Business Statistics and Analysis

Business Statistics and Analysis

In the previous assignment…

Business Statistics and Analysis

In the previous assignment…

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- Using the 2013 data.
- Basic cleaning of data.
- Using data on ‘Single Family Housing’.

  TYPE = 1 and STRUCTURETYPE = 1

Business Statistics and Analysis

In the previous assignment…

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- Using the 2013 data.
- Basic cleaning of data.
- Using data on ‘Single Family Housing’.

  TYPE = 1 and STRUCTURETYPE = 1

Business Statistics and Analysis

In the previous assignment…

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- ✓ Using the 2013 data.
- Basic cleaning of data.
- Using data on ‘Single Family Housing’.

  TYPE = 1 and STRUCTURETYPE = 1

Business Statistics and Analysis

In the previous assignment…

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- ✓ Using the 2013 data.
- ✓ Basic cleaning of data.
- Using data on ‘Single Family Housing’.

  TYPE = 1 and STRUCTURETYPE = 1

Business Statistics and Analysis

In the previous assignment…

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- ✓ Using the 2013 data.
- ✓ Basic cleaning of data.
- ✓ Using data on ‘Single Family Housing’.

  TYPE = 1 and STRUCTURETYPE = 1

Business Statistics and Analysis

In the previous assignment…

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- ✓ Using the 2013 data.
- ✓ Basic cleaning of data.
- ✓ Using data on ‘Single Family Housing’.

  TYPE = 1 and STRUCTURETYPE = 1

Business Statistics and Analysis

In the previous assignment…

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- ✓ Using the 2013 data.
- ✓ Basic cleaning of data.
- ✓ Using data on ‘Single Family Housing’.

  TYPE = 1 and STRUCTURETYPE = 1

Business Statistics and Analysis

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- **Concurrent Y and X variables.**

Business Statistics and Analysis

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- **Concurrent Y and X variables.**
- **Y and X variables belong to the same year.**

Business Statistics and Analysis

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- **Concurrent Y and X variables.**
- **Y and X variables belong to the same year.**

Using the regression model to predict market value two years from now may be problematic.

Business Statistics and Analysis

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- Concurrent Y and X variables.
- Y and X variables belong to the same year.

Using the regression model to predict market value two years from now may be problematic.

- Use the lag of X variables.

Business Statistics and Analysis

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- Concurrent Y and X variables.
- Y and X variables belong to the same year.

Using the regression model to predict market value two years from now may be problematic.

- Use the lag of X variables.
- Y variable from 2013 and X variables from 2011.

Business Statistics and Analysis

$$VALUE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- Concurrent Y and X variables.
- Y and X variables belong to the same year.

Using the regression model to predict market value two years from now may be problematic.

- Use the lag of X variables.
- Y variable from 2013 and X variables from 2011.
- This will allow predictions about future market value of housing units.

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Merge the VALUE variable from 2013 data into the 2011 data.

  (*VLOOKUP command using the CONTROL variable to match the data files*)

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Merge the VALUE variable from 2013 data into the 2011 data.

  (*VLOOKUP command using the CONTROL variable to match the data files*)

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Merge the VALUE variable from 2013 data into the 2011 data.
  (*VLOOKUP command using the CONTROL variable to match the data files*)
- Data cleaning.
  - Delete all rows for housing units that are not common across the two years.
  - Consider only 'Single Family Housing'
    TYPE = 1 and STRUCTURETYPE = 1
  - Delete all Housing units which have a market value of less than $1000, that is, delete units with VALUE < $1000

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Merge the VALUE variable from 2013 data into the 2011 data.
  (*VLOOKUP command using the CONTROL variable to match the data files*)
- Data cleaning.
  - Delete all rows for housing units that are not common across the two years.
  - Consider only 'Single Family Housing'
    TYPE = 1 and STRUCTURETYPE = 1
  - Delete all Housing units which have a market value of less than $1000, that is, delete units with VALUE < $1000

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Merge the VALUE variable from 2013 data into the 2011 data.
  (*VLOOKUP command using the CONTROL variable to match the data files*)
- Data cleaning.
  - Delete all rows for housing units that are not common across the two years.
  - Consider only 'Single Family Housing'
    TYPE = 1 and STRUCTURETYPE = 1
  - Delete all Housing units which have a market value of less than $1000, that is, delete units with VALUE < $1000

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Merge the VALUE variable from 2013 data into the 2011 data.
  (*VLOOKUP command using the CONTROL variable to match the data files*)
- Data cleaning.
  - Delete all rows for housing units that are not common across the two years.
  - Consider only 'Single Family Housing'
    TYPE = 1 and STRUCTURETYPE = 1
  - Delete all Housing units which have a market value of less than $1000, that is, delete units with VALUE < $1000

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Merge the VALUE variable from 2013 data into the 2011 data.
  (*VLOOKUP command using the CONTROL variable to match the data files*)
- Data cleaning.
  - Delete all rows for housing units that are not common across the two years.
  - Consider only 'Single Family Housing'
    TYPE = 1 and STRUCTURETYPE = 1
  - Delete all Housing units which have a market value of less than $1000, that is, delete units with VALUE < $1000

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Estimate a regression model using the 2013 VALUE variable and the X variables from the 2011 data.

Business Statistics and Analysis

Use the VALUE variable from 2013 data and X variables from 2011 data.

- Estimate a regression model using the 2013 VALUE variable and the X variables from the 2011 data.

$$VALUE^{(2013)} = \beta_0 + \beta_1 X_1^{(2011)} + \beta_2 X_2^{(2011)} + \cdots + + \beta_k X_k^{(2011)}$$

Business Statistics and Analysis

$$VALUE^{(2013)} = \beta_0 + \beta_1 X_1^{(2011)} + \beta_2 X_2^{(2011)} + \cdots + +\beta_k X_k^{(2011)}$$

Business Statistics and Analysis

$$VALUE^{(2013)} = \beta_0 + \beta_1 X_1^{(2011)} + \beta_2 X_2^{(2011)} + \cdots + + \beta_k X_k^{(2011)}$$

How well does the model predict future Market values?

Business Statistics and Analysis

$$VALUE^{(2013)} = \beta_0 + \beta_1 X_1^{(2011)} + \beta_2 X_2^{(2011)} + \cdots + + \beta_k X_k^{(2011)}$$

How well does the model predict future Market values?

- R-square measure.
- 'Holdout Analysis'.
  - Hold out some data and not include it in the regression.
  - Use the estimated regression model to predict the held out data.

Business Statistics and Analysis

$$VALUE^{(2013)} = \beta_0 + \beta_1 X_1^{(2011)} + \beta_2 X_2^{(2011)} + \cdots + + \beta_k X_k^{(2011)}$$

How well does the model predict future Market values?

- R-square measure.
- 'Holdout Analysis'.
  - Hold out some data and not include it in the regression.
  - Use the estimated regression model to predict the held out data.

Business Statistics and Analysis

$$VALUE^{(2013)} = \beta_0 + \beta_1 X_1^{(2011)} + \beta_2 X_2^{(2011)} + \cdots + + \beta_k X_k^{(2011)}$$

How well does the model predict future Market values?

- R-square measure.
- 'Holdout Analysis'.
  - **Hold out some data and not include it in the regression.**
  - **Use the estimated regression model to predict the held out data.**

Business Statistics and Analysis

$$VALUE^{(2013)} = \beta_0 + \beta_1 X_1^{(2011)} + \beta_2 X_2^{(2011)} + \cdots + + \beta_k X_k^{(2011)}$$

How well does the model predict future Market values?

- R-square measure.
- 'Holdout Analysis'.
  - **Hold out some data and not include it in the regression.**
  - **Use the estimated regression model to predict the held out data.**

Business Statistics and Analysis

'Holdout Analysis'

- From the data that you create, select 1000 Housing Units at random. This is your 'Holdout Sample'.
- To select a random sample…
  - Use the =RAND() function in Excel.
  - Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.
  - Estimate regression model on the remaining set of Housing Units.
- Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.

Business Statistics and Analysis

'Holdout Analysis'

- **From the data that you create, select 1000 Housing Units at random.** This is your 'Holdout Sample'.

- To select a random sample…
  - Use the =RAND() function in Excel.
  - Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.
  - Estimate regression model on the remaining set of Housing Units.

- Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.

Business Statistics and Analysis

'Holdout Analysis'

- **From the data that you create, select 1000 Housing Units at random. This is your 'Holdout Sample'.**

- **To select a random sample…**
  - **Use the =RAND() function in Excel.**
  - **Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.**
  - **Estimate regression model on the remaining set of Housing Units.**

- **Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.**

Business Statistics and Analysis

'Holdout Analysis'

- **From the data that you create, select 1000 Housing Units at random. This is your 'Holdout Sample'.**

- **To select a random sample…**
  - **Use the =RAND() function in Excel.**
  - **Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.**
  - **Estimate regression model on the remaining set of Housing Units.**

- **Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.**

Business Statistics and Analysis

'Holdout Analysis'

- **From the data that you create, select 1000 Housing Units at random. This is your 'Holdout Sample'.**

- **To select a random sample…**
  - **Use the =RAND() function in Excel.**
  - **Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.**
  - **Estimate regression model on the remaining set of Housing Units.**

- **Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.**

Business Statistics and Analysis

'Holdout Analysis'

- From the data that you create, select 1000 Housing Units at random. This is your 'Holdout Sample'.

- To select a random sample…
  - Use the =RAND() function in Excel.
  - Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.
  - Estimate regression model on the remaining set of Housing Units.

- Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.

Business Statistics and Analysis

'Holdout Analysis'

- From the data that you create, select 1000 Housing Units at random. This is your 'Holdout Sample'.

- To select a random sample…
  - Use the =RAND() function in Excel.
  - Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.
  - Estimate regression model on the remaining set of Housing Units.

- Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.

Business Statistics and Analysis

'Holdout Analysis'

- From the data that you create, select 1000 Housing Units at random. This is your 'Holdout Sample'.

- To select a random sample…
  - Use the =RAND() function in Excel.
  - Sort the data on this random value and select the top 1000 Housing Units for 'Holdout Sample'.
  - Estimate regression model on the remaining set of Housing Units.

- Using 'β' coefficients from regression model and X variables from 'Holdout Sample', predict VALUE for each Housing Unit in the Holdout Sample.

Business Statistics and Analysis

'Holdout Analysis'

- **Compare these thousand predictions with the actual market value of those housing units.**
  - Create a 'Mean Absolute Deviation' measure.
  - It is average of the absolute difference between *actual* and the *predicted* value.

Business Statistics and Analysis

'Holdout Analysis'

- **Compare these thousand predictions with the actual market value of those housing units.**
  - **Create a 'Mean Absolute Deviation' measure.**
  - **It is average of the absolute difference between *actual* and the *predicted* value.**

Business Statistics and Analysis

'Holdout Analysis'

- **Compare these thousand predictions with the actual market value of those housing units.**
  - **Create a 'Mean Absolute Deviation' measure.**
  - **It is average of the absolute difference between *actual* and the *predicted* value.**

Mean Absolute Deviation = $\left(\sum_{i=1}^{1000} |Actual_i - Predicted_i|\right) \Big/ 1000$

Business Statistics and Analysis

Your report, at the minimum, should cover the following...

1. **The regression output.**

2. **The calculation of 'Mean Absolute Deviation'.**

3. ***You may extend analysis to multiple years.***

Business Statistics and Analysis

Your report, at the minimum, should cover the following...

1. **The regression output.**
2. **The calculation of 'Mean Absolute Deviation'.**
3. ***You may extend analysis to multiple years.***

Business Statistics and Analysis

Your report, at the minimum, should cover the following...

1. **The regression output.**

2. **The calculation of 'Mean Absolute Deviation'.**

3. ***You may extend analysis to multiple years.***

Business Statistics and Analysis

Your report, at the minimum, should cover the following...

1. **The regression output.**

2. **The calculation of 'Mean Absolute Deviation'.**

3. ***You may extend analysis to multiple years.***