Homework 10

Code

6210422036: Tanat Iempreedee

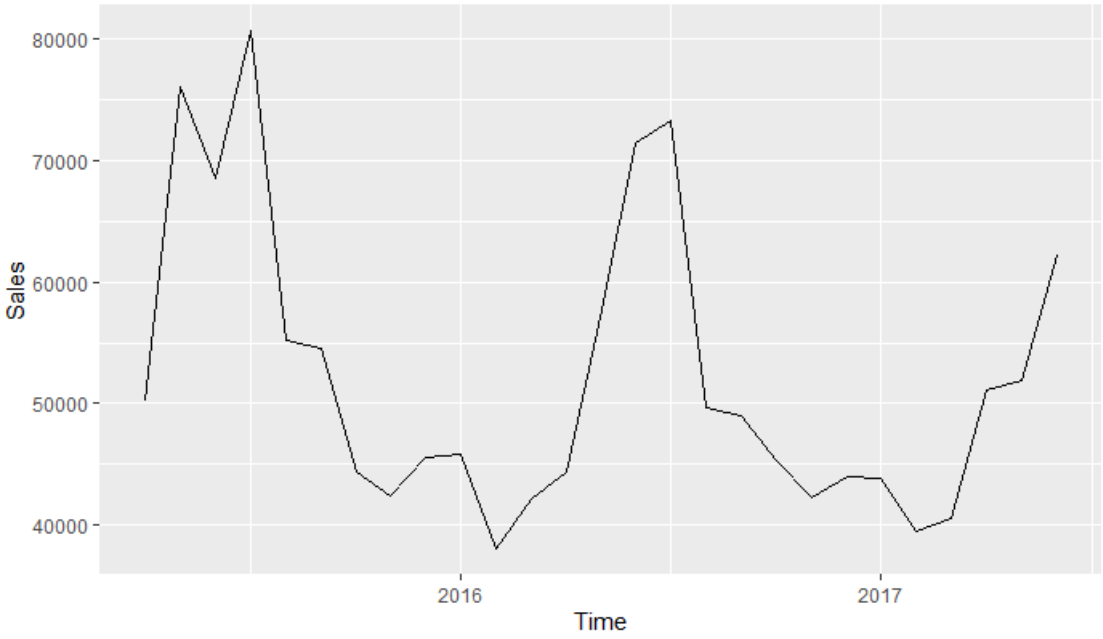
Hide

```
library(fpp2)
library(forecast)


d <- read.table('c3t13.txt', header=TRUE,
                colClasses = c('numeric','factor'))
str(d)
```

```
'data.frame':   27 obs. of  2 variables:
 $ Mustard: num  50137 76030 68590 80681 55228 ...
 $ Event  : Factor w/ 3 levels "1","2","3": 3 3 2 1 1 1 1 1 2 1 ...
```

Hide

```
dts <- ts(data=d, frequency=12, start=c(2015,4))
autoplot(dts[,'Mustard']) +
  ylab('Sales')
```



Prepare data to forecast

Hide

```
newdata <- data.frame(trend=28:33,
                      season=factor(7:12, levels=1:12),
                      fevent=factor(rep(1,6), levels=1:3))
newdata
```

trend <int>	season <fctr>	fevent <fctr>
28	7	1
29	8	1
30	9	1
31	10	1
32	11	1
33	12	1

6 rows

Model 1: Mustard ~ trend

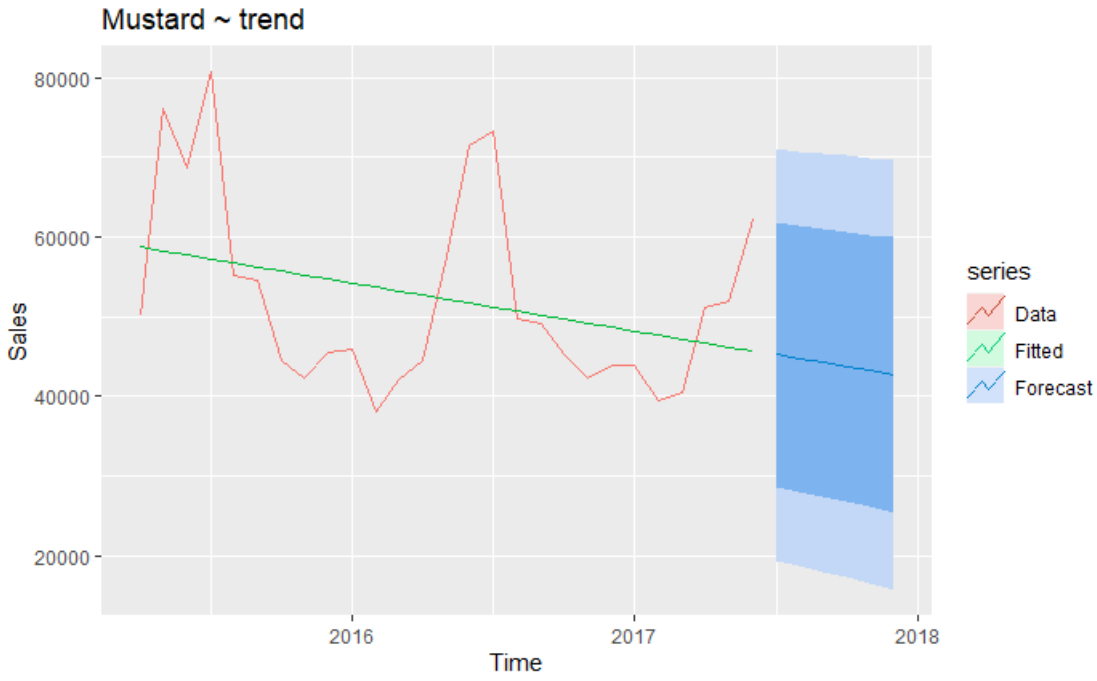
Hide

```
# Model 1
reg1 <- tslm(Mustard ~ trend, data=dts)
summary(reg1)
```

```

Call:
tslm(formula = Mustard ~ trend, data = dts)

Residuals:
   Min     1Q Median     3Q    Max
-15662  -8347  -4418   5591  23466

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  59219.8     4622.5  12.811 1.74e-12 ***
trend         -501.1      288.5  -1.737   0.0947 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11680 on 25 degrees of freedom
Multiple R-squared:  0.1077,    Adjusted R-squared:  0.07197
F-statistic: 3.016 on 1 and 25 DF,  p-value: 0.09474
```

Hide

```
fc <- forecast(reg1, newdata=newdata)
autoplot(dts[,'Mustard'], series='Data') +
  autolayer(fitted(reg1), series='Fitted') +
  autolayer(fc, series='Forecast') +
  ylab('Sales') +
  ggtitle('Mustard ~ trend')
```



Model 2: Mustard ~ trend + season + Event

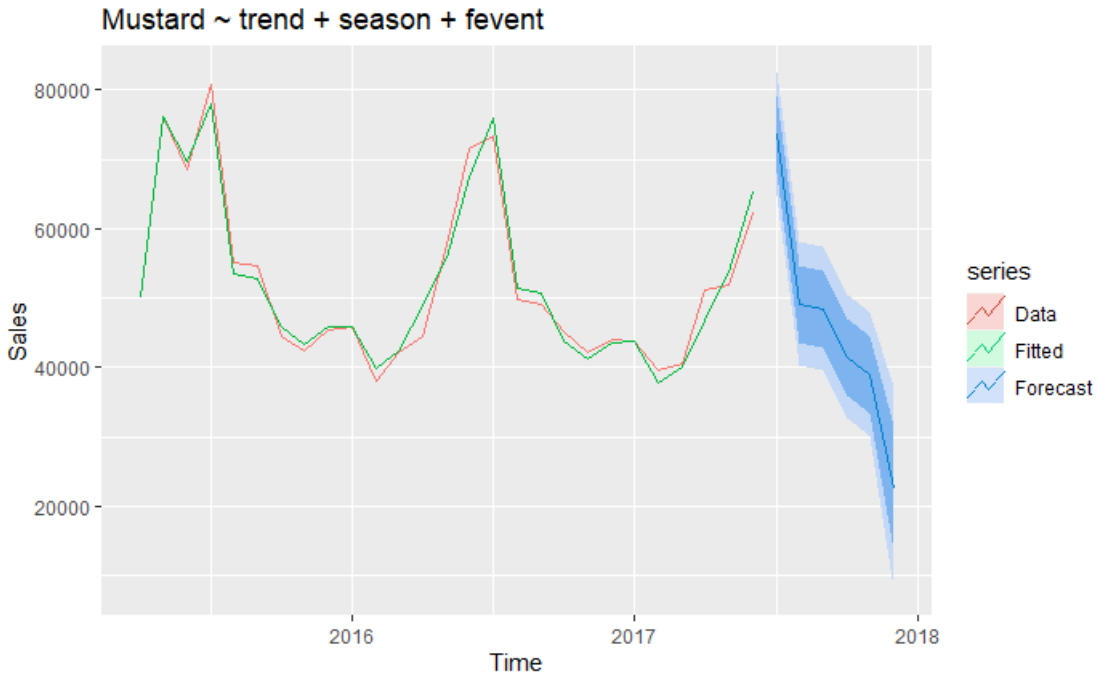
Hide

```
# Model 2
fevent <- factor(dts[,'Event'], levels=1:3)

reg2 <- tslm(Mustard ~ trend + season + fevent, data=dts)
summary(reg2)
```

```
Call:
tslm(formula = Mustard ~ trend + season + fevent, data = dts)

Residuals:
   Min     1Q Median     3Q    Max
 -4476  -1603      0   1603   4476

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  47731.7     2609.7  18.290 3.95e-10 ***
trend         -184.0       92.5  -1.989 0.069991 .
season2      -5819.6     3041.2  -1.914 0.079822 .
season3      -3118.9     3045.4  -1.024 0.325980
season4     -15340.0     5272.4  -2.910 0.013093 *
season5      10736.6     3062.2   3.506 0.004332 **
season6       3560.1     5952.3   0.598 0.560884
season7      31083.2     3090.0  10.059 3.36e-07 ***
season8       6753.6     3074.8   2.196 0.048443 *
season9       6275.0     3062.2   2.049 0.062966 .
season10      -516.3     3052.4  -0.169 0.868491
season11     -2882.4     3045.4  -0.946 0.362580
season12    -19130.1     6080.3  -3.146 0.008434 **
fevent2      18898.4     5265.1   3.589 0.003718 **
fevent3      17929.5     4078.3   4.396 0.000871 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3040 on 12 degrees of freedom
Multiple R-squared:  0.971, Adjusted R-squared:  0.9371
F-statistic: 28.67 on 14 and 12 DF,  p-value: 4.75e-07
```

Hide

```
fc <- forecast(reg2, newdata=newdata)
autoplot(dts[,'Mustard'], series='Data') +
  autolayer(fitted(reg2), series='Fitted') +
  autolayer(fc, series='Forecast') +
  ylab('Sales') +
  ggtitle('Mustard ~ trend + season + fevent')
```



Model 3: Mustard ~ trend + season

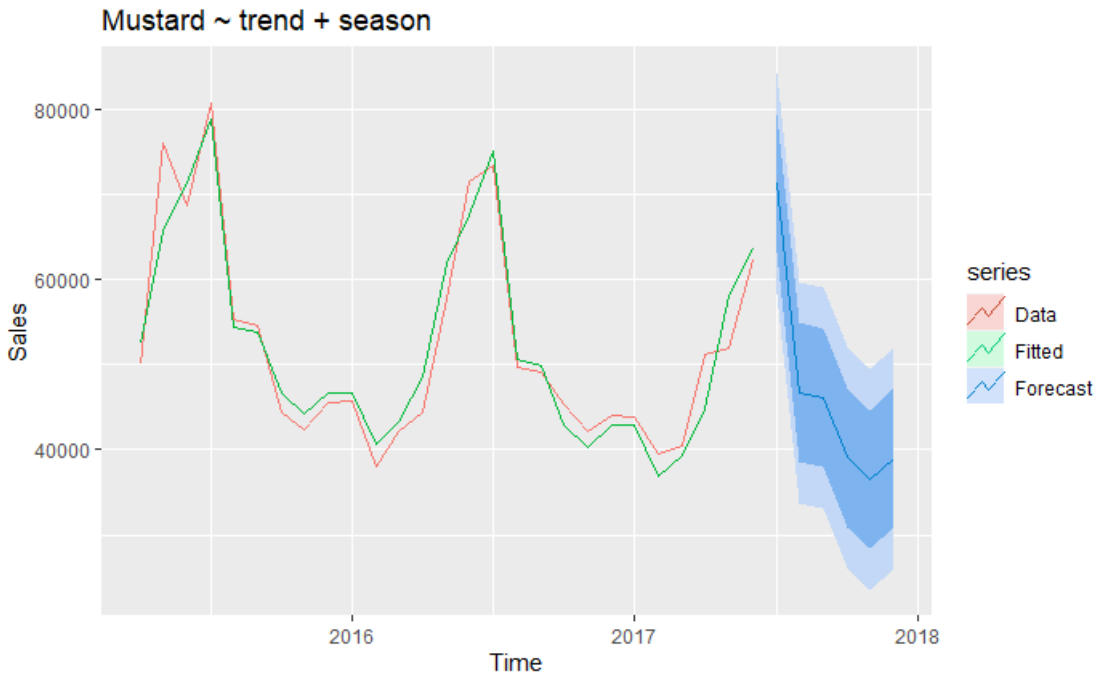
Hide

```
reg3 <- tslm(Mustard ~ trend + season, data=dts)
summary(reg3)
```

```

Call:
tslm(formula = Mustard ~ trend + season, data = dts)

Residuals:
    Min      1Q  Median      3Q     Max 
-6130.2 -2054.7  -866.3  1466.9 10337.7 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  49885.4     3777.2  13.207 2.71e-09 ***
trend         -318.6      118.7  -2.683 0.017834 *  
season2      -5685.0     4618.5  -1.231 0.238628    
season3      -2849.7     4623.0  -0.616 0.547520    
season4       2831.6     4229.7   0.669 0.514100    
season5      16443.9     4221.3   3.895 0.001616 ** 
season6      22323.9     4216.3   5.295 0.000113 ***
season7      30275.5     4671.6   6.481 1.45e-05 ***
season8       6080.6     4654.9   1.306 0.212518    
season9       5736.5     4641.3   1.236 0.236804    
season10      -920.2     4630.7  -0.199 0.845345    
season11     -3151.6     4623.0  -0.682 0.506540    
season12      -366.3     4618.5  -0.079 0.937906    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4617 on 14 degrees of freedom
Multiple R-squared:  0.9219,    Adjusted R-squared:  0.8549 
F-statistic: 13.77 on 12 and 14 DF,  p-value: 9.874e-06
```

Hide

```
fc <- forecast(reg3, newdata=newdata)
autoplot(dts[,'Mustard'], series='Data') +
  autolayer(fitted(reg3), series='Fitted') +
  autolayer(fc, series='Forecast') +
  ylab('Sales') +
  ggtitle('Mustard ~ trend + season')
```



Model 4: Mustard ~ trend + Event

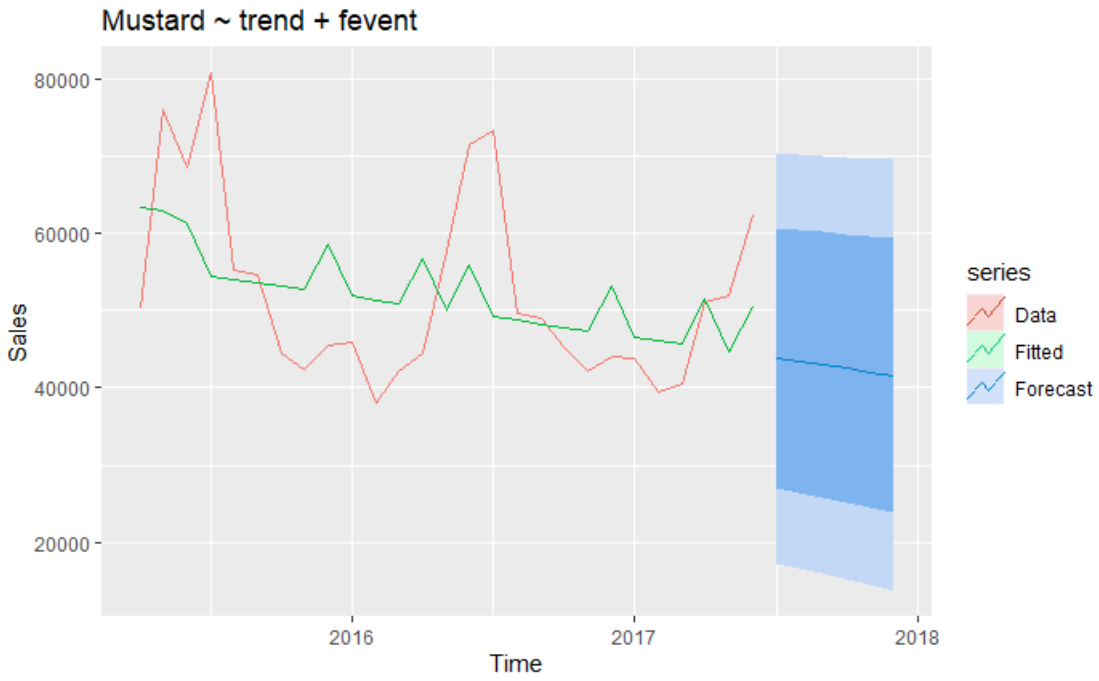
Hide

```
reg4 <- tslm(Mustard ~ trend + fevent, data=dts)
summary(reg4)
```

```

Call:
tslm(formula = Mustard ~ trend + fevent, data = dts)

Residuals:
   Min     1Q Median     3Q    Max
-13310  -8766  -2542   7319  26219

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  56237.2     5494.6  10.235 4.92e-10 ***
trend         -443.8      326.4  -1.360    0.187
fevent2       6264.5     5242.8   1.195    0.244
fevent3       7512.1     9714.4   0.773    0.447
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11710 on 23 degrees of freedom
Multiple R-squared:  0.1741,    Adjusted R-squared:  0.06637
F-statistic: 1.616 on 3 and 23 DF,  p-value: 0.213
```

Hide

```
fc <- forecast(reg4, newdata=newdata)
autoplot(dts[,'Mustard'], series='Data') +
  autolayer(fitted(reg4), series='Fitted') +
  autolayer(fc, series='Forecast') +
  ylab('Sales') +
  ggtitle('Mustard ~ trend + fevent')
```



Compare AIC

Based on the AIC, Model 2 is selected.

Hide

```
data.frame(MD1=AIC(reg1),MD2=AIC(reg2),MD3=AIC(reg3),MD4=AIC(reg4))
```

MD1 <dbl>	MD2 <dbl>	MD3 <dbl>	MD4 <dbl>
586.2768	519.7829	542.5138	588.1878

1 row

Forecast

Hide

```
fc <- forecast(reg2, newdata=newdata)
fc
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Jul 2017	73663.48	68132.42	79194.55	64777.626	82549.34
Aug 2017	49149.93	43618.86	54681.00	40264.071	58035.79
Sep 2017	48487.30	42956.24	54018.37	39601.446	57373.16
Oct 2017	41512.02	35980.95	47043.09	32626.161	50397.88
Nov 2017	38962.00	33430.93	44493.07	30076.141	47847.86
Dec 2017	22530.34	13498.15	31562.54	8019.798	37040.89

6 rows

Question

If we want to optimize the campaign spending, what are other possible independent variables? Indicate whether they are numerical or categorical.

- Campaign Spending (numerical) plan for each month
- A flag indicate Campaign sensitivity (categorical)?
- If $Profit = Sales - Qty(UnitCost) - CampaignSpending - \ldots$, Unit cost could affect finding an optimum campaign spending as well, we may want to use costs as well, e.g., estimated raw material cost (numerical) for each month?