Assignment 3

Presented by:

**Question1**

Figure 1.1 Time series plot of data

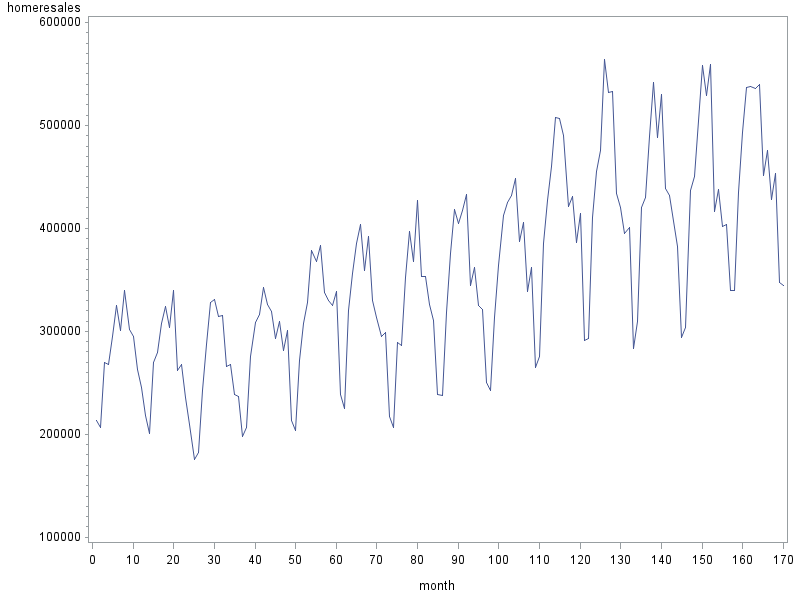
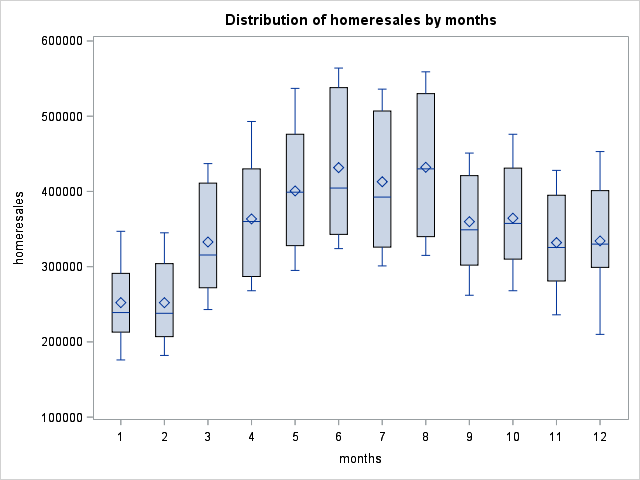


Figure 1.2 Box plot for each month



Observing Figure 1.1, we got the general idea that the data is drifting in an upward trend, and in a regular seasonal pattern. According to the box plot (Figure 1.2), the level of series––the home resales––is at its highest in June (the 6th month) with the peak more than 500000 and drop close to 300000 in January (the 1st month). Also, as time passes the seasonal swings are getting larger and larger, suggesting increasing variability in the time series.

**Question 2**

Table 2.1 Output of Regression Model

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: homeresales**

|  |  |
| --- | --- |
| **Number of Observations Read** | 170 |
| **Number of Observations Used** | 170 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 12 | 1.331273E12 | 1.109394E11 | 162.44 | <.0001 |
| **Error** | 157 | 1.072212E11 | 682937337 |  |  |
| **Corrected Total** | 169 | 1.438494E12 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 26133 | **R-Square** | 0.9255 |
| **Dependent Mean** | 354465 | **Adj R-Sq** | 0.9198 |
| **Coeff Var** | 7.37254 |  |  |

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **95% Confidence Limits** | |
| **Intercept** | 1 | 315539 | 7820.09421 | 40.35 | <.0001 | 300093 | 330985 |
| **m1** | 1 | -178516 | 9711.44001 | -18.38 | <.0001 | -197698 | -159334 |
| **m2** | 1 | -179938 | 9711.35388 | -18.53 | <.0001 | -199120 | -160756 |
| **m3** | 1 | -92511 | 9879.48879 | -9.36 | <.0001 | -112024 | -72997 |
| **m4** | 1 | -63223 | 9878.72678 | -6.40 | <.0001 | -82735 | -43710 |
| **m5** | 1 | -27363 | 9878.13407 | -2.77 | 0.0063 | -46875 | -7852.30870 |
| **m6** | 1 | 2281.48148 | 9877.71068 | 0.23 | 0.8176 | -17229 | 21792 |
| **m7** | 1 | -17788 | 9877.45664 | -1.80 | 0.0736 | -37298 | 1722.01462 |
| **m9** | 1 | -73641 | 9877.45664 | -7.46 | <.0001 | -93151 | -54131 |
| **m10** | 1 | -70353 | 9877.71068 | -7.12 | <.0001 | -89863 | -50843 |
| **m11** | 1 | -104208 | 9878.13407 | -10.55 | <.0001 | -123719 | -84697 |
| **m12** | 1 | -103277 | 9878.72678 | -10.45 | <.0001 | -122790 | -83765 |
| **month** | 1 | 1355.02646 | 40.90088 | 33.13 | <.0001 | 1274.23948 | 1435.81343 |

**Regression model:**

The residuals are normally distributed, and the R-square is equal to 0.9225. Together they suggest that the regression model fits the data relatively well.

**Overall hypothesis testing:**

H0: βi = 0, Ɐ i ≠ 0;

(Variability in the resales does not exist.)

Ha: ⱻ i ≠0, s.t. βi ≠ 0.

(Variability exists.)

According to analysis of variance (ANVOA) table, we rejected the null hypothesis at 95% confidence level based on P-value < 0.0001 and statistical F = 162.44 > 1.93, and concluded there is at least one significant coefficient.

**Individual component hypothesis testing:**

Ɐ i ≠ 0：

H0: βi = 0;

(Coefficient is not significant.)

Ha: βi ≠ 0.

(Coefficient is significant.)

According to the table of parameter estimates, P-value for every component in the model except M6 (June) and M7 (July) is less than 0.0001; therefore, coefficients M1 to M5 and M9 to M12 are significant, and M6 and M7 are not significant. This conclusion is consistent with the analysis from 95% confidence intervals––the only two CI that contains zero are CI for coefficients of M6 and M7.

The significant coefficients are those of: M1, M2, M3, M4, M5, M9, M10, M11, M12, and Month.

**Question 3**

The constant (intercept) of the model, β0=315539, is the intercept of deseasonalized trend line with Y-axis. This value is determined by the reference month we chose, i.e. August.

January coefficient is -178516, and November coefficient is -104208. These coefficients describe the seasonal difference between the indicated month––January or November––and the reference month––August.

**Question 4**

**Seasonality:**

Table 4.1 Testing Seasonality

| **Test 1 Results for Dependent Variable homeresales** | | | | |
| --- | --- | --- | --- | --- |
| **Source** | **DF** | **Mean Square** | **F Value** | **Pr > F** |
| **Numerator** | 11 | 54078421110 | 79.19 | <.0001 |
| **Denominator** | 157 | 682937337 |  |  |

H0: βi = 0, for all seasonal terms.

(Seasonality does not exist.)

Ha: At least one βi ≠ 0, among all seasonal terms.

(Seasonality exists.)

According to the above table, P-value is smaller than 0.05 and the statistical F = 79.19 > 1.93 = F0.05 (11,157) ; therefore, we rejected the null hypothesis at 95% confidence level, and concluded that there is significant evidence for seasonality to exist in home resales data.

**Upward trend**:

Table 4.2 Estimate of Trend Coefficient

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **95% Confidence Limits** | |
| **month** | 1 | 1355.02646 | 40.90088 | 33.13 | <.0001 | 1274.23948 | 1435.81343 |

The t-test given here is for a two-tailed test; however, to test for upward trending, it needs to be adjusted for a one-tailed test.

H0: βt = 0;

Ha: βt > 0.

For one-tailed hypothesis testing, since t = 33.13 > 0, p(one-tail) = ½ p(two-tail) < 0.0001. Therefore, we rejected the null hypothesis at 95% confidence level, and an upward trend in home resales is confirmed.

**Question 5**

**Hypothesis testing for equality of June and July coefficients:**

Table 5.1 Test for June and July Coefficients

| **Test 3 Results for Dependent Variable homeresales** | | | | |
| --- | --- | --- | --- | --- |
| **Source** | **DF** | **Mean Square** | **F Value** | **Pr > F** |
| **Numerator** | 1 | 2819392693 | 4.13 | 0.0439 |
| **Denominator** | 157 | 682937337 |  |  |

H0: β6 = β7;

Ha: β6 ≠ β7.

Using the test function in SAS, P-value is smaller than 0.05 and statistical F = 4.13 > 3.94 = F0.05 (1,157), and reject the null hypothesis at 95% confidence level; therefore, the June coefficient is significantly different from the July coefficient.

**Question 6**

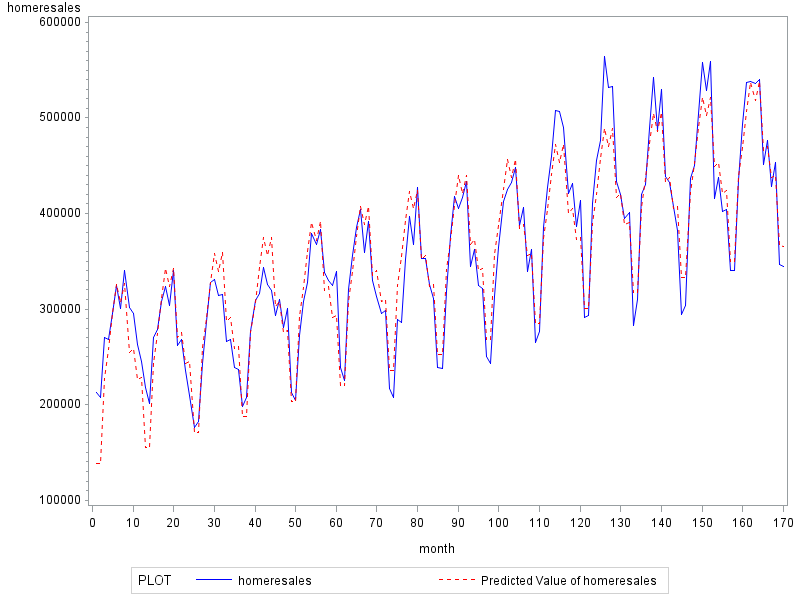


Figure 6.1 Time Series Plot of Actual Data and Model Prediction

The above figure shows that the model has a quite good fit for the original data. However, as time passes, the model fails to grasp the increasing variability, and consequent higher differences are observed near both ends of the graph compared to months close to center. R-square of the model is 0.9255, precisely denoting that 92.55% of variation in the data can be explained by the regression model.

**Question 7**

Model with reference to August:

Model with reference to January：

Model result is independent of the reference, i.e.

(7.3)

Both seasonal terms have period of 12, i.e.

(7.4)

Substitute (7.1) and (7.2) into (7.4) separately, and the seasonal term would be eliminated on both sides:

(7.6)

(7.5)

Expand equations Y(1)=Y(1)’ and Y(8)=Y(8)’:

(7.8)

(7.7)

Since βt=βt’, they can be eliminated:

(7.10)

(7.9)

(7.10) minus (7.9):

(7.11)

That is, with the reference month being January, the August coefficient β8’ would be negative value of the January coefficient with reference to August, β1.