" dummy, seasonal dummies from "Monday"

## [1] "Sales"      "Promotion" "Monday"     "Tuesday"    "Wednesday"  "Thursday"
## [7] "Friday"     "Saturday"   "prolag1"
```

```
add.fit.2 <- lm(Sales ~ ., data = bluestem_1)
summary(add.fit.2)
```

```
##
## Call:
## lm(formula = Sales ~ ., data = bluestem_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1021.78 -213.09  -45.67   198.63  1140.74
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      941.2      152.6   6.167 2.94e-09 ***
## Promotion1       533.9      160.3   3.330 0.00101 **
## Monday1        -457.7      162.8  -2.811 0.00535 **
## Tuesday1       -383.9      161.7  -2.375 0.01836 *
## Wednesday1     -316.6      161.9  -1.956 0.05165 .
## Thursday1      -329.0      161.7  -2.035 0.04296 *
## Friday1         461.2      161.8   2.850 0.00476 **
## Saturday1       895.9      161.9   5.533 8.27e-08 ***
## prolag1         326.3      160.7   2.031 0.04336 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 341.3 on 239 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.7192, Adjusted R-squared:  0.7098
## F-statistic: 76.53 on 8 and 239 DF, p-value: < 2.2e-16
```

```
coef(summary(add.fit.2))[c(2,9),] # check promotion effect
```

```
##           Estimate Std. Error  t value    Pr(>|t|)
## Promotion1 533.8707   160.3261 3.329906 0.001006169
## prolag11   326.3359   160.6732 2.031054 0.043356439
```

```
# additive model 0.2 with lag 1 and lag 2
```

```
bluestem_1 <- bluestem[,c(1:8,13,14)]
```

```
names(bluestem_1) # a subset of bluestem with "Sales", "Promotion" dummy, seasonal dummies from "Monday
```

```
## [1] "Sales"      "Promotion" "Monday"     "Tuesday"    "Wednesday"
## [6] "Thursday"     "Friday"     "Saturday"   "prolag1"    "prolag2"
```

```
add.fit.3 <- lm(Sales ~ ., data = bluestem_1)
```

```
summary(add.fit.3)
```

```
##
```

```
## Call:
```

```
## lm(formula = Sales ~ ., data = bluestem_1)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -1022.37  -210.83   -44.28   205.01  1140.74
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  941.21     152.54   6.170 2.93e-09 ***
## Promotion1    536.99     160.34   3.349 0.000943 ***
## Monday1     -461.20     162.88  -2.832 0.005031 **
## Tuesday1    -390.53     162.00  -2.411 0.016688 *
## Wednesday1  -304.67     162.03  -1.880 0.061283 .
## Thursday1   -328.96     161.58  -2.036 0.042873 *
## Friday1      461.04     161.77   2.850 0.004756 **
## Saturday1    896.51     161.86   5.539 8.07e-08 ***
## prolag11     309.30     163.37   1.893 0.059538 .
## prolag21      91.28     160.47   0.569 0.570032
```

```
## ---
```

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

```
##
```

```
## Residual standard error: 341.1 on 237 degrees of freedom
```

```
## (2 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.7201, Adjusted R-squared:  0.7095
```

```
## F-statistic: 67.75 on 9 and 237 DF, p-value: < 2.2e-16
```

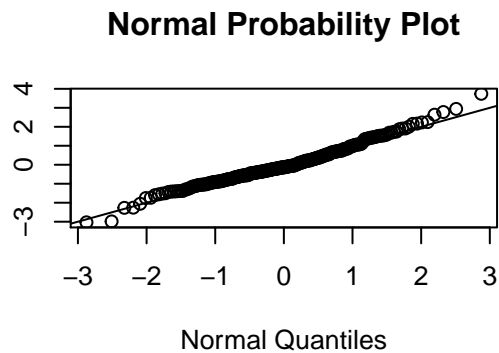
```
coef(summary(add.fit.3))[c(2,9,10),] # check promotion effect
```

```
##           Estimate Std. Error  t value    Pr(>|t|)
## Promotion1 536.98590   160.3422 3.3489986 0.0009433981
## prolag11   309.29720   163.3655 1.8932834 0.0595383850
## prolag21    91.27735   160.4743 0.5687973 0.5700323091
```

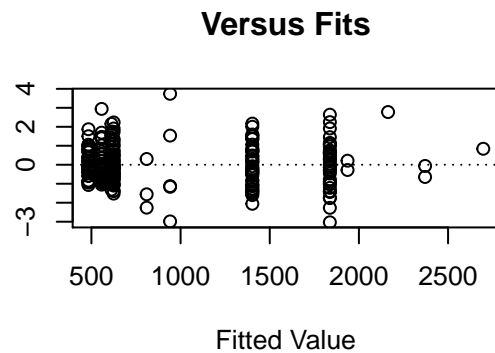
```
# check model assumptions
```

```
check_assumption(add.fit.2)
```

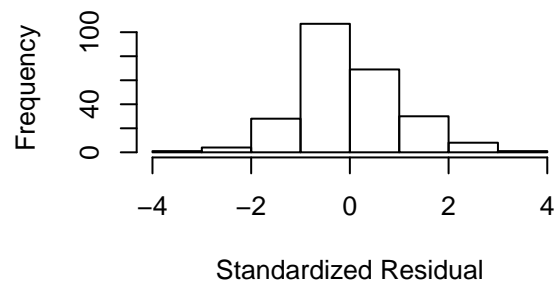
Standardized Residual Quantiles



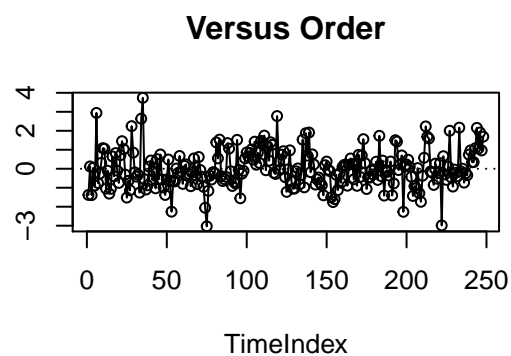
Standardized Residual



**Histogram**



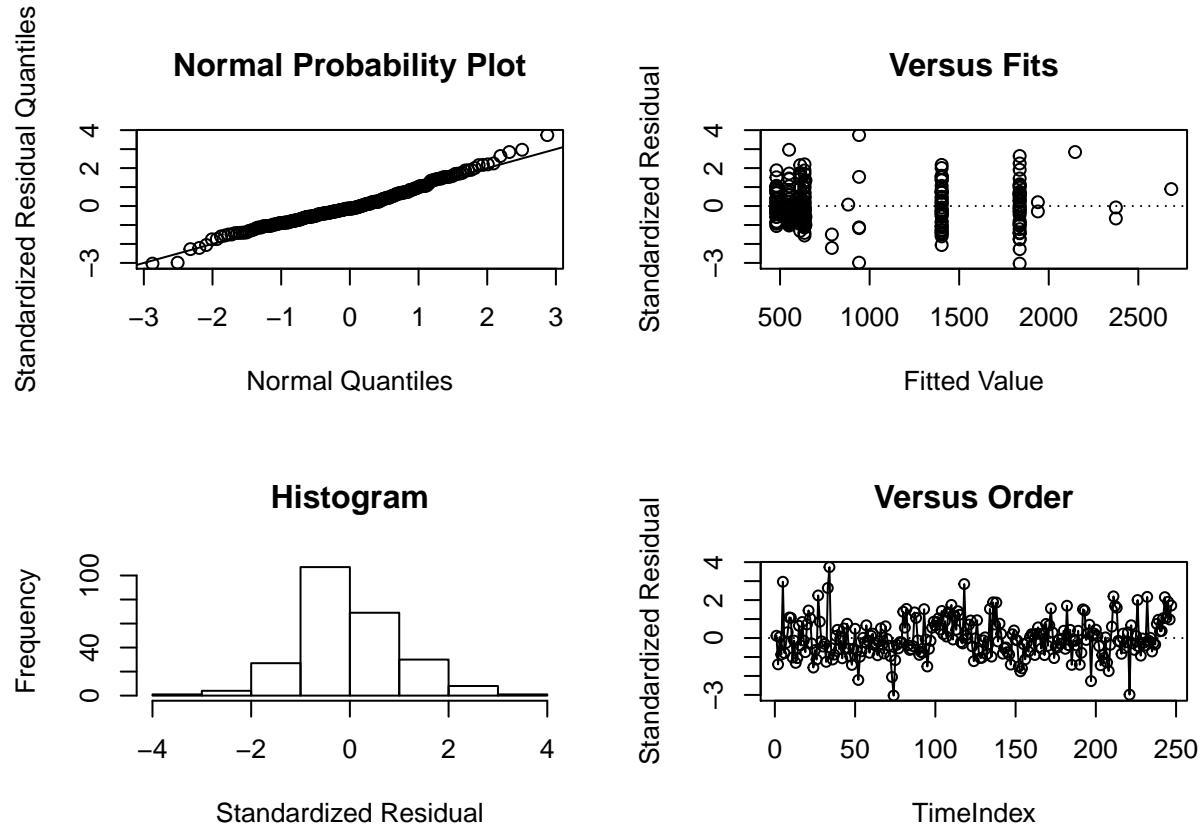
Standardized Residual



```
##
## =====Breusch-Pagan test=====
##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 12.5225, Df = 1, p = 0.00040208
##
## =====Checking multicollinearity for independent variables=====
##
## VIF:
## Promotion    Monday    Tuesday Wednesday  Thursday    Friday    Saturday
## 1.081318    7.326714    7.679032    7.548387    7.679032    7.697005    7.855268
##   prolag1
## 1.086005
##
## =====Cochrane-Orcutt test=====
##
## Call:
## lm(formula = testfit$residuals ~ Lag(testfit$residuals, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -939.75 -228.87  -34.71   194.26 1066.11
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)          2.48040    20.63791    0.120    0.904
## Lag(testfit$residuals, 1) 0.25886    0.06184    4.186 3.96e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 324.3 on 245 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.06675,    Adjusted R-squared:  0.06294
## F-statistic: 17.52 on 1 and 245 DF,  p-value: 3.961e-05
```

```
check_assumption(add.fit.3)
```



```
##
## =====Breusch-Pagan test=====
##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 13.06085, Df = 1, p = 0.00030153
##
## =====Checking multicollinearity for independent variables=====
##
## VIF:
## Promotion    Monday    Tuesday Wednesday Thursday    Friday    Saturday
## 1.082491    7.331567    7.713395    7.410526    7.672874    7.690854    7.849105
##   prolage1   prolage2
## 1.123697    1.084275
##
```

```
## =====Cochrane-Orcutt test=====
##
## Call:
## lm(formula = testfit$residuals ~ Lag(testfit$residuals, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -937.7 -224.2  -35.8   194.3  1074.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.44470     20.70414    0.021   0.983
## Lag(testfit$residuals, 1)  0.26019     0.06222    4.182 4.03e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 324.7 on 244 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.06688,    Adjusted R-squared:  0.06306
## F-statistic: 17.49 on 1 and 244 DF,  p-value: 4.031e-05
```

### 3.2 Multiplicative Model

#### 3.2.1 Model 1.1

```
# multiplicative model 1.1 with promotion dummy only
mtp.fit.1 <- lm(desSales ~ Promotion, data = bluestem)
summary(mtp.fit.1)
```

```
##
## Call:
## lm(formula = desSales ~ Promotion, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -799.52 -274.27  -68.99   218.01  2679.87
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    850.39      26.14   32.527 <2e-16 ***
## Promotion1     435.62     184.50    2.361   0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 408.4 on 247 degrees of freedom
## Multiple R-squared:  0.02207,    Adjusted R-squared:  0.01811
## F-statistic: 5.575 on 1 and 247 DF,  p-value: 0.019
mtp.fit.1$coefficients[2] # check promotion effect on deseasonalized sales
```

```
## Promotion1
##      435.6229
```

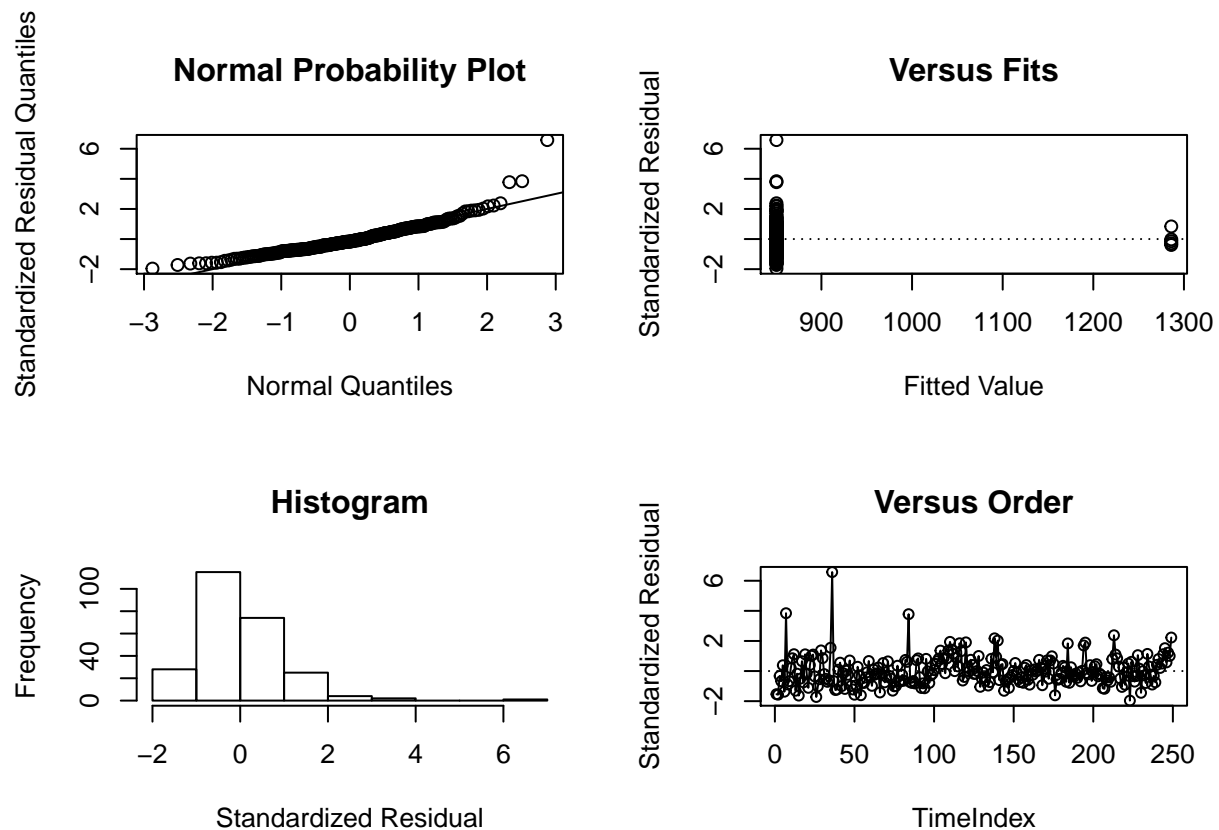
```
# The coefficient of Promotion stands for the boost in revenues without "days of week" effect. Such bo
mtp.fit.1$coefficients[2] * index_table$Index[which(index_table$Weekday=="Friday")]
```

```
## Promotion1
```

```
##      706.4911
mtp.fit.1$coefficients[2] * index_table$Index[which(index_table$Weekday=="Saturday")]

## Promotion1
##      801.4345
# Checking model assumptions
check_assumption(mtp.fit.1)

##
## =====Breusch-Pagan test=====
##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 1.822548, Df = 1, p = 0.17701
##
## =====Checking multicollinearity for independent variables=====
##
## VIF:
## Original error message:
## Error in vif.default(testfit): model contains fewer than 2 terms
```



```
##
## =====Cochrane-Orcutt test=====
##
## Call:
```



```
## lm(formula = testfit$residuals ~ Lag(testfit$residuals, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -844.21 -250.99  -54.69   216.92  2547.78
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.26896    25.32362   0.129  0.89739
## Lag(testfit$residuals, 1)  0.20413     0.06277   3.252  0.00131 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 398.8 on 246 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.04122,    Adjusted R-squared:  0.03733
## F-statistic: 10.58 on 1 and 246 DF,  p-value: 0.001305
```

### 3.2.2 Model 1.2 and Model 1.3 with Lagged Predictors

```
# multiplicative model 1.2 with promotion dummy and lag 1
mtp.fit.2 <- lm(desSales ~ Promotion + prolag1, data = bluestem)
summary(mtp.fit.2)
```

```
##
## Call:
## lm(formula = desSales ~ Promotion + prolag1, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -799.83 -274.51  -65.88   206.38  2679.57
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    850.70      26.33   32.307  <2e-16 ***
## Promotion1     407.73     187.35    2.176   0.0305 *
## prolag11       137.95     187.35    0.736   0.4623
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 407.6 on 245 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.02418,    Adjusted R-squared:  0.01621
## F-statistic: 3.036 on 2 and 245 DF,  p-value: 0.04986
```

```
# multiplicative model 1.3 with promotion dummy, lag 1 and lag 2
mtp.fit.3 <- lm(desSales ~ Promotion + prolag1 + prolag2, data = bluestem)
summary(mtp.fit.3)
```

```
##
## Call:
## lm(formula = desSales ~ Promotion + prolag1 + prolag2, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -799.30 -274.01  -68.77   218.37  2680.09
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)   850.17     26.51  32.069  <2e-16 ***
## Promotion1    415.82    187.10   2.222  0.0272 *
## prolag11      100.12    190.29   0.526  0.5993
## prolag21      183.66    187.10   0.982  0.3273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 406.4 on 243 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.02794,    Adjusted R-squared:  0.01594
## F-statistic: 2.328 on 3 and 243 DF,  p-value: 0.07514
```

### 3.2.3 Model 1.4 & Model 1.5: Using the Logged Response to Solve Autocorrelation

```
# multiplicative model 1.4: try log of deseasonalized sales
bluestem$log_desSales <- log(bluestem$desSales)
mtp.fit.4 <- lm(log_desSales ~ Promotion, data = bluestem)
summary(mtp.fit.4)
```

```
##
## Call:
## lm(formula = log_desSales ~ Promotion, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7001 -0.2730  0.0291  0.3303  1.5398
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.62935    0.03257 203.573  <2e-16 ***
## Promotion1   0.52284    0.22981   2.275  0.0238 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5087 on 247 degrees of freedom
## Multiple R-squared:  0.02053,    Adjusted R-squared:  0.01656
## F-statistic: 5.176 on 1 and 247 DF,  p-value: 0.02376
```

```
coef(summary(mtp.fit.4))[2] # check promotion effect on deseasonalized sales as %
```

```
## [1] 0.5228393
```

```
check_assumption(mtp.fit.4)
```

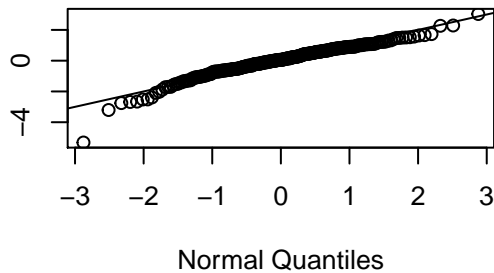
```
##
## =====Breusch-Pagan test=====
##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 2.287233, Df = 1, p = 0.13044
##
## =====Checking multicollinearity for independent variables=====
##
## VIF:
```

```
## Original error message:
```

```
## Error in vif.default(testfit): model contains fewer than 2 terms
```

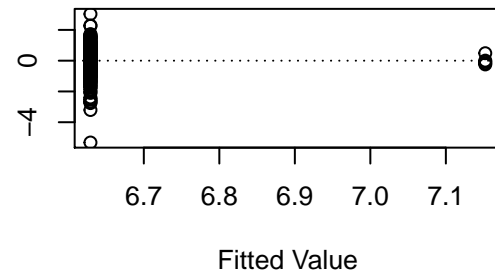
Standardized Residual Quantiles

**Normal Probability Plot**

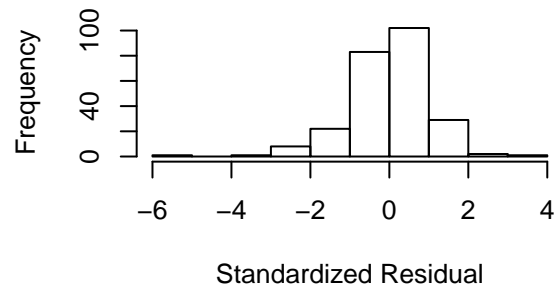


Standardized Residual

**Versus Fits**

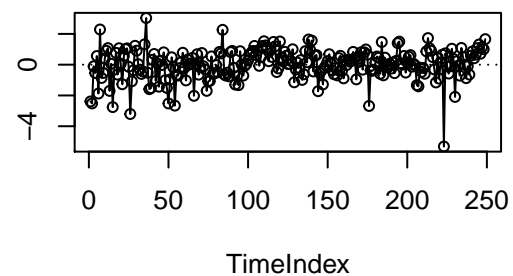


**Histogram**



Standardized Residual

**Versus Order**



```
##
## =====Cochrane-Orcutt test=====
##
## Call:
## lm(formula = testfit$residuals ~ Lag(testfit$residuals, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.76379 -0.26466  0.03335  0.32849  1.41596
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.00549    0.03148   0.174  0.86170
## Lag(testfit$residuals, 1) 0.17623    0.06237   2.826  0.00511 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4958 on 246 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.03144,    Adjusted R-squared:  0.0275
## F-statistic: 7.984 on 1 and 246 DF,  p-value: 0.005105
```

```

# additive model 1.5: try log of sales
bluestem$log_Sales <- log(bluestem$Sales)
bluestem_1 <- bluestem[,c(17,2:8,13)]
names(bluestem_1) # a subset of bluestem with "log_Sales", "Promotion" dummy, seasonal dummies from "Mo

## [1] "log_Sales" "Promotion" "Monday"      "Tuesday"      "Wednesday" "Thursday"
## [7] "Friday"      "Saturday"    "prolag1"

add.fit.5 <- lm(log_Sales ~ ., data = bluestem_1)
summary(add.fit.5)

##
## Call:
## lm(formula = log_Sales ~ ., data = bluestem_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.81094 -0.24174  0.01675  0.26429  1.42892
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.212139   0.219260  28.332 < 2e-16 ***
## Promotion1    0.340352   0.230334   1.478  0.141
## Monday1     -0.114131   0.233950  -0.488  0.626
## Tuesday1    -0.006467   0.232245  -0.028  0.978
## Wednesday1   0.087983   0.232560   0.378  0.706
## Thursday1    0.119954   0.232245   0.516  0.606
## Friday1      0.998658   0.232516   4.295 2.54e-05 ***
## Saturday1    1.276468   0.232644   5.487 1.04e-07 ***
## prolag11     -0.032282   0.230833  -0.140  0.889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4903 on 239 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.5657, Adjusted R-squared:  0.5512
## F-statistic: 38.92 on 8 and 239 DF, p-value: < 2.2e-16

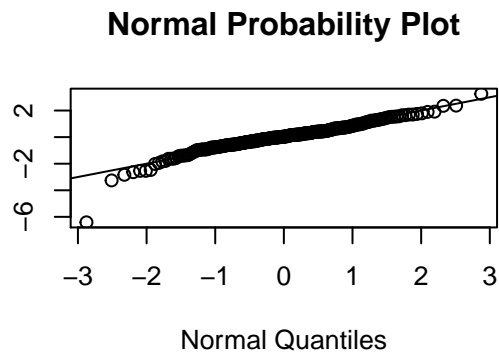
coef(summary(add.fit.5))[c(2,9),] #check promotion effect on sales as %

##              Estimate Std. Error  t value  Pr(>|t|)
## Promotion1  0.34035196  0.2303342  1.477644 0.1408196
## prolag11   -0.03228221  0.2308329 -0.139851 0.8888955

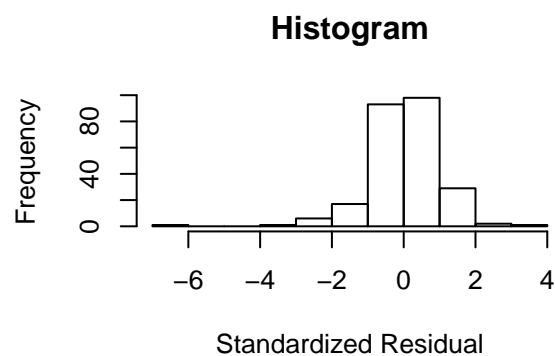
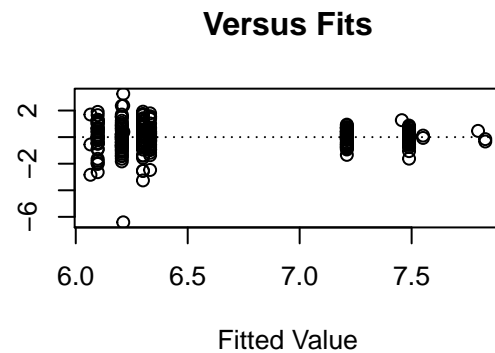
check_assumption(add.fit.5)

```

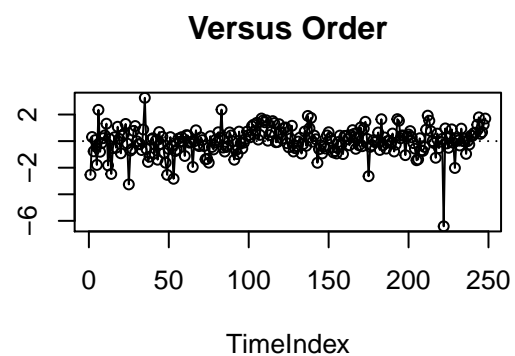
Standardized Residual Quantiles



Standardized Residual



Standardized Residual



```
##
## =====Breusch-Pagan test=====
##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 33.03497, Df = 1, p = 9.0516e-09
##
## =====Checking multicollinearity for independent variables=====
##
## VIF:
## Promotion    Monday    Tuesday Wednesday  Thursday    Friday    Saturday
## 1.081318  7.326714  7.679032  7.548387  7.679032  7.697005  7.855268
##   prolag1
## 1.086005
##
## =====Cochrane-Orcutt test=====
##
## Call:
## lm(formula = testfit$residuals ~ Lag(testfit$residuals, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8312 -0.2146  0.0166  0.2595  1.3468
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)          0.005577    0.029891    0.187  0.85214
## Lag(testfit$residuals, 1) 0.181488    0.062346    2.911  0.00393 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4698 on 245 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.03343,    Adjusted R-squared:  0.02949
## F-statistic: 8.474 on 1 and 245 DF,  p-value: 0.003935
```

### 3.3 Model Accuracy

```
## 2 models are selected in first round selection: add.fit.2 and mtp.fit.1
## L1 L2 percentage---- these three criteria are used to evaluate them

Error <- matrix(0,nrow=2,ncol=3)
colnames(Error) <- c("Mean Squared Error","Absolute Error","Absolute Error Percentage")
rownames(Error) <- c("add.fit.2","mtp.fit.1")

## add.fit.2
bluestem$prolag1[1]<-0
add.pd.2 <- predict(add.fit.2,bluestem)
dif <- bluestem$Sales-add.pd.2
Error[1,1] <- crossprod(dif,dif)/249
Error[1,2] <- sum(abs(dif))/249
Error[1,3] <- sum(abs(dif/bluestem$Sales))/249

## mtp.fit.1
mtp.pd.1 <- predict(mtp.fit.1,bluestem)*bluestem$Index
dif <- bluestem$Sales-mtp.pd.1
Error[2,1] <- crossprod(dif,dif)/249
Error[2,2] <- sum(abs(dif))/249
Error[2,3] <- sum(abs(dif/bluestem$Sales))/249

knitr::kable(Error,format = "markdown")
```

	Mean Squared Error	Absolute Error	Absolute Error Percentage
add.fit.2	112245.7	259.9391	0.5517314
mtp.fit.1	135555.3	285.8413	0.5322004

### 3.4 Conclusion

```
# selected additive model
summary(add.fit.2)

##
## Call:
## lm(formula = Sales ~ ., data = bluestem_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1021.78  -213.09   -45.67   198.63  1140.74
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    941.2      152.6   6.167 2.94e-09 ***
## Promotion1     533.9      160.3   3.330 0.00101 **
## Monday1       -457.7      162.8  -2.811 0.00535 **
## Tuesday1      -383.9      161.7  -2.375 0.01836 *
## Wednesday1    -316.6      161.9  -1.956 0.05165 .
## Thursday1     -329.0      161.7  -2.035 0.04296 *
## Friday1        461.2      161.8   2.850 0.00476 **
## Saturday1      895.9      161.9   5.533 8.27e-08 ***
## prolag11       326.3      160.7   2.031 0.04336 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 341.3 on 239 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.7192, Adjusted R-squared:  0.7098
## F-statistic: 76.53 on 8 and 239 DF,  p-value: < 2.2e-16
coef(summary(add.fit.2))[c(2,9),] # check promotion effect

##           Estimate Std. Error t value Pr(>|t|)
## Promotion1 533.8707    160.3261 3.329906 0.001006169
## prolag11   326.3359    160.6732 2.031054 0.043356439
# selected multiplicative model
summary(mtp.fit.1)

##
## Call:
## lm(formula = desSales ~ Promotion, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -799.52 -274.27  -68.99   218.01  2679.87
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    850.39      26.14  32.527 <2e-16 ***
## Promotion1     435.62     184.50   2.361  0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 408.4 on 247 degrees of freedom
## Multiple R-squared:  0.02207, Adjusted R-squared:  0.01811
## F-statistic: 5.575 on 1 and 247 DF,  p-value: 0.019
mtp.fit.1$coefficients[2] # check promotion effect on deseasonalized sales

## Promotion1
##    435.6229
index_table$Index[which(index_table$Weekday=="Saturday")] # standardized nightly index on Saturdays

## [1] 1.839744
mtp.fit.1$coefficients[2] * index_table$Index[which(index_table$Weekday=="Saturday")]

## Promotion1
```

```
## 801.4345
```

Note: The following part is not included in the report. It records our exploration of different methods to solve the autocorrelation problem including adding lagged predictors and logging the response variable, but these methods did not turn to be effective to eliminate the problem.

```
## additive model with timeindex
bluestem_1_t <- bluestem[,c(1:8,15)]
names(bluestem_1_t)

## [1] "Sales"      "Promotion" "Monday"    "Tuesday"   "Wednesday" "Thursday"
## [7] "Friday"     "Saturday"  "timeindex"

add.fit.1_t <- lm(Sales ~ ., data = bluestem_1_t)
summary(add.fit.1_t)
```

```
##
## Call:
## lm(formula = Sales ~ ., data = bluestem_1_t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1014.02  -211.05   -49.12   203.77  1169.43
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  899.5445   157.4615   5.713 3.28e-08 ***
## Promotion1    604.7302   158.8117   3.808 0.000178 ***
## Monday1     -444.3871   163.0727  -2.725 0.006902 **
## Tuesday1    -388.3039   162.6430  -2.387 0.017740 *
## Wednesday1  -319.6732   162.8411  -1.963 0.050790 .
## Thursday1   -331.9393   162.6192  -2.041 0.042324 *
## Friday1      454.5622   162.8138   2.792 0.005662 **
## Saturday1   902.4351   162.8330   5.542 7.85e-08 ***
## timeindex      0.3604     0.3033   1.188 0.235847
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 343.3 on 240 degrees of freedom
## Multiple R-squared:  0.7166, Adjusted R-squared:  0.7072
## F-statistic: 75.86 on 8 and 240 DF, p-value: < 2.2e-16
```

```
## add lag to additive model with timeindex
# add lag 1
bluestem_1_t <- bluestem[,c(1:8,15,13)]
names(bluestem_1_t)

## [1] "Sales"      "Promotion" "Monday"    "Tuesday"   "Wednesday"
## [6] "Thursday"   "Friday"    "Saturday"  "timeindex" "prolag1"

add.fit.2_t <- lm(Sales ~ ., data = bluestem_1_t)
summary(add.fit.2_t)
```

```
##
## Call:
## lm(formula = Sales ~ ., data = bluestem_1_t)
##
## Residuals:
```



```
##      Min      1Q   Median      3Q      Max
## -1000.65 -209.87   -42.65   202.72  1171.78
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   896.1280    156.3237   5.733 2.97e-08 ***
## Promotion1    543.1787    160.2851   3.389 0.000821 ***
## Monday1      -470.9206    162.3649  -2.900 0.004074 **
## Tuesday1     -388.6690    161.4593  -2.407 0.016833 *
## Wednesday1   -319.9251    161.6560  -1.979 0.048958 *
## Thursday1    -332.1839    161.4357  -2.058 0.040705 *
## Friday1       457.3042    161.6339   2.829 0.005062 **
## Saturday1     890.2363    161.7494   5.504 9.56e-08 ***
## timeindex      0.3900      0.3014   1.294 0.196946
## prolag11      341.6778    160.4975   2.129 0.034288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 340.8 on 239 degrees of freedom
## Multiple R-squared:  0.7219, Adjusted R-squared:  0.7114
## F-statistic: 68.93 on 9 and 239 DF,  p-value: < 2.2e-16
```

```
# add lag 1 and lag 2
bluestem_1_t <- bluestem[,c(1:8,15,13,14)]
names(bluestem_1_t)
```

```
## [1] "Sales"      "Promotion" "Monday"     "Tuesday"    "Wednesday"
## [6] "Thursday"   "Friday"     "Saturday"   "timeindex"  "prolag1"
## [11] "prolag2"
```

```
add.fit.3_t <- lm(Sales ~ ., data = bluestem_1_t)
summary(add.fit.3_t)
```

```
##
## Call:
## lm(formula = Sales ~ ., data = bluestem_1_t)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -1004.75 -208.60   -41.82   197.35  1166.92
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   903.1836    156.5366   5.770 2.48e-08 ***
## Promotion1    545.9020    160.5032   3.401 0.000788 ***
## Monday1      -466.0824    162.8872  -2.861 0.004597 **
## Tuesday1     -395.2491    162.0096  -2.440 0.015438 *
## Wednesday1   -308.5077    162.0131  -1.904 0.058098 .
## Thursday1    -331.6787    161.5442  -2.053 0.041158 *
## Friday1       457.7094    161.7432   2.830 0.005058 **
## Saturday1     891.9159    161.8617   5.510 9.35e-08 ***
## timeindex      0.3289      0.3058   1.076 0.283129
## prolag11      316.4133    163.4452   1.936 0.054075 .
## prolag21      100.2319    160.6369   0.624 0.533253
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 341 on 236 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.7215, Adjusted R-squared:  0.7097
## F-statistic: 61.13 on 10 and 236 DF,  p-value: < 2.2e-16

## multiplicative model with timeindex
mtp.fit.1_t <- lm(desSales ~ Promotion + timeindex, data = bluestem)
summary(mtp.fit.1_t)

##
## Call:
## lm(formula = desSales ~ Promotion + timeindex, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -848.33 -255.71  -60.79   205.55  2724.75
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  787.4759    52.1946   15.087 <2e-16 ***
## Promotion1   450.1433   184.4450    2.441  0.0154 *
## timeindex     0.5010     0.3599    1.392  0.1652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 407.6 on 246 degrees of freedom
## Multiple R-squared:  0.02971, Adjusted R-squared:  0.02183
## F-statistic: 3.767 on 2 and 246 DF,  p-value: 0.02447

## add lag to multiplicative model with timeindex
# add lag 1
mtp.fit.2_t <- lm(desSales ~ Promotion + timeindex + prolag1, data = bluestem)
summary(mtp.fit.2)

##
## Call:
## lm(formula = desSales ~ Promotion + prolag1, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -799.83 -274.51  -65.88   206.38  2679.57
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   850.70     26.33   32.307 <2e-16 ***
## Promotion1    407.73    187.35    2.176  0.0305 *
## prolag1       137.95    187.35    0.736  0.4623
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 407.6 on 245 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.02418, Adjusted R-squared:  0.01621
## F-statistic: 3.036 on 2 and 245 DF,  p-value: 0.04986

```

```
# add lag 1 and lag 2
mtp.fit.3_t <- lm(desSales ~ Promotion + timeindex + prolag1 + prolag2, data = bluestem)
summary(mtp.fit.3)
```

```
##
## Call:
## lm(formula = desSales ~ Promotion + prolag1 + prolag2, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -799.30 -274.01  -68.77   218.37 2680.09
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   850.17      26.51   32.069  <2e-16 ***
## Promotion1    415.82     187.10    2.222   0.0272 *
## prolag1       100.12     190.29    0.526   0.5993
## prolag2       183.66     187.10    0.982   0.3273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 406.4 on 243 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.02794,    Adjusted R-squared:  0.01594
## F-statistic: 2.328 on 3 and 243 DF,  p-value: 0.07514
```

```
## try log(desSales)
mtp.fit.4_t <- lm(log_desSales ~ Promotion + timeindex, data = bluestem)
summary(mtp.fit.4_t)
```

```
##
## Call:
## lm(formula = log_desSales ~ Promotion + timeindex, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.80156 -0.25066  0.03369  0.32535  1.63309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.4985339  0.0645538 100.668  <2e-16 ***
## Promotion1   0.5530291  0.2281199   2.424   0.0161 *
## timeindex    0.0010417  0.0004452   2.340   0.0201 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5041 on 246 degrees of freedom
## Multiple R-squared:  0.04185,    Adjusted R-squared:  0.03406
## F-statistic: 5.372 on 2 and 246 DF,  p-value: 0.005203
```

```
mtp.fit.5_t <- lm(log_desSales ~ Promotion + timeindex + prolag1, data = bluestem)
summary(mtp.fit.5_t)
```

```
##
## Call:
```

```

## lm(formula = log_desSales ~ Promotion + timeindex + prolag1,
##     data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.80154 -0.25062  0.03373  0.32541  1.63315
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.4984708  0.0649949  99.984  <2e-16 ***
## Promotion1    0.5526096  0.2324274   2.378   0.0182 *
## timeindex     0.0010419  0.0004465   2.333   0.0204 *
## prolag11      0.0023163  0.2324016   0.010   0.9921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5052 on 245 degrees of freedom
## Multiple R-squared:  0.04185,    Adjusted R-squared:  0.03012
## F-statistic: 3.567 on 3 and 245 DF,  p-value: 0.01479

mtp.fit.6_t <- lm(log_desSales ~ Promotion + timeindex + prolag1 + prolag2, data = bluestem)
summary(mtp.fit.6_t)

##
## Call:
## lm(formula = log_desSales ~ Promotion + timeindex + prolag1 +
##     prolag2, data = bluestem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.78700 -0.24422  0.01996  0.33396  1.61067
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.5280809  0.0652477 100.051  <2e-16 ***
## Promotion1    0.5555123  0.2287133   2.429   0.0159 *
## timeindex     0.0008439  0.0004446   1.898   0.0589 .
## prolag11     -0.0646126  0.2324547  -0.278   0.7813
## prolag21      0.2803163  0.2286684   1.226   0.2214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4962 on 242 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.04021,    Adjusted R-squared:  0.02434
## F-statistic: 2.534 on 4 and 242 DF,  p-value: 0.0409

```