STATS 326
Applied Time Series
ASSIGNMENT THREE
ANSWER GUIDE

Question One:

```
> summary(SF.Barrow.fit1)

Call:
lm(formula = red.CO2.ts[-1] ~ Time[-1] + Quarter[-1] + red.CO2.ts[-68])

Residuals:
     Min       1Q   Median       3Q      Max 
-1.91405 -0.53160 -0.04121  0.45716  2.43353 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     203.99637   47.48575   4.296 6.35e-05 ***
Time[-1]          0.28656    0.06754   4.243 7.64e-05 ***
Quarter[-1]2     -2.92314    0.72227  -4.047 0.000148 ***
Quarter[-1]3    -16.03066    0.65394 -24.514  < 2e-16 ***
Quarter[-1]4     -0.92157    1.16188  -0.793 0.430753    
red.CO2.ts[-68]   0.46122    0.12888   3.579 0.000684 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7858 on 61 degrees of freedom
Multiple R-squared:  0.9958,    Adjusted R-squared:  0.9955 
F-statistic:  2912 on 5 and 61 DF,  p-value: < 2.2e-16

> t.69.sf.pred = SF.Barrow.fit1$coef[1]+SF.Barrow.fit1$coef[2]*69+
SF.Barrow.fit1$coef[6]*red.CO2.ts[68]
> t.69.sf.pred
(Intercept) 
   411.7948 
> t.70.sf.pred = SF.Barrow.fit1$coef[1]+SF.Barrow.fit1$coef[2]*70+
  SF.Barrow.fit1$coef[3]+SF.Barrow.fit1$coef[6]*t.69.sf.pred
> t.70.sf.pred
(Intercept) 
   411.0606 
> t.71.sf.pred = SF.Barrow.fit1$coef[1]+SF.Barrow.fit1$coef[2]*71+
  SF.Barrow.fit1$coef[4]+SF.Barrow.fit1$coef[6]*t.70.sf.pred
> t.71.sf.pred
(Intercept) 
    397.901 
> t.72.sf.pred = SF.Barrow.fit1$coef[1]+SF.Barrow.fit1$coef[2]*72+
  SF.Barrow.fit1$coef[5]+SF.Barrow.fit1$coef[6]*t.71.sf.pred
> t.72.sf.pred
(Intercept) 
   407.2272 

> SF.pred = c(t.69.sf.pred,t.70.sf.pred,t.71.sf.pred,t.72.sf.pred)
> names(SF.pred) = c("2017.1","2017.2","2017.3","2017.4")
> SF.pred
  2017.1   2017.2   2017.3   2017.4 
411.7948 411.0606 397.9010 407.2272 
> RMSEP.SF.Barrow = sqrt(1/4*sum((actual-SF.pred)^2))
> RMSEP.SF.Barrow
[1] 1.366159
```

The Seasonal Factor model included a Time variable, a seasonal factor and a lagged response variable. The Residual Series showed reasonably constant scatter about 0 with a larger residual for Quarter 4, 2016. The plot of the autocorrelation function of the Residual Series showed no significant lags. The residuals appeared to follow a normal distribution (Shapiro-Wilk *P-value* = 0.237) although the 5-number summary of the residuals showed slight right skew (min = -1.91, max = 2.43).

Quarters 2 – 4 CO_2 concentrations were all lower than the omitted baseline level (Quarter 1) with Quarter 3 being the lowest (16.03 ppm below Quarter 1).

The RMSEP was 1.37 ppm.

<u>Question Two</u>:

```
> summary(FH.Barrow.fit1)

Call:
lm(formula = red.CO2.ts[-1] ~ Time[-1] + c1[-1] + s1[-1] + c2[-1] +
    red.CO2.ts[-68])

Residuals:
     Min       1Q   Median       3Q      Max 
-1.91405 -0.53160 -0.04121  0.45716  2.43353 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     199.02753   47.51872   4.188 9.19e-05 ***
Time[-1]          0.28656    0.06754   4.243 7.64e-05 ***
c1[-1]            1.00078    0.90884   1.101 0.275150    
s1[-1]            8.01533    0.32697  24.514  < 2e-16 ***
c2[-1]            3.04649    0.28062  10.856 7.00e-16 ***
red.CO2.ts[-68]   0.46122    0.12888   3.579 0.000684 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7858 on 61 degrees of freedom
Multiple R-squared:  0.9958,    Adjusted R-squared:  0.9955 
F-statistic:  2912 on 5 and 61 DF,  p-value: < 2.2e-16

> t.69.fh.pred = FH.Barrow.fit1$coef[1]+FH.Barrow.fit1$coef[2]*69+
FH.Barrow.fit1$coef[3]*cos(2*pi*69*(1/4))+
FH.Barrow.fit1$coef[4]*sin(2*pi*69*(1/4))+
FH.Barrow.fit1$coef[5]*cos(2*pi*69*(2/4))+
FH.Barrow.fit1$coef[6]*red.CO2.ts[68]
> t.69.fh.pred
(Intercept) 
   411.7948 

> t.70.fh.pred = FH.Barrow.fit1$coef[1]+FH.Barrow.fit1$coef[2]*70+
FH.Barrow.fit1$coef[3]*cos(2*pi*70*(1/4))+
FH.Barrow.fit1$coef[4]*sin(2*pi*70*(1/4))+
FH.Barrow.fit1$coef[5]*cos(2*pi*70*(2/4))+
FH.Barrow.fit1$coef[6]*t.69.fh.pred
> t.70.fh.pred
(Intercept) 
   411.0606 

> t.71.fh.pred = FH.Barrow.fit1$coef[1]+FH.Barrow.fit1$coef[2]*71+
FH.Barrow.fit1$coef[3]*cos(2*pi*71*(1/4))+
FH.Barrow.fit1$coef[4]*sin(2*pi*71*(1/4))+
FH.Barrow.fit1$coef[5]*cos(2*pi*71*(2/4))+
FH.Barrow.fit1$coef[6]*t.70.fh.pred
> t.71.fh.pred
(Intercept) 
    397.901 

> t.72.fh.pred = FH.Barrow.fit1$coef[1]+FH.Barrow.fit1$coef[2]*72+
FH.Barrow.fit1$coef[3]*cos(2*pi*72*(1/4))+
FH.Barrow.fit1$coef[4]*sin(2*pi*72*(1/4))+
FH.Barrow.fit1$coef[5]*cos(2*pi*72*(2/4))+
FH.Barrow.fit1$coef[6]*t.71.fh.pred
> t.72.fh.pred
(Intercept) 
   407.2272 

> FH.pred = c(t.69.fh.pred,t.70.fh.pred,t.71.fh.pred,t.72.fh.pred)
> names(FH.pred) = c("2017.1","2017.2","2017.3","2017.4")
> FH.pred
  2017.1   2017.2   2017.3   2017.4 
411.7948 411.0606 397.9010 407.2272 

> RMSEP.FH.Barrow = sqrt(1/4*sum((actual-FH.pred)^2))
> RMSEP.FH.Barrow
[1] 1.366159
```

The Full Harmonic model produced the same results as the Seasonal Factor model, as was expected. It had the smallest RMSEP (1.37 ppm) of all the Harmonic models.

The Full Harmonic model included a Time variable, 3 harmonics (c1 with a *P-value* = 0.28 being non-significant) and a lagged response variable. The Residual Series showed reasonably constant scatter about 0 with a larger residual for Quarter 4, 2016. The plot of the autocorrelation function of the Residual Series showed no significant lags. The residuals appeared to follow a normal distribution (Shapiro-Wilk *P-value* = 0.237) although the 5-number summary of the residuals showed slight right skew (min = -1.91, max = 2.43).

A model retaining all pairs where 1 harmonic from the pair was significant is the same as the Full Harmonic model. The Reduced Harmonic model, removing c1, produced an RMSEP of 1.68 ppm so it was rejected. A single cosine model was not appropriate as the observations did not follow a smooth (harmonic) curve for each year.

<u>Question Three</u>: (Can use either Seasonal Factor model or Full Harmonic model)

The Seasonal Factor model included a Time variable, a seasonal factor and a lagged response variable to take care of autocorrelation.

The Residual Series shows reasonably constant scatter about 0 with a large positive residual for the last observation. The plot of the autocorrelation function of the Residual Series shows no significant lags. The residuals appear to follow a normal distribution (Shapiro-Wilk *P-value* = 0.24). The 5-number summary of the residuals shows slight right skew (min = -1.91, max = 2.43).

We have very strong evidence that the Time variable (*P-value* = 7.64×10^{-5}) is not equal to 0.

We have strong evidence that Quarter 2 is lower than the omitted baseline (Quarter 1) level (*P-value* = 0.00015) and extremely strong evidence that Quarter 3 is below the omitted baseline level (*P-value* ≈ 0). We have no evidence against the hypothesis that Quarter 4 is no different to Quarter 1 (*P-value* = 0.43).

We have strong evidence against the hypothesis of no autocorrelation (*P-value* = 0.00068).

The F-statistic provides extremely strong evidence against the hypothesis that none of the variables are related to the CO2 concentration (*P-value* ≈ 0). The Multiple R^2 is 0.996 indicating that nearly all the variation in the CO2 concentration is explained by the model.

The Residual Standard Error is 0.79 ppm so prediction intervals should be reasonably narrow. The model predictions can be relied on as the assumptions appear to be satisfied.

The RMSEP for the 2017 predictions was 1.37 which was smaller than the Reduced Harmonic model (1.68). It was the same as that of the Full Harmonic model, as expected.

Our predictions for 2017 were:

Quarter 1: 411.79 ppm
Quarter 2: 411.06 ppm
Quarter 3: 397.90 ppm
Quarter 4: 407.23 ppm

<u>Question Four</u>:

```
> Quarter.F = factor(rep(1:4,18))
> Time.F = 1:72

> SF.Barrow.Full.fit1 = lm(full.CO2.ts[-1]~Time.F[-1]+
  Quarter.F[-1]+full.CO2.ts[-72])
> summary(SF.Barrow.Full.fit1)

Call:
lm(formula = full.CO2.ts[-1] ~ Time.F[-1] + Quarter.F[-1] + full.CO2.ts[-
72])

Residuals:
     Min       1Q   Median       3Q      Max 
-1.90255 -0.57295 -0.06903  0.45185  2.21890 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      161.22208   37.53521   4.295 5.95e-05 ***
Time.F[-1]         0.22819    0.05463   4.177 9.00e-05 ***
Quarter.F[-1]2    -3.61378    0.59512  -6.072 7.27e-08 ***
Quarter.F[-1]3   -16.67694    0.53594 -31.117  < 2e-16 ***
Quarter.F[-1]4     0.04538    0.94576   0.048    0.962    
full.CO2.ts[-72]   0.57734    0.10194   5.663 3.64e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7952 on 65 degrees of freedom
Multiple R-squared:  0.9962,    Adjusted R-squared:  0.9959 
F-statistic:  3404 on 5 and 65 DF,  p-value: < 2.2e-16

> t.73.sf.pred = SF.Barrow.Full.fit1$coef[1]+
  SF.Barrow.Full.fit1$coef[2]*73+
  SF.Barrow.Full.fit1$coef[6]*full.CO2.ts[72]
> t.73.sf.pred
(Intercept) 
   413.6884 

> t.74.sf.pred = SF.Barrow.Full.fit1$coef[1]+
  SF.Barrow.Full.fit1$coef[2]*74+SF.Barrow.Full.fit1$coef[3]*1+
  SF.Barrow.Full.fit1$coef[6]*t.73.sf.pred
> t.74.sf.pred
(Intercept) 
   413.3328 

> t.75.sf.pred = SF.Barrow.Full.fit1$coef[1]+
  SF.Barrow.Full.fit1$coef[2]*75+SF.Barrow.Full.fit1$coef[4]*1+
  SF.Barrow.Full.fit1$coef[6]*t.74.sf.pred
> t.75.sf.pred
(Intercept) 
   400.2926 

> t.76.sf.pred = SF.Barrow.Full.fit1$coef[1]+
  SF.Barrow.Full.fit1$coef[2]*76+SF.Barrow.Full.fit1$coef[5]*1+
  SF.Barrow.Full.fit1$coef[6]*t.75.sf.pred
> t.76.sf.pred
(Intercept) 
   409.7145 
```

```
> SF.Full.pred = c(t.73.sf.pred,t.74.sf.pred,t.75.sf.pred,t.76.sf.pred)
> names(SF.Full.pred) = c("2018.1","2018.2","2018.3","2018.4")
> SF.Full.pred
  2018.1   2018.2   2018.3   2018.4 
413.6884 413.3328 400.2926 409.7145 
```

The model including the 2017 data has similar estimates to our previous model (2000 – 2016). The intercept is larger (204 compared to 161) while the estimate for Quarter 2 is lower (-3.61 compared to -2.92) while the autocorrelation estimate is higher (0.58 compared to 0.46). The estimate for Q4 changes sign from -0.92 to 0.05. The Residual Standard Error is (0.8 ppm) so the prediction intervals should be reasonably narrow. Our predictions should be reliable.

The prediction intervals are between 3.1 and 3.8 ppm.

Question Five:

The best predicting model is the Holt-Winters model as it has the lowest RMSEP (0.96). We should be able to rely on any predictions.

The prediction intervals for the Holt-Winters model (3.5 – 5.6) are wider than for the Seasonal Factor model (3.1 – 3.8).