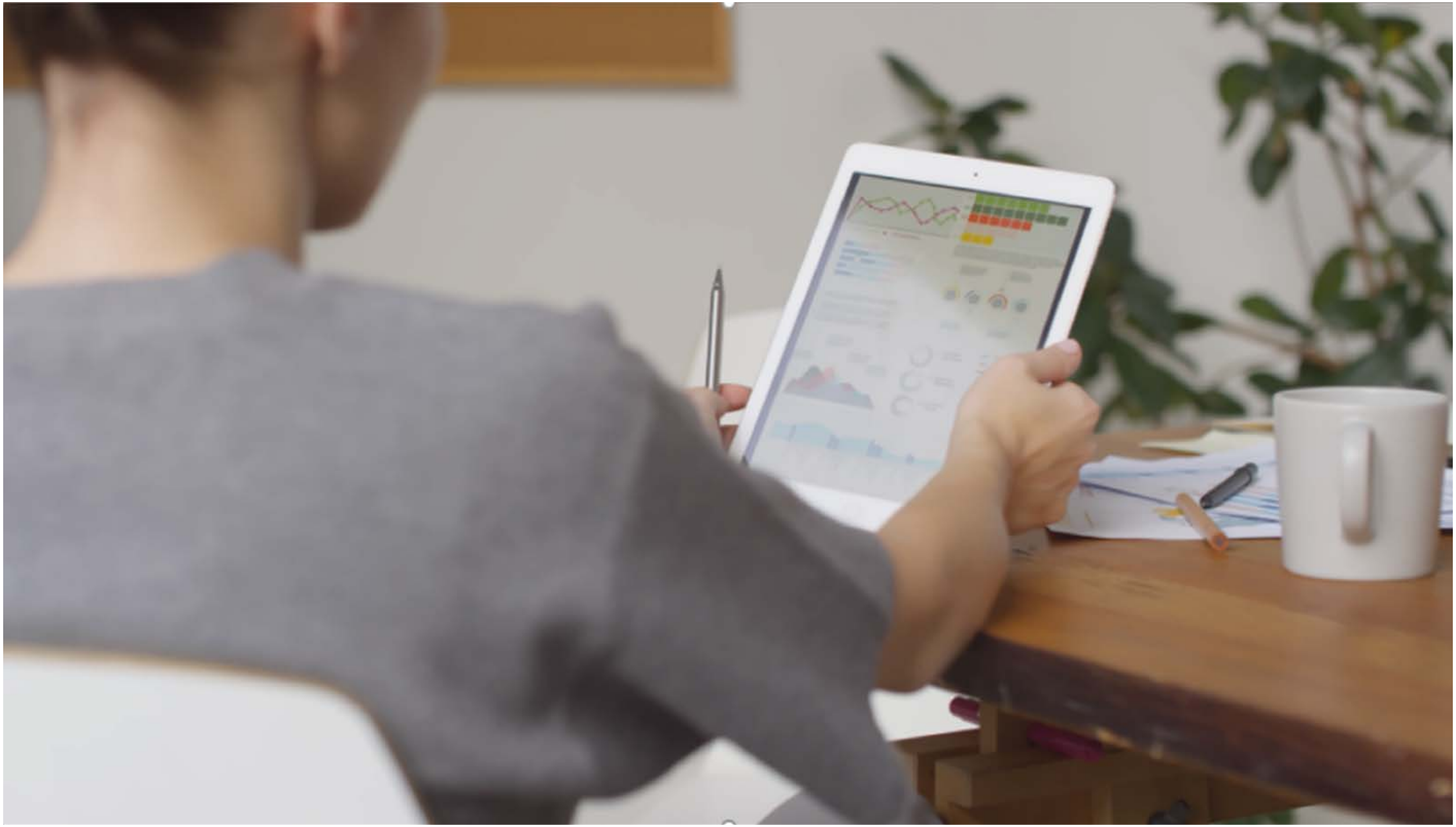


TBA2102 2020/2021 Semester 2
Tutorial 8: Time series forecasting

STRUCTURE OF TUTORIAL

Duration:

45 mins

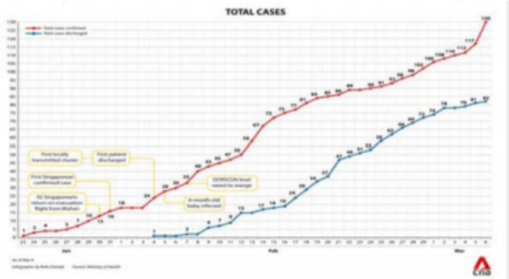
Content:

- Time series concepts
- Tutorial 8 (Questions 1 & 2)

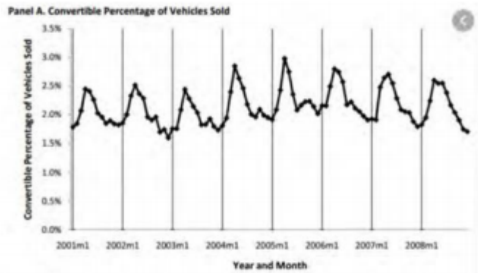
Time Series Concepts

KEY CONCEPTS

A gradual upward or downward movement of a time series over time.

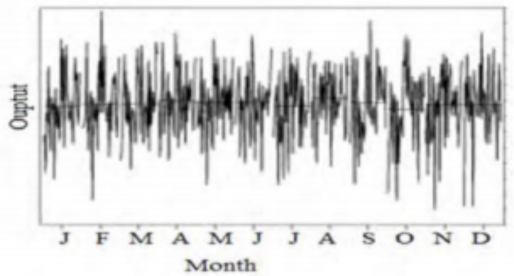


An effect that occurs/repeats at a fixed time interval (e.g. day, week, monthly, year)

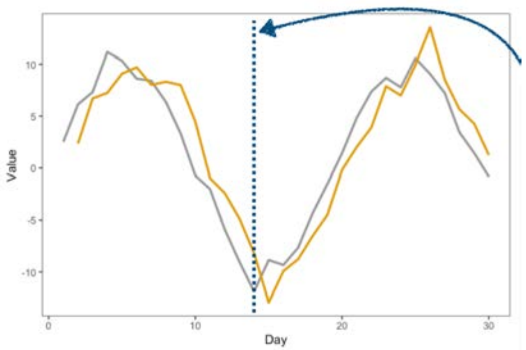


Ups & downs over much longer time frame that do not have a fixed interval/length

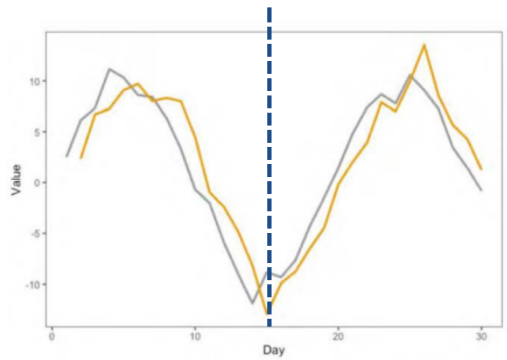
Statistical properties of the time series (e.g. mean, variance) do not change over time.



If changes in X_t precedes changes in Y_t.



A variable that follows movements of other variables.



Tutorial 8 Question 1

TUTORIAL 8 QUESTION 1

Dataset required: T8hdb.csv

- Note: This dataset comes from publically available data from the Singapore Department of Statistics, or SingStat. https://data.gov.sg/dataset/hdb-resale-price-index

- First, load in the dataset for this question. There is only one variable, which is the average HDB resale price index. Q1 (or first quarter) of 2009 is set as the "base" period, and thus has by definition an index value of 100. The index values of the rest of the years are relative to this base value.

- The code below creates a "train" and "test" dataset. The "train" dataset comprises of data from all years except 2018 and 2019. The "test" data then contains data in Years 2018 and 2019, to test the predictions of our model. This means we can fit the model using the "train" dataset, and then once we have the fitted model, we test the fitted model with the "test" dataset to see how well our fitted model perform against the real data (in 2018 and 2019).

TUTORIAL 8 QUESTION 1

- The code below creates a “train” and “test” dataset.

```
hdb_wide = read.csv('T8hdb.csv', header=T, na.strings = "NA")
```

```
# removing unused columns
hdb_wide=hdb_wide[,2:119]

# convert to a `ts` object:
hdb_ts = ts(unlist(hdb_wide[1,1:ncol(hdb_wide)], use.names=F), frequency=4, start = c(1990, 1))

# We also create a long form data frame. You can try understanding what each step in this code does by running each line sep
arately (without the last %>%), and inspecting the resulting file using head(hdb_long)
hdb_long <- hdb_wide %>%
  # gather() converts wide-form to long-form.
  gather(key="YearQuarter", value="PriceIndex") %>%
  # remove "X"
  mutate_at("YearQuarter", function(x) {sub(pattern="X", replacement="", x)}) %>%
  # split "YearQuarter" into a "Year" variable and a "Quarter" variable
  # and make a variable called "TimeIndex" that just goes 1, 2, 3, 4...
  mutate( Year = as.numeric(substr(YearQuarter, start=1, stop=4)),
          Quarter = substr(YearQuarter, start=6, stop=7),
          TimeIndex = 1:length(YearQuarter)) %>%
  # Rearrange the columns in a nicer order
  select("TimeIndex", "YearQuarter", "Year", "Quarter", "PriceIndex")
```

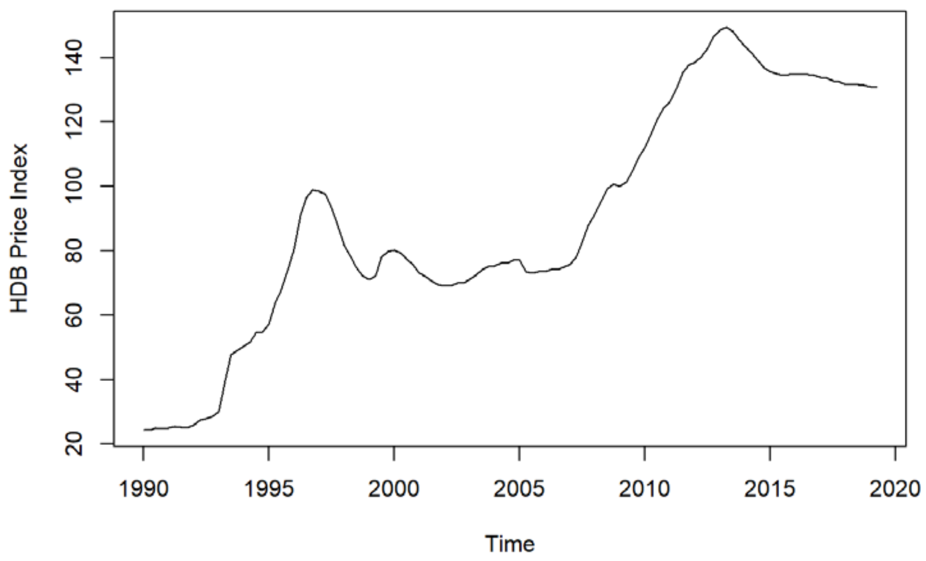
```
hdb_test = hdb_long[113:118,] # exclude values in 2018 and 2019 for testing later
hdb_train = hdb_long[1:112,] # keeping values up to and including 2016
```

```
hdbtest_ts<- window(hdb_ts, start=2018)
hdbtrain_ts<- window(hdb_ts,start=1990,end=2017)
```

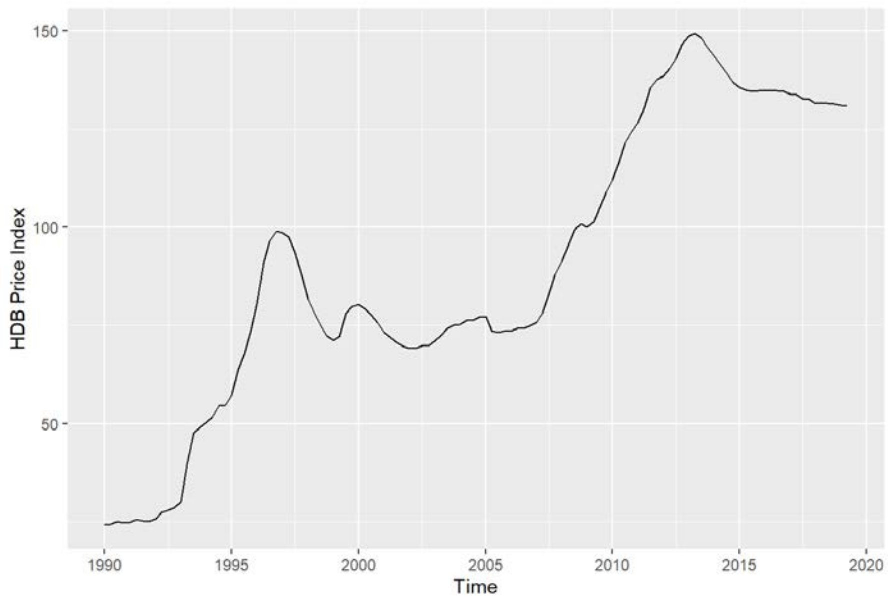
QUESTION 1A

- First, plot the data. There is only one variable, so just plot this against time on the horizontal axis.

```r
plot(hdb_ts,ylab="HDB Price Index")
```




```r
# or can use functions in "Forecast" package
autoplot(hdb_ts)+ylab("HDB Price Index")
```



How would you describe the time series?

- Stationary? Trend? Seasons? Cycles?

There seems to be an increasing trend but no seasonal effects observed in the HDB price index.

QUESTION 1B

Using the Simple Moving Average model, where:

$$\widehat{Y}_{t+1} = \frac{1}{k}\left(Y_t + Y_{t-1} + \cdots + Y_{t-(k-1)}\right)$$

i

Forecast the HDB Price Index during the observed period.

Calculate one forecast with a window size of 4 periods (1 year) and assign it to "hdb_long$SMApred4".

Next calculate a second forecast with a window size of 16 periods (4 years) and assign it to "hdb_long$SMApred16".

ii

Plot these two forecasts (and the actual data) on the same plot. Discuss what you see.

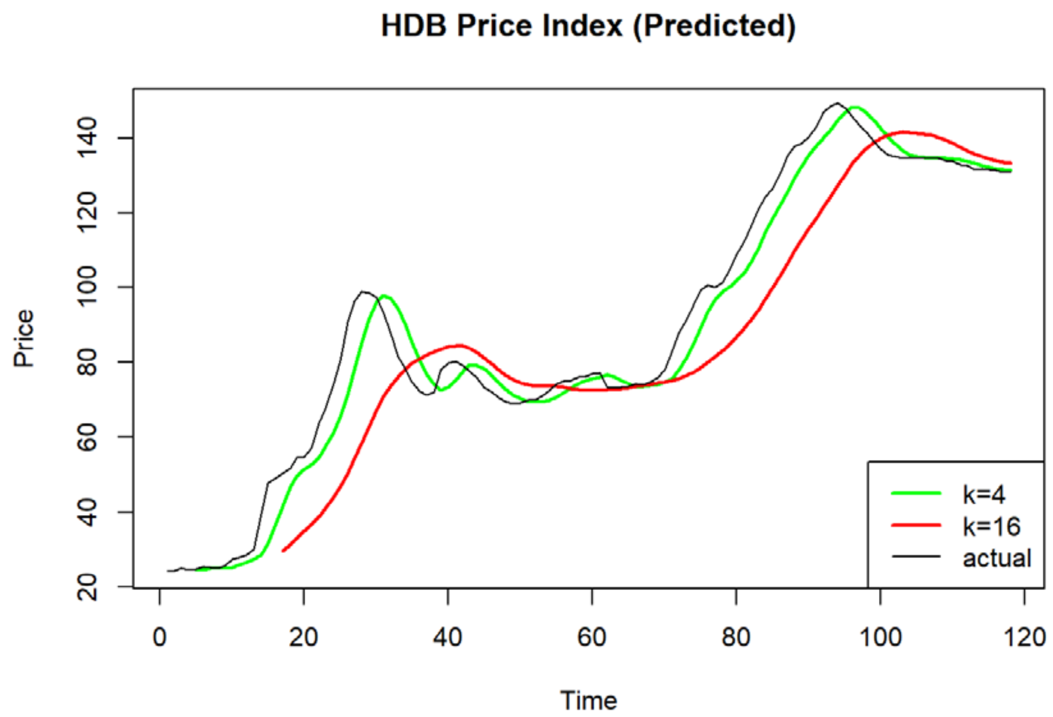
Which window size produced a better forecast? Evaluate by computing the RMSE for each model.

$$RMSE = \sqrt{\frac{1}{n}\Sigma(\widehat{y}_i - y_i)^2}$$

QUESTION 1B

```
hdb_long$SMApred4 = dplyr::lag(SMA(hdb_long$PriceIndex, n=4), 1)
hdb_long$SMApred16 = dplyr::lag(SMA(hdb_long$PriceIndex, n=16), 1)

# using base R to plot the original, pred4 & pred16 values
plot(hdb_long$TimeIndex,hdb_long$SMApred4,xlab="Time", ylab="Price",type="l", col="green", lwd=2)
lines(hdb_long$TimeIndex,hdb_long$SMApred16, col="red", lwd=2)
lines(hdb_long$TimeIndex,hdb_long$PriceIndex, col="black", lwd=1)
title("HDB Price Index (Predicted)")
legend("bottomright",c("k=4","k=16","actual"), lwd=c(2,2,1), col=c("green","red","black"))
```



- Why are the SMA lines are slightly "offset" to the right?
- What happens when we increase the window size?

QUESTION 1B

```
#computing RMSE for window=4
rmse_sma4<-sqrt(mean((hdb_long$SMApred4-hdb_long$PriceIndex)^2,na.rm=TRUE)) #need to set na.rm=TRUE
rmse_sma4
```

6.456354

```
#computing RMSE for window=16
rmse_sma16<-sqrt(mean((hdb_long$SMApred16-hdb_long$PriceIndex)^2,na.rm=TRUE))
rmse_sma16
```

16.53608

- Which is the better model? why?

QUESTION 1C

Based on what you observed about the time-series in Q1a, fit a HoltWinters model to the HDB train dataset, hdbtrain_ts. Use the model to predict the next 6 periods (6 quarters), and plot the predictions.

1

Simple exponential smoothing

- **Stationary time series**
- **Smoothing parameter: α**

2

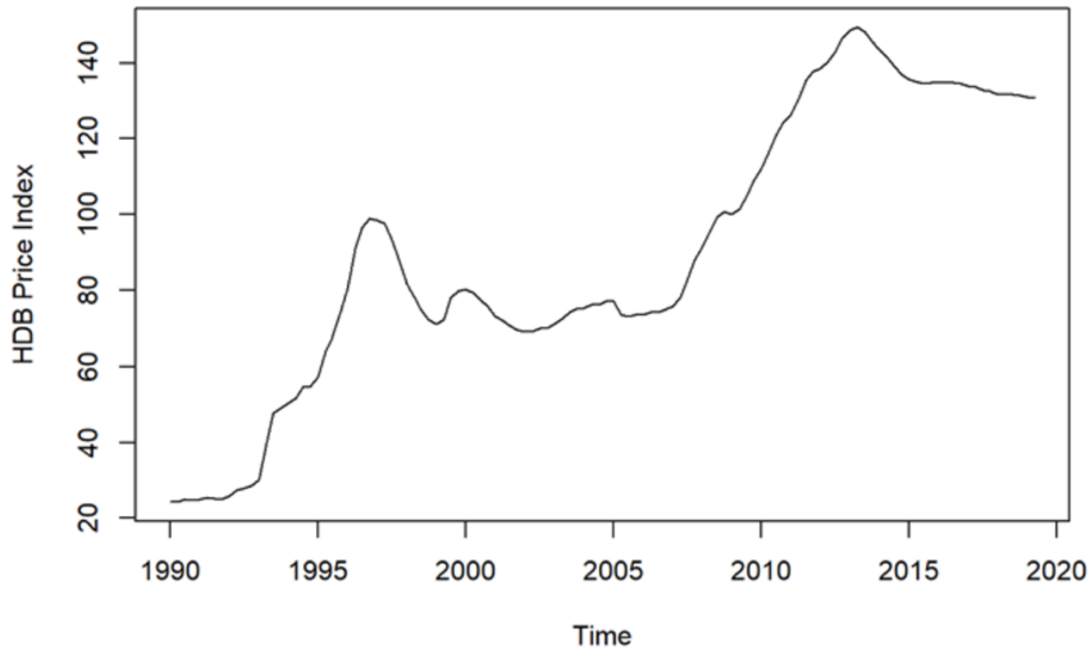
Double exponential smoothing

- **Time series with a trend.**
- **Smoothing parameter: α and β**

3

Triple exponential smoothing

- **Time series with seasonality.**
- **Smoothing parameter: α, β, Gamma**



- Which Holtwinters model is appropriate?

QUESTION 1C

Based on what you observed about the time-series in Q1a, fit a HoltWinters model to the HDB train dataset, hdbtrain_ts. Use the model to predict the next 6 periods (6 quarters), and plot the predictions.

```
hdbtrain_hw <- HoltWinters(hdbtrain_ts,gamma=FALSE)
hdbtrain_hw
```

- Why is gamma=FALSE?
- What happens if we remove the gamma=FALSE argument?

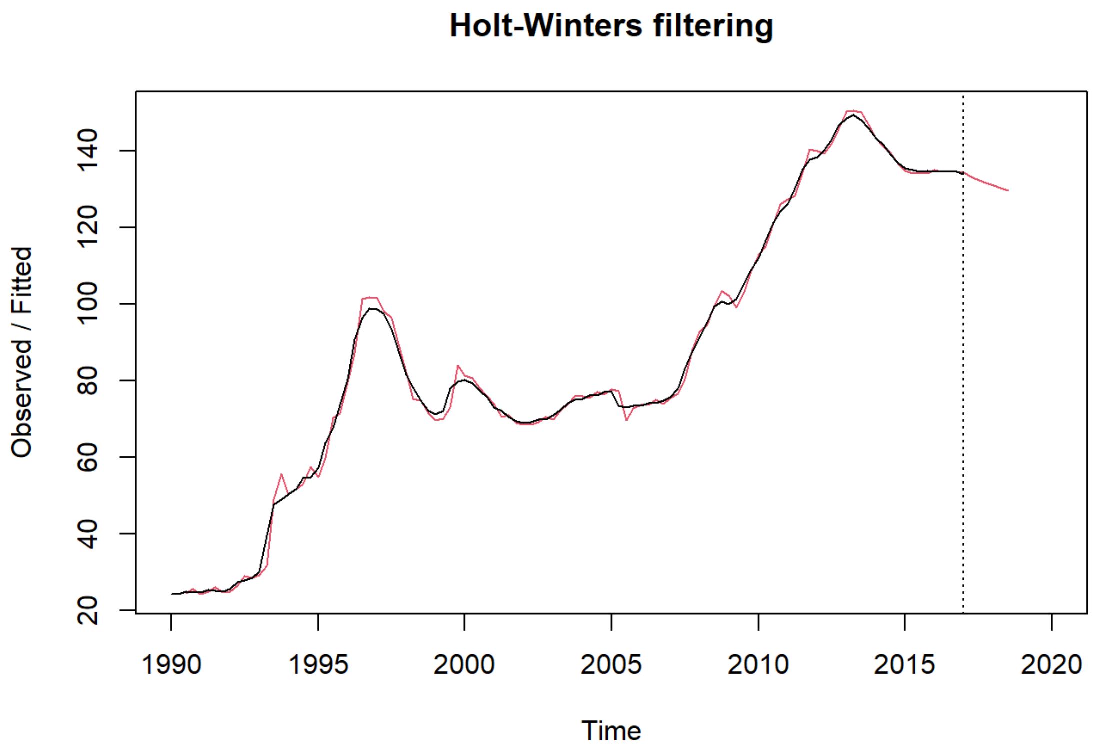
```
## Holt-Winters exponential smoothing with trend and without seasonal component.
##
## Call:
## HoltWinters(x = hdbtrain_ts, gamma = FALSE)
##
## Smoothing parameters:
##  alpha: 1
##  beta : 1
##  gamma: FALSE
##
## Coefficients:
##     [,1]
## a 133.9
## b  -0.7
```

What does it mean when α=1?

QUESTION 1C

Based on what you observed about the time-series in Q1a, fit a HoltWinters model to the HDB train dataset, hdbtrain_ts. Use the model to predict the next 6 periods (6 quarters), and plot the predictions.

```
hdbtrain_hw_pred<-predict(hdbtrain_hw,n.ahead=6)
plot(hdbtrain_hw,hdbtrain_hw_pred, xlim=c(1990,2020))
```



QUESTION 1D

Compare the HoltWinters model’s (Q1c) predictions with hdb_test, which contains the actual values for 2018/2019. (Hint: use XXX[1:Y] to extract the first few values from the predict object).

- Make a plot of the Holt-Winters predictions and the actual values in hdb_test, both on the y axis and with time on the horizontal axis. Use colors and/or linetypes to differentiate, and include a legend.

- What is the RMSE for these 6 predicted data points?

QUESTION 1D

What is the RMSE for these 6 predicted data points?

$$RMSE = \sqrt{\frac{1}{n}\Sigma(\hat{y}_i - y_i)^2}$$

```
rmse_hdbtrain_hw =sqrt(mean((hdbtrain_hw_pred[1:6] - hdb_test$PriceIndex)^2))
rmse_hdbtrain_hw
```

0.9036961

What is the RMSE for these 6 predicted data points if Holtwinter is ran with gamma=TRUE?

```
hdbtrain_hw_with_gamma_pred = predict(HoltWinters(hdbtrain_ts), n.ahead=6)
rmse_hdbtrainhw_with_gamma_pred = sqrt(mean((hdbtrain_hw_with_gamma_pred[1:6] - hdb_test$PriceIndex)^2))
rmse_hdbtrainhw_with_gamma_pred
```

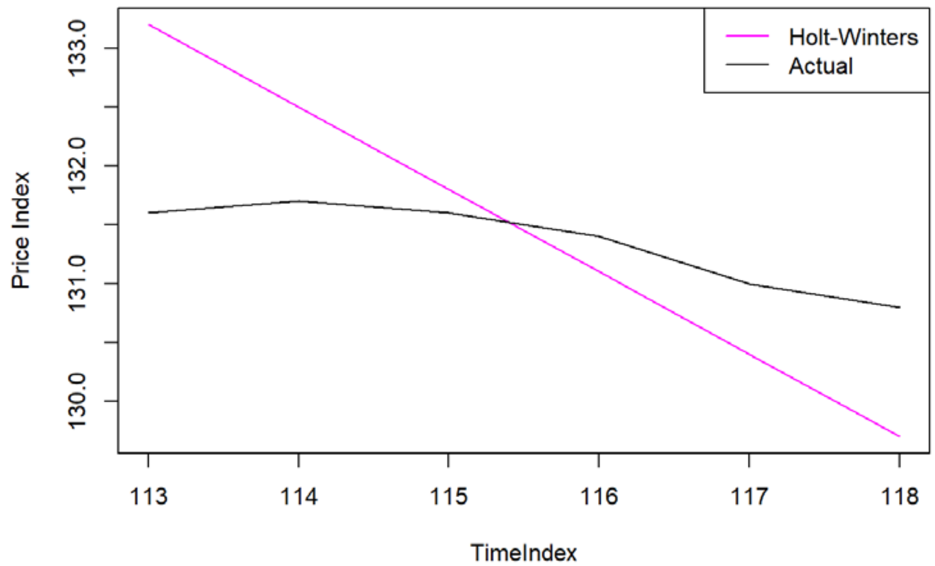
1.064116

QUESTION 1D

Make a plot of the Holt-Winters predictions and the actual values in hdb_test, both on the y axis and with time on the horizontal axis. Use colors and/or linetypes to differentiate, and include a legend.

```
plot_min_value = min(c(hdbtrain_hw_pred[1:6], hdb_test$PriceIndex))
plot_max_value = max(c(hdbtrain_hw_pred[1:6], hdb_test$PriceIndex))

plot(113:118, hdbtrain_hw_pred[1:6], type='l', col="magenta", ylim=c(plot_min_value, plot_max_value),ylab="Price Index", xla
b="TimeIndex")
lines(113:118, hdb_test$PriceIndex, type="l", col="black")
legend("topright", legend=c("Holt-Winters", "Actual"), col=c("magenta", "black"), lty=1)
```



QUESTION 1E

Compute also the RMSE for the 6 predicted data points using SMA4 (Simple Moving Average with Window of 4) against the hdb_test.

Add the line for SMA4 to the plot in Q1d. What do you notice? Does HoltWinters model perform better than the SMA4 model?

(Discuss: The RMSE for SMA4 is computed based on the complete dataset. What if you have computed the SMA4 using the hdb_train dataset instead?)

QUESTION 1E

```
#compute rmse for SMA for last 6 observations
rmse_sma4_last6<-sqrt(mean((hdb_long$SMApred4[113:118] - hdb_test$PriceIndex)^2))

rmse_hdbtrain_hw
```

```
## [1] 0.9036961
```

```
rmse_sma4_last6
```

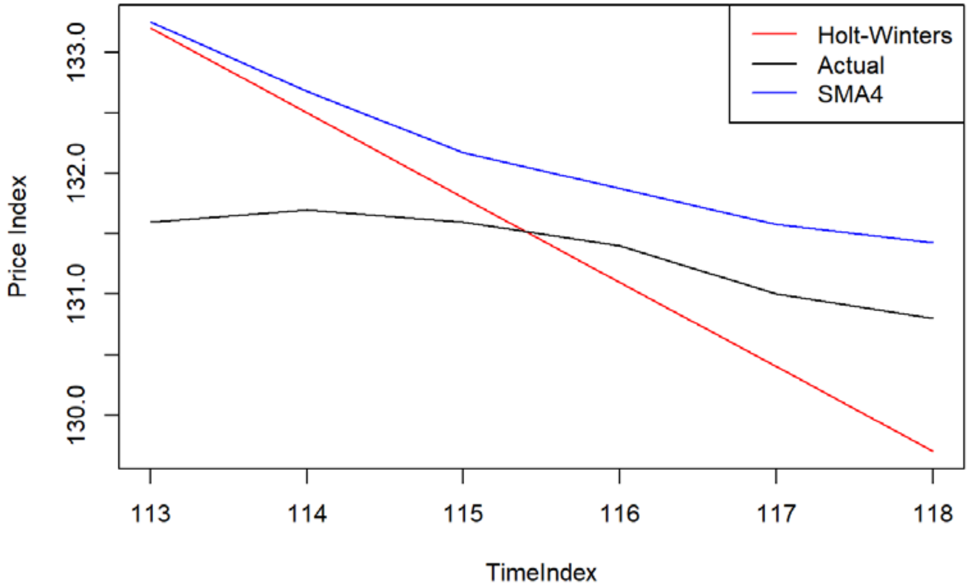
```
## [1] 0.9083524
```

- Comparing the RMSE, we can see that the HW model has a slightly lower RMSE hence performs better.

QUESTION 1E

```
plot_min_value = min(c(hdbtrain_hw_pred[1:6], hdb_test$PriceIndex))
plot_max_value = max(c(hdb_long$SMApred4[113:118], hdbtrain_hw_pred[1:6], hdb_test$PriceIndex))

plot(113:118, hdbtrain_hw_pred[1:6], type='l', col="red", ylim=c(plot_min_value, plot_max_value),ylab="Price Index", xlab="T
imeIndex")
lines(113:118, hdb_test$PriceIndex, type="l", col="black")
lines(113:118,hdb_long$SMApred4[113:118], type="l", col="blue")
legend("topright", legend=c("Holt-Winters", "Actual","SMA4"), col=c("red", "black","blue"), lty=1)
```



- The SMA4 model predicts values that are all greater than the actual values whereas the HW model has predictions that are above and below the actual values.

Tutorial 8 Question 2

TUTORIAL 8 QUESTION 2

Dataset required: data('fertil3) in Wooldridge package

Note: This dataset comes from a publically available dataset from Jeffery Wooldridge Textbook. See data description here: https://rdrr.io/cran/wooldridge/man/fertil3.html

First, load in the time series data for this question. There are 72 observations on 24 variables about women fertility rate between year 1913 and 1984. Key variables are listed below:

gfr	births per 1000 women between age 15 and 44.
pe	real value personal tax exemption in US dollars.
t	time trend, t=1,…,72
ww2	a binary variable = 1 during World War 2 between 1941 and 1945.
pill	a binary variable = 1 from 1963 on when the birth control pill was made available for contraception.

DATA PREPARATION

```
# read dataset into workplace, note that you need library(wooldridge) to load this data set
data('fertil3')
# convert the data to ts object, with frequency = 1 and start = 1913
fertil = ts(fertil3, frequency = 1, start = 1913)
```

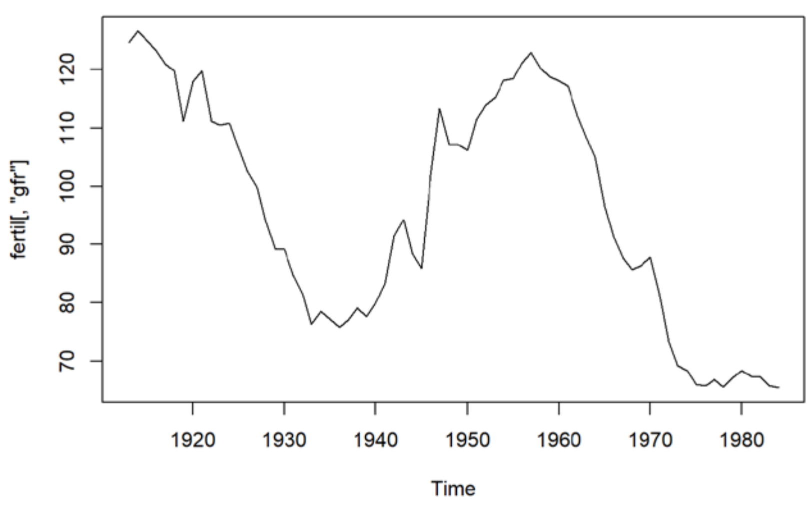
High fertility rate is essential for long-term growth in any economy. Many countries are bothered with low or even negative fertility rate. For example, the fertility rate in Singapore in general shows a decreasing trend in recent decades: https://www.channelnewsasia.com/news/singapore/number-of-babies-born-in-singapore-falls-to-lowest-in-8-years-11743722

Fertility is affected by many socio-economic factors, including single rate, family disposable income, level of tax duty, war attrition, contraception technology, etc. fertil3 data contains information about women’s fertility rate and personal tax exemption in U.S. in early-mid 1900s.

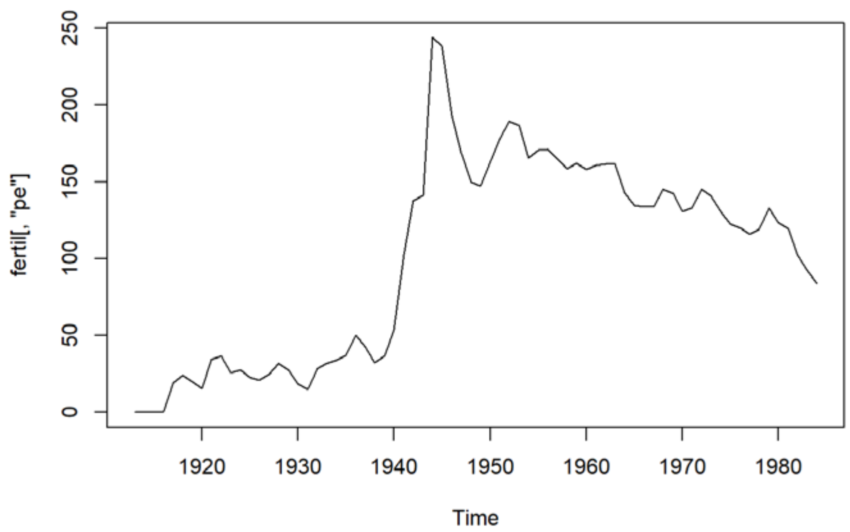
QUESTION 2A

Start off by plotting gfr (fertility rate) and pe (personal tax exemption) against time. What do you observe from the time series plots alone? Do you see any trend or seasonality? Is gfr time series stationary?

```r
plot(fertil[,"gfr"])
```



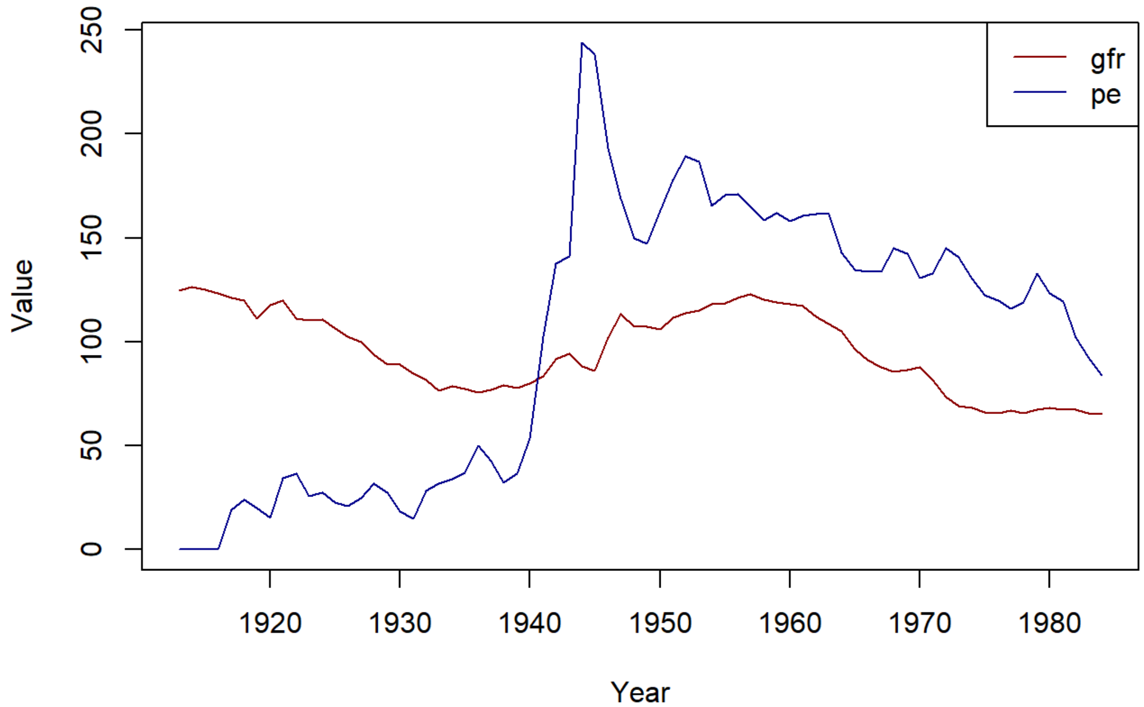

```r
plot(fertil[,"pe"])
```



QUESTION 2A

Start off by plotting gfr (fertility rate) and pe (personal tax exemption) against time. What do you observe from the time series plots alone? Do you see any trend or seasonality? Is gfr time series stationary?

```
# Or we can plot two time series together
ts.plot(fertil[,"gfr"], fertil[,"pe"], gpars = list(xlab = "Year", ylab = "Value", col = c("darkred","darkblue")))
legend("topright", legend = c("gfr", "pe"), col = c("darkred","darkblue"), lty = 1)
```



QUESTION 2B

To study the relationship between personal tax exemption (economic factor) and fertility rate, run a linear regression of gfr on pe.

$$Fertility\ rate = \beta_0 + \beta_1 Personal\ tax\ exemption + \epsilon$$

```
# run a OLS linear regression of 'gfr' on 'pe'
fit1 = lm(gfr ~ pe, data = fertil3)
summary(fit1)
```

QUESTION 2B

Interpret the coefficient before pe. Is it statistically significant? What's your conclusion about whether tax exemption improves fertility rate?

```
# run a OLS linear regression of 'gfr' on 'pe'
fit1 = lm(gfr ~ pe, data = fertil3)
summary(fit1)
```

```
##
## Call:
## lm(formula = gfr ~ pe, data = fertil3)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -30.35 -17.82  -1.68  18.34  30.26
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 96.344294   4.304734  22.381   <2e-16 ***
## pe          -0.007095   0.035923  -0.198    0.844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.94 on 70 degrees of freedom
## Multiple R-squared:  0.000557,   Adjusted R-squared:  -0.01372
## F-statistic: 0.03901 on 1 and 70 DF,  p-value: 0.844
```

What can you tell from the magnitude of the pe coefficient, t-value and p-value?

QUESTION 2C

Now, include the time trend variable t into the regression model. Interpret the coefficient before pe and t. What is the change in the regression result compared to the previous one? Which model should we choose and why?

```
fit2 = lm(gfr ~ pe + t, data = fertil3)
summary(fit2)
```

```
##
## Call:
## lm(formula = gfr ~ pe + t, data = fertil3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.659  -9.934   1.841  11.027  22.882
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 109.93016    3.47526  31.632  < 2e-16 ***
## pe            0.18666    0.03463   5.391 9.23e-07 ***
## t            -0.90519    0.10899  -8.305 5.53e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.2 on 69 degrees of freedom
## Multiple R-squared:  0.5002, Adjusted R-squared:  0.4857
## F-statistic: 34.53 on 2 and 69 DF,  p-value: 4.064e-11
```

What can you tell from the magnitude of the pe coefficient, t-value and p-value?

What is spurious regression?

QUESTION 2D

Up to now, we have applied static model on gfr and pe. Many have argued that fertility rate might respond to the tax exemption in previous periods (pe1, pe2), ww2 (war attrition) and pill (introduction of contraception technology).

Run a linear regression model with lag terms of pe, i.e. regress gfr ~ pe + pe_1 + pe_2 + ww2 + pill + t.

Interpret the coefficient before pe_1 and ww2.

Do you think fertility rate responses to personal tax exemption in previous periods?

QUESTION 2D

```
fit3 = lm(gfr ~ pe + pe_1 + pe_2 + ww2 + pill + t, data = fertil3)
summary(fit3)
```

```
##
## Call:
## lm(formula = gfr ~ pe + pe_1 + pe_2 + ww2 + pill + t, data = fertil3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.7343  -9.0588   0.3934  10.2415  18.6394
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 109.67794    3.60769  30.401  < 2e-16 ***
## pe            0.19286    0.10475   1.841  0.07029 .
## pe_1         -0.04129    0.12740  -0.324  0.74691
## pe_2          0.12773    0.10451   1.222  0.22619
## ww2         -26.44527    8.80545  -3.003  0.00383 **
## pill         -0.37731    6.30726  -0.060  0.95249
## t            -1.10578    0.19316  -5.725  3.1e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.67 on 63 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6702, Adjusted R-squared:  0.6388
## F-statistic: 21.33 on 6 and 63 DF,  p-value: 1.689e-13
```

QUESTION 2E

Check the linear regression model in (2d) for multicollinearity. Are there any independent variables that are highly correlated or do any independent variables create issues of multicollinearity? What would you do to address the multicollinearity issues, if any?

```
# check correlation of the IVs
cor_iv<-fertil3%>%select(pe,pe_1,pe_2,ww2,pill,t)
cor(cor_iv)
```

```
##             pe pe_1 pe_2         ww2       pill           t
## pe   1.0000000   NA   NA  0.30201821  0.2750324  0.67376166
## pe_1        NA    1   NA          NA         NA          NA
## pe_2        NA   NA    1          NA         NA          NA
## ww2  0.3020182   NA   NA  1.00000000 -0.1812066 -0.07229534
## pill 0.2750324   NA   NA -0.18120657  1.0000000  0.79793289
## t    0.6737617   NA   NA -0.07229534  0.7979329  1.00000000
```

Do we have possible multicollinearity issues?

Correlation of t and pill to be > 0.7 hence suggesting possible multicollinearity issues.

QUESTION 2E

Check the linear regression model in (2d) for multicollinearity. Are there any independent variables that are highly correlated or do any independent variables create issues of multicollinearity? What would you do to address the multicollinearity issues, if any?

```
fit3 = lm(gfr ~ pe + pe_1 + pe_2 + ww2 + pill + t, data = fertil3)
summary(fit3)
```

```
#check vif of the model
vif(fit3)
```

```
##        pe      pe_1      pe_2       ww2      pill         t
## 23.169776 35.491843 24.701175  2.645401  4.410117  7.835632
```

Any multicollinearity issues?

- 5 variables have VIF > 3.
- This suggests the model has serious multicollinearity issues.

QUESTION 2E: USE STEPWISE REGRESSION TO SELECT BEST MODEL

```
# use stepwise regression to check for best model
ols_step_best_subset(fit3)
```

```
##        Best Subsets Regression
## --------------------------------------
## Model Index    Predictors
## --------------------------------------
##      1         pill
##      2         pe_2 t
##      3         pe ww2 t
##      4         pe pe_2 ww2 t
##      5         pe pe_1 pe_2 ww2 t
##      6         pe pe_1 pe_2 ww2 pill t
## --------------------------------------
##
##                                                    Subsets Regression Summary
## ------------------------------------------------------------------------------------------------------------------------
------------
##                         Adj.        Pred
## Model      R-Square   R-Square    R-Square     C(p)        AIC         SBIC        SBC          MSEP         FPE        HS
P       APC
## ------------------------------------------------------------------------------------------------------------------------
------------
##   1         0.3693     0.3603      0.3352     61.0607     606.1015    399.6149    612.9315    18064.6785    257.8648     3.6
362    0.6667
##   2         0.6201     0.6088      0.5716      8.5647     553.0976    354.2317    562.0916    10319.2335    153.6994     2.2
331    0.4139
##   3         0.6621     0.6472      0.6066      5.1517     565.1749    361.1739    576.5583     9968.1252    146.0737     2.0
655    0.3777
##   4         0.6696     0.6493      0.5922      3.1105     547.3263    349.7427    560.8173     9255.2472    141.5626     2.0
645    0.3812
##   5         0.6702     0.6444      0.5031      5.0036     549.2076    351.8641    564.9470     9386.2225    145.4446     2.1
264    0.3917
##   6         0.6702     0.6388      0.4837      7.0000     551.2036    354.0831    569.1916     9537.0714    149.6889     2.1
949    0.4031
## ------------------------------------------------------------------------------------------------------------------------
------------
## AIC: Akaike Information Criteria
##  SBIC: Sawa's Bayesian Information Criteria
##  SBC: Schwarz Bayesian Criteria
##  MSEP: Estimated error of prediction, assuming multivariate normality
##  FPE: Final Prediction Error
##  HSP: Hocking's Sp
##  APC: Amemiya Prediction Criteria
```

QUESTION 2E: USE STEPWISE REGRESSION TO SELECT BEST MODEL

```
fit4 = lm(gfr ~ pe + pe_2 + ww2  + t, data = fertil3)
summary(fit4)
```

Remove *pill* from the predictors.

```
##
## Call:
## lm(formula = gfr ~ pe + pe_2 + ww2 + t, data = fertil3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.5082  -9.0034   0.1651  10.3193  18.5341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 109.81071    3.01894  36.374  < 2e-16 ***
## pe            0.17436    0.08303   2.100  0.03963 *
## pe_2          0.10592    0.07688   1.378  0.17299
## ww2         -26.64595    8.65585  -3.078  0.00305 **
## t            -1.11495    0.09820 -11.354  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.49 on 65 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6696, Adjusted R-squared:  0.6493
## F-statistic: 32.93 on 4 and 65 DF,  p-value: 5.326e-15
```

```
vif(fit4)
```

Any multicollinearity issues?

```
##        pe      pe_2       ww2         t
## 14.995797 13.767308  2.632809  2.085573
```

QUESTION 2E: USE STEPWISE REGRESSION TO SELECT BEST MODEL

```
fit5 = lm(gfr ~ pe  + ww2  + t, data = fertil3)
summary(fit5)
```

Remove *pe_2* from the predictors.

```
##
## Call:
## lm(formula = gfr ~ pe + ww2 + t, data = fertil3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.5377  -8.3846   0.3382   9.8265  17.3456
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 111.50327    2.89160  38.561  < 2e-16 ***
## pe            0.27524    0.03261   8.441 3.47e-12 ***
## ww2         -35.48625    6.21706  -5.708 2.73e-07 ***
## t            -1.12441    0.09811 -11.461  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.76 on 68 degrees of freedom
## Multiple R-squared:  0.6621, Adjusted R-squared:  0.6472
## F-statistic: 44.41 on 3 and 68 DF,  p-value: 5.2e-16
```

```
vif(fit5)
```

```
##       pe      ww2        t
## 2.367487 1.299547 2.162841
```

Any multicollinearity issues?

THANK YOU.

SEE YOU NEXT WEEK.