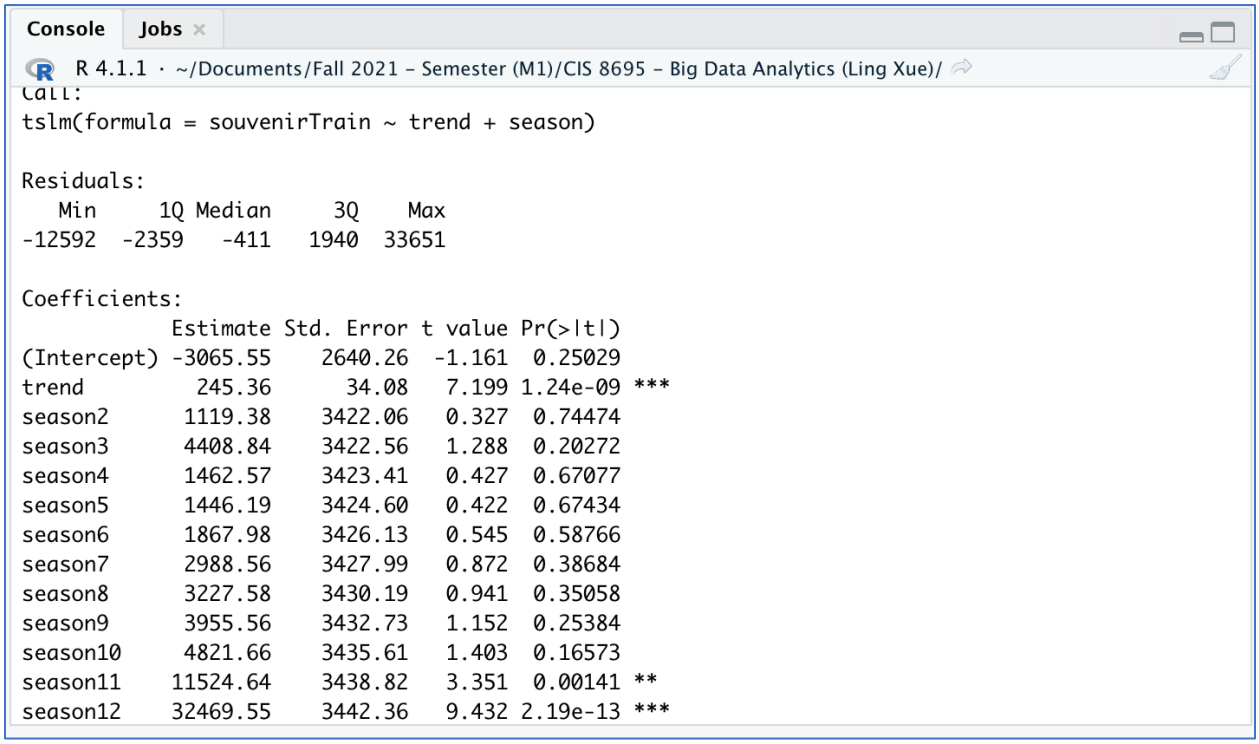
Problem-1: Time Series (Souvenir Sales)

17.6 b. Run a regression model with Sales (in Australian dollars) as the outcome variable and with a linear trend and monthly seasonality. Remember to fit only the training period. Call this model A.
i. Examine the coefficients: Which month tends to have the highest average sales during the year? Why is this reasonable?
ii. The estimated trend coefficient in model A is 245.36. What does this mean?

Ans 17.6 b i: December has the highest average sales during the year, followed by November. This seems reasonable as sales increase near the holiday season, when people purchase more. Also, Decembers in Australia are generally warmer, and it may be assumed that tourists are flocking that time of the year.
Ans 17.6 b ii. The estimated trend coefficient (245.36) in model A indicates "average linear increase" over each month, holding other variables constant.



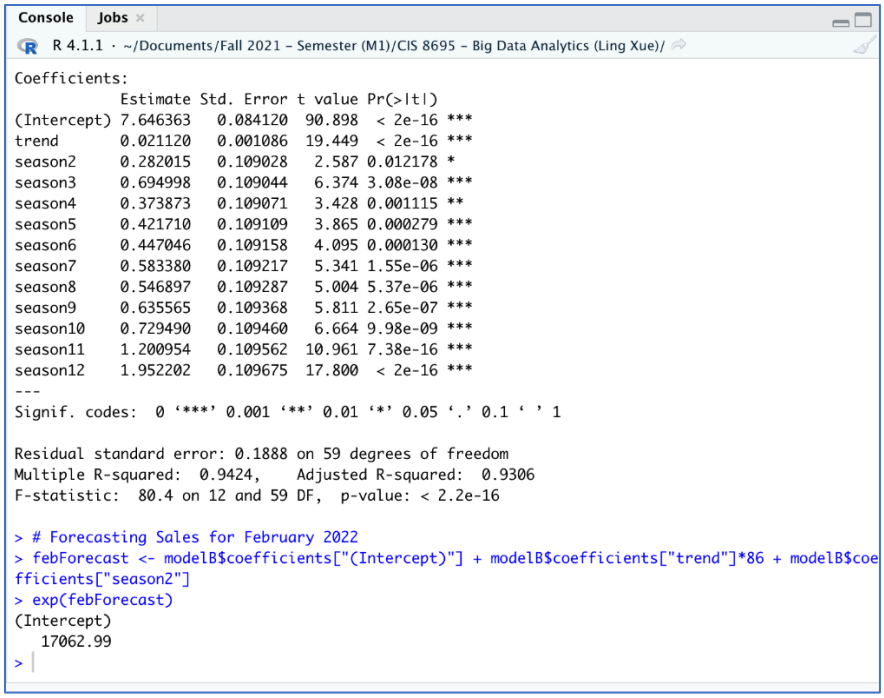

```
Call:
tslm(formula = souvenirTrain ~ trend + season)

Residuals:
   Min     1Q Median     3Q    Max
-12592  -2359   -411   1940  33651

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3065.55    2640.26  -1.161  0.25029
trend         245.36      34.08   7.199 1.24e-09 ***
season2      1119.38    3422.06   0.327  0.74474
season3      4408.84    3422.56   1.288  0.20272
season4      1462.57    3423.41   0.427  0.67077
season5      1446.19    3424.60   0.422  0.67434
season6      1867.98    3426.13   0.545  0.58766
season7      2988.56    3427.99   0.872  0.38684
season8      3227.58    3430.19   0.941  0.35058
season9      3955.56    3432.73   1.152  0.25384
season10     4821.66    3435.61   1.403  0.16573
season11    11524.64    3438.82   3.351  0.00141 **
season12    32469.55    3442.36   9.432 2.19e-13 ***
```

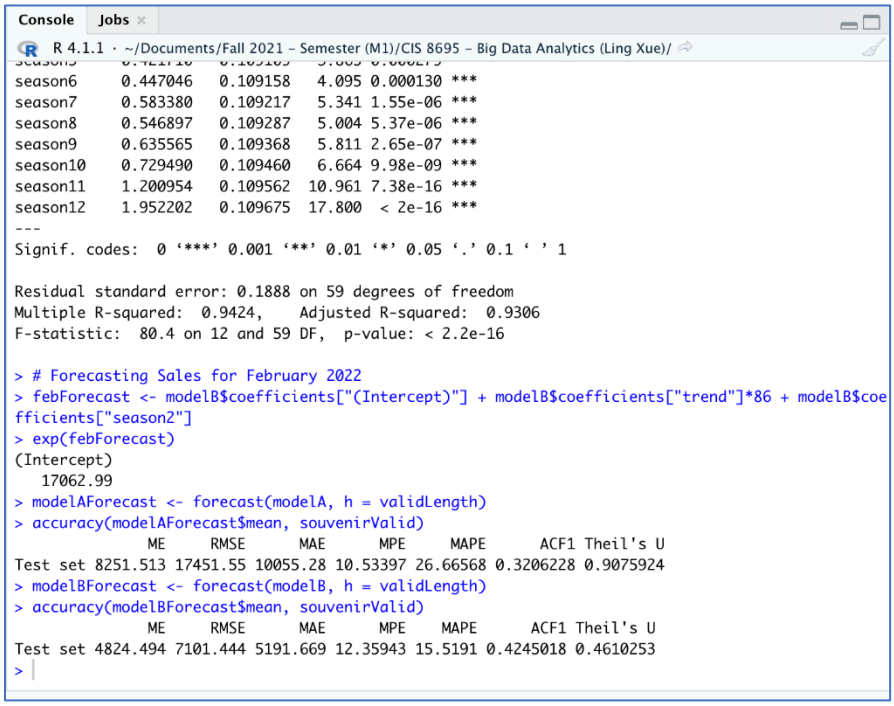
c. Run a regression model with an exponential trend and multiplicative seasonality. Remember to fit only training data. Call this model B.
i. Fitting a model to log(Sales) with a linear trend is equivalent to fitting a model to Sales (in dollars) with what type of trend?
ii. The estimated trend coefficient in model B is 0.02 What does this mean?
iii. Use this model to forecast the sales in February 2002.

Ans 17.6 c i: Fitting a model to log(Sales) with linear trend is equivalent to fitting a model to Sales (in dollars) with "exponential trend".
Ans 17.6 c ii: The estimated trend coefficient (0.02) in model B indicates an overall increase "~2.1% each month" for every successive time period.
Ans 17.6 c iii. The forecasted sales in February 2002 using this model are "$17062.99".



d. Compare the two regression models (A and B) in terms of forecast performance. Which model is preferable for forecasting? Mention at least two reasons based on the information in the outputs.

Ans 17.6 d: When comparing both models in terms of forecast performance, "model B" is preferrable as it appears to be a better fit. Looking at accuracy, we can find that MAPE (Mean Absolute Percentage Error) for model B is 15.5191 whereas model A is 26.6656. When we compare how well they explain data, model B has adjusted R-Squared value of "0.9306" (93%) while model A has "0.7476" (75%).



Sample Code (Problem-1: Time Series)

P1-Time Series.txt

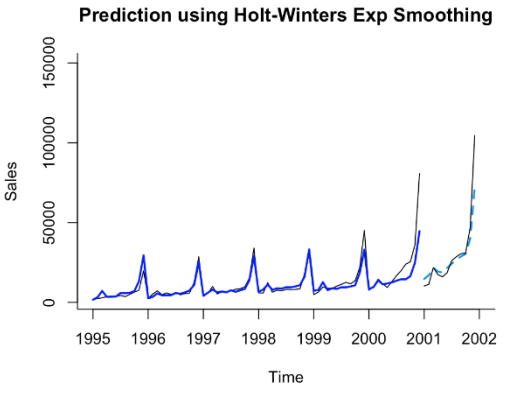
Problem-2: Time Series (Souvenir Sales)

Holt Winter Exponential Smoothing

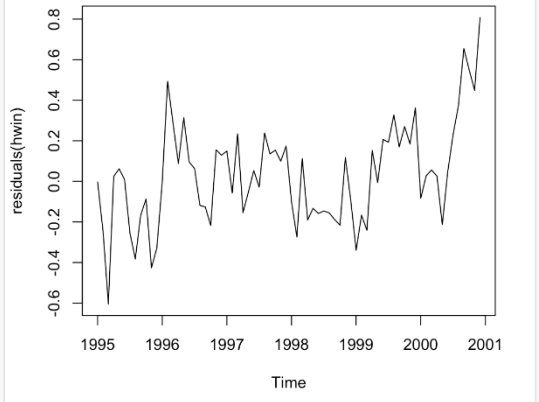
After applying the Holt Winter Exponential Smoothing method, we can see the predicted values of given data, forecasted values for 2001 and 2002 are displayed. The values are calculated at 80% and 95% confidence intervals. We can also see the residuals plot on and don't see any pattern in it, indicating forecasted values are right.

```
> #install.packages("zoo")
> library("zoo")
> # Running Holt-Winters Exponential Smoothing
> hwin <- ets(souvenirTrain, model = "MAA")
> # Creating the Predictions
> hwin.pred <- forecast(hwin, h = validLength, level = 0)
> # Plotting the Series
> plot(hwin.pred, ylim = c(1600, 150000),  ylab = "Sales", xlab = "Time",
+      bty = "l", xaxt = "n", xlim = c(1995,2002), main = "Prediction using Holt-Winters Exp Smoothing", flty = 2)
> axis(1, at = seq(1995, 2002, 1), labels = format(seq(1995, 2002, 1)))
> lines(hwin.pred$fitted, lwd = 2, col = "blue")
> lines(souvenirValid)
> accuracy(hwin.pred, souvenirValid)
                    ME      RMSE       MAE       MPE     MAPE      MASE      ACF1 Theil's U
Training set 1021.335 5398.946 2629.436 -3.392261 20.68743 0.7021564 0.3822527        NA
Test set     2317.689 9893.032 5255.931 -5.786291 17.06143 1.4035277 0.2352126 0.5571445
> hwin
ETS(M,A,A)

Call:
 ets(y = souvenirTrain, model = "MAA")
```




```
> forecast(hwin,24)
         Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
Jan 2001       14695.05  9285.564  20104.53  6421.957  22968.14
Feb 2001       16882.21 10660.783  23103.63  7367.360  26397.05
Mar 2001       21708.34 13703.779  29712.90  9466.421  33950.25
Apr 2001       19291.92 12159.303  26424.55  8383.522  30200.33
May 2001       18725.19 11783.715  25666.67  8109.118  29341.27
Jun 2001       21501.85 13528.253  29475.45  9307.285  33696.42
Jul 2001       24292.81 15279.592  33306.03 10508.283  38077.34
Aug 2001       26576.42 16706.268  36446.56 11481.328  41671.51
Sep 2001       28786.38 18083.634  39489.13 12417.943  45154.82
Oct 2001       30312.77 19023.502  41602.04 13047.323  47578.22
Nov 2001       39741.98 24985.201  54498.75 17173.435  62310.52
Dec 2001       72329.89 45612.174  99047.60 31468.671 113191.10
Jan 2002       21660.72 12514.995  30806.44  7673.543  35647.89
Feb 2002       23847.87 13725.073  33970.68  8366.386  39329.36
Mar 2002       28674.00 16465.211  40882.80 10002.266  47345.74
Apr 2002       26257.59 14980.602  37534.58  9010.924  43504.26
May 2002       25690.86 14562.225  36819.49  8671.081  42710.64
Jun 2002       28467.52 16110.811  40824.22  9569.565  47365.47
Jul 2002       31258.48 17660.954  44856.00 10462.859  52054.10
Aug 2002       33542.08 18908.402  48175.77 11161.799  55922.37
Sep 2002       35752.05 20108.664  51395.43 11827.558  59676.54
Oct 2002       37278.44 20905.209  53651.67 12237.745  62319.13
Nov 2002       46707.64 26294.721  67120.57 15488.772  77926.52
Dec 2002       79295.55 44846.772 113744.34 26610.689 131980.42
> plot(residuals(hwin))
>
```



Sample Code (Problem-2: Time Series)

P2-Time Series.txt