



## Factorial ANOVA analysis on housing prices based on amenities and environment

This dataset contains house prices and the corresponding amenities, properties of the buildings and surrounding area. We will analyze the effect of such properties on the price as well as potential interactions between them.

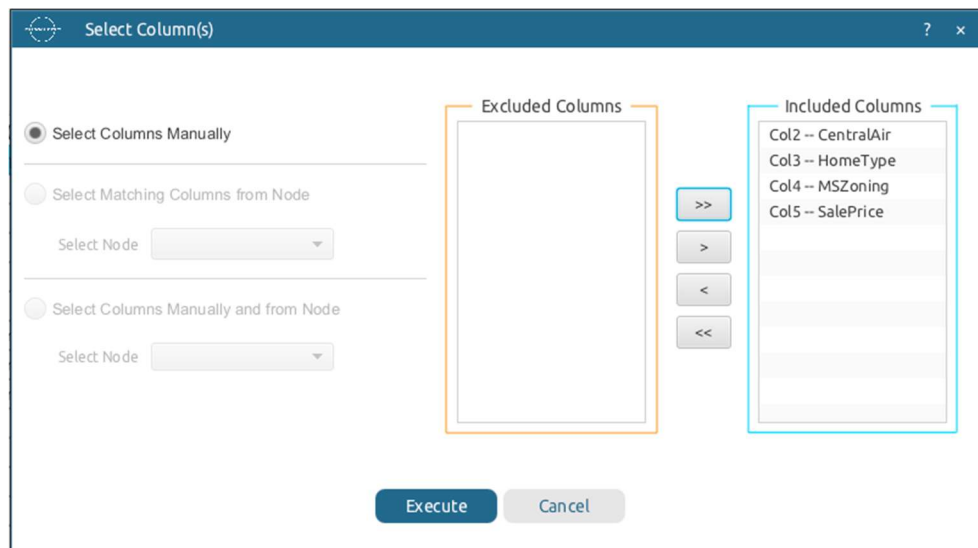
The variables examined are Sale Price as a dependent variable and Central AC, Building Type and Residential Density as factors. Dataset is modified to be used for Factorial ANOVA

*Isalos version used: 2.1.5*

Dataset: <https://www.kaggle.com/datasets/prevek18/ames-housing-dataset>

### Step 1: Data preparation

Import dataset and rename the tab to “Data”. Go to Data Transformation → Data Manipulation → Select Column(s), transfer all columns to the “Included Columns” and click “Execute”



|             | Col1        | Col2 (S)   | Col3 (S) | Col4 (S) | Col5 (I)  |
|-------------|-------------|------------|----------|----------|-----------|
| User Header | User Row ID | CentralAir | HomeType | MSZoning | SalePrice |
| 1           | 526301100   | Y          | 1Fam     | RL       | 215000    |
| 2           | 526351010   | Y          | 1Fam     | RL       | 172000    |
| 3           | 526353030   | Y          | 1Fam     | RL       | 244000    |
| 4           | 527105010   | Y          | 1Fam     | RL       | 189900    |
| 5           | 527105030   | Y          | 1Fam     | RL       | 195500    |
| 6           | 527127150   | Y          | MultiFam | RL       | 213500    |
| 7           | 527145080   | Y          | MultiFam | RL       | 191500    |
| 8           | 527146030   | Y          | MultiFam | RL       | 236500    |
| 9           | 527162130   | Y          | 1Fam     | RL       | 189000    |
| 10          | 527163010   | Y          | 1Fam     | RL       | 175900    |
| 11          | 527165230   | Y          | 1Fam     | RL       | 185000    |
| 12          | 527166040   | Y          | 1Fam     | RL       | 180400    |
| 13          | 527180040   | Y          | 1Fam     | RL       | 171500    |
| 14          | 527182190   | Y          | MultiFam | RL       | 212000    |
| 15          | 527216070   | Y          | 1Fam     | RL       | 538000    |
| 16          | 527225035   | Y          | 1Fam     | RL       | 164000    |
| 17          | 527258010   | Y          | 1Fam     | RL       | 394432    |
| 18          | 527276150   | Y          | 1Fam     | RL       | 141000    |

## Step 2a: Full Factorial ANOVA

Create a new tab named “Full Factorial” and import the spreadsheet from the “Data” tab. To perform the analysis, go to **Statistics** → **Analysis of (Co)Variance** → **Factorial ANOVA** and enter “95” for Confidence Level (%), “SalePrice” as Dependent Variable, move all columns to Factors and select “Full Factorial” as Formula Type.

Factorial ANOVA

Confidence Level (%)

95

Dependent Variable

Col5 -- SalePrice

Sum of Squares for Tests

Adjusted (Type III)

Coding for Factors

(-1, 0, +1)

Excluded Columns

Factors

Col2 -- CentralAir

Col3 -- HomeType

Col4 -- MSZoning

Formula Type:

☐ Custom
☐ Main Effects
☒ Full Factorial

Formula:

CentralAir + HomeType + MSZoning + CentralAir:HomeType + Centra

Residual Analysis & Assumption's Checks

Execute

Cancel

## Step 2b: Residual Analysis & Assumption's Checks

Click on Residual Analysis & Assumption's Checks and check the Shapiro-Wilk Test box under "Statistical Tests", Normal Q-Q Plots and Density Plot of Residuals boxes under "Plots". Check the Brown-Forsythe Test box under "Statistical Tests" and Violin Plot by Group box under "Plots". Click Save and then Execute.

**Residual Analysis & Assumption's Checks**

**Normality of Residuals**

Statistical Tests

- ☒ Shapiro-Wilk Test
- ☐ Anderson-Darling Test
- ☐ Kolmogorov-Smirnov Test

Plots

- ☒ Normal Q-Q Plots
- ☐ Histogram of Residuals
- ☒ Density Plot of Residuals

**Homogeneity of Variance (Homoscedasticity)**

Statistical Tests

- ☐ Levene's Test
- ☒ Brown-Forsythe Test
- ☐ Bartlett's Test (for normal data)

Plots

- ☐ Residuals vs Fitted Values
- ☐ Residuals vs Dependent Variable
- ☐ Boxplots by Group
- ☒ Violin Plot by Group

**Residual Diagnostics**

Plots

- ☐ Residuals vs Observation Order

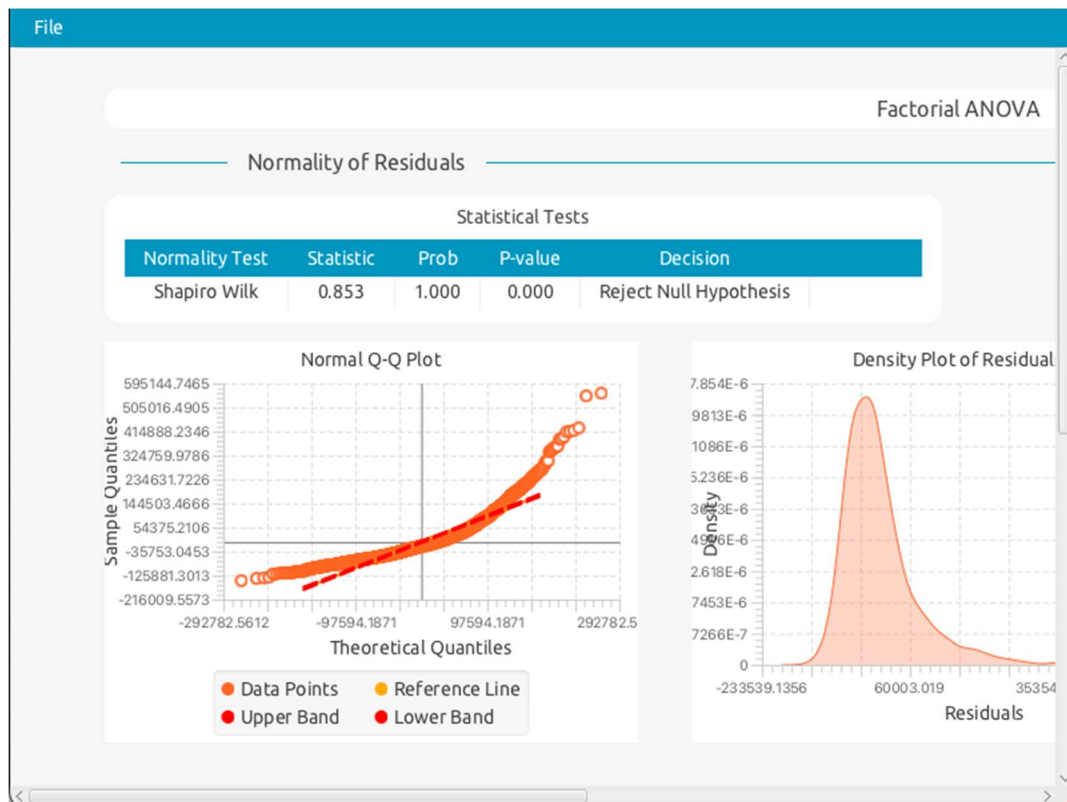
**Model Adequacy**

Plots

- ☐ Residuals vs Fitted values
- ☐ Residuals vs Each Factors

Save Cancel

A window with the tests results and requested plots appears:



Expand window to view all of the results and plots. Double-clicking on any plot opens it in a new window. The results are shown in the right spreadsheet:

|             | Col1        | Col2 (S)                     | Col3 (I) | Col4 (D)               | Col5 (D)               | Col6 (D)   | Col7 (D)  | Col8 (S)    |
|-------------|-------------|------------------------------|----------|------------------------|------------------------|------------|-----------|-------------|
| User Header | User Row ID | Source                       | DF       | Adj SS                 | Adj MS                 | F-Value    | P-Value   | Significant |
| 1           |             | CentralAir                   | 1        | 364930245<br>901.74805 | 364930245<br>901.74805 | 65.2528546 | 0E-7      | Yes         |
| 2           |             | HomeType                     | 1        | 125432669.1<br>171875  | 125432669.1<br>171875  | 0.0224285  | 0.8809639 | No          |
| 3           |             | MSZoning                     | 1        | 1151728402<br>35.92773 | 1151728402<br>35.92773 | 20.5939537 | 0.0000059 | Yes         |
| 4           |             | CentralAir*HomeType          | 1        | 601608641.<br>0761719  | 601608641.<br>0761719  | 0.1075731  | 0.7429494 | No          |
| 5           |             | CentralAir*MSZoning          | 1        | 6661748221<br>3.77539  | 6661748221<br>3.77539  | 11.9118131 | 0.0005663 | Yes         |
| 6           |             | HomeType*MSZoning            | 1        | 4134983601<br>6445312  | 4134983601<br>6445312  | 0.7393728  | 0.3899375 | No          |
| 7           |             | CentralAir*HomeType*MSZoning | 1        | 4395051546<br>.089844  | 4395051546<br>.089844  | 0.7858753  | 0.3754284 | No          |
| 8           |             | Error                        | 2727     | 1525090030<br>1916.748 | 5592556032<br>.972772  |            |           |             |
| 9           |             | Total                        | 2734     | 1768169007<br>4617.58  |                        |            |           |             |

## Step 3: Reduced Factorial ANOVA

Create a new tab named “Reduced” and import the spreadsheet from the “Data” tab. To perform the analysis, go to Statistics → Analysis of (Co)Variance → Factorial ANOVA and enter “95” for Confidence Level (%), “SalePrice” as Dependent Variable, move all columns to Factors and select “Custom” as Formula Type. Click Specify Custom Model Formula, select all columns from Included Columns and click Add interactions up to order 2, then, click Save. Click on Residual Analysis & Assumption's Checks and perform the tests like in Step 2b. Click Execute

Factorial ANOVA

Confidence Level (%)

95

Dependent Variable

Col5 -- SalePrice

Sum of Squares for Tests

Adjusted (Type III)

Coding for Factors

(-1, 0, +1)

Excluded Columns

Factors

Col2 -- CentralAir

Col3 -- HomeType

Col4 -- MSZoning

Formula Type:

☒ Custom
☐ Main Effects
☐ Full Factorial

Specify Custom Model Formula

Formula:

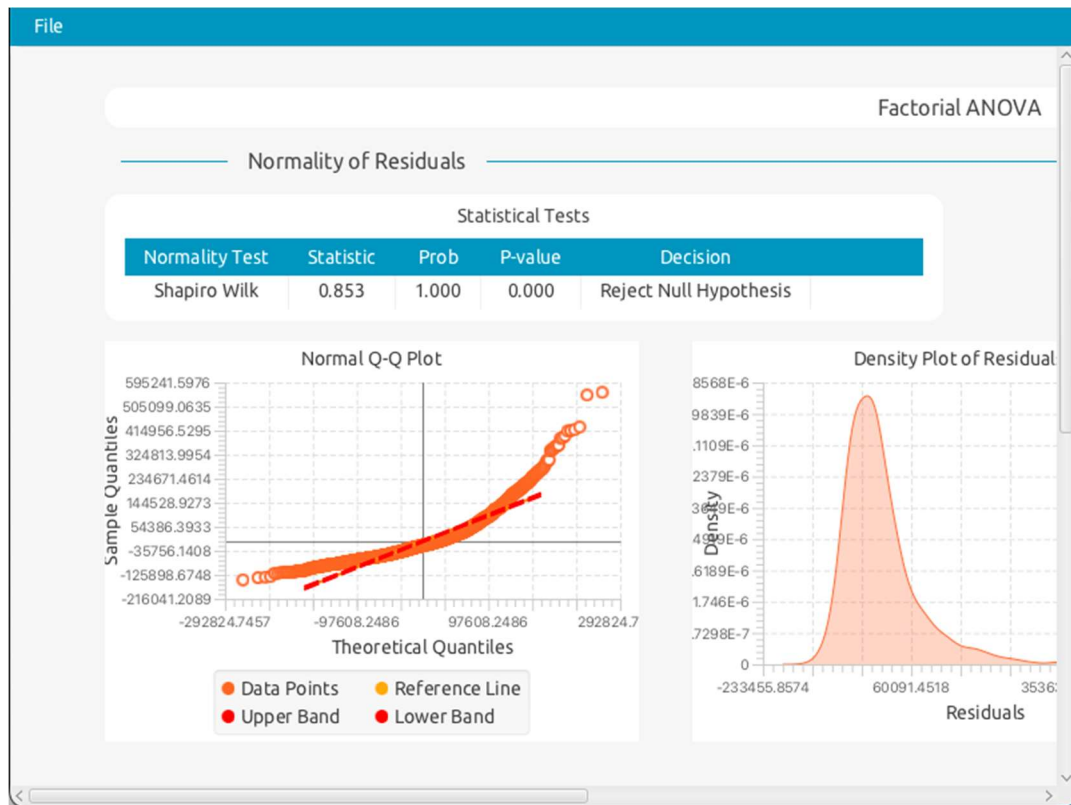
CentralAir + HomeType + MSZoning + CentralAir:HomeType + Centra

Residual Analysis & Assumption's Checks

Execute

Cancel

Test results and plots appear in a new window, analysis results are shown in right spreadsheet:



|             | Col1        | Col2 (S)            | Col3 (I) | Col4 (D)               | Col5 (D)               | Col6 (D)   | Col7 (D)  | Col8 (S)    |
|-------------|-------------|---------------------|----------|------------------------|------------------------|------------|-----------|-------------|
| User Header | User Row ID | Source              | DF       | Adj SS                 | Adj MS                 | F-Value    | P-Value   | Significant |
| 1           |             | CentralAir          | 1        | 3611421436<br>83.7832  | 3611421436<br>83.7832  | 64.5805765 | 0E-7      | Yes         |
| 2           |             | HomeType            | 1        | 2499177734<br>375      | 2499177734<br>375      | 0.0000447  | 0.9946666 | No          |
| 3           |             | MSZoning            | 1        | 1150355458<br>54.60352 | 1150355458<br>54.60352 | 20.5710189 | 0.0000060 | Yes         |
| 4           |             | CentralAir*HomeType | 1        | 1800813097<br>2089844  | 1800813097<br>2089844  | 0.3220271  | 0.5704387 | No          |
| 5           |             | CentralAir*MSZoning | 1        | 1070309652<br>45.23828 | 1070309652<br>45.23828 | 19.1396146 | 0.0000126 | Yes         |
| 6           |             | HomeType*MSZoning   | 1        | 3362037355<br>6.14453  | 3362037355<br>6.14453  | 6.0121012  | 0.0142704 | Yes         |
| 7           |             | Error               | 2728     | 1525529535<br>3462.838 | 5592117065<br>05236    |            |           |             |
| 8           |             | Total               | 2734     | 1768169007<br>4617.58  |                        |            |           |             |

## Step 4: Main Effects Factorial ANOVA

Create a new tab named “Main Effects” and import the spreadsheet from the “Data” tab. To perform the analysis, go to **Statistics** → **Analysis of (Co)Variance** → **Factorial ANOVA** and enter “95” for **Confidence Level (%)**, “SalePrice” as **Dependent Variable**, move all columns to **Factors** and select “Main Effects” as **Formula Type**. Click on **Residual Analysis & Assumption’s Checks** and perform the tests like in Step 2b. Click **Execute**

**Factorial ANOVA**

Confidence Level (%)

Dependent Variable

Sum of Squares for Tests

Coding for Factors

Excluded Columns

Factors

Col2 -- CentralAir  
Col3 -- HomeType  
Col4 -- MSZoning

Specify Reference Le...

Formula Type: ☐ Custom ☒ Main Effects ☐ Full Factorial

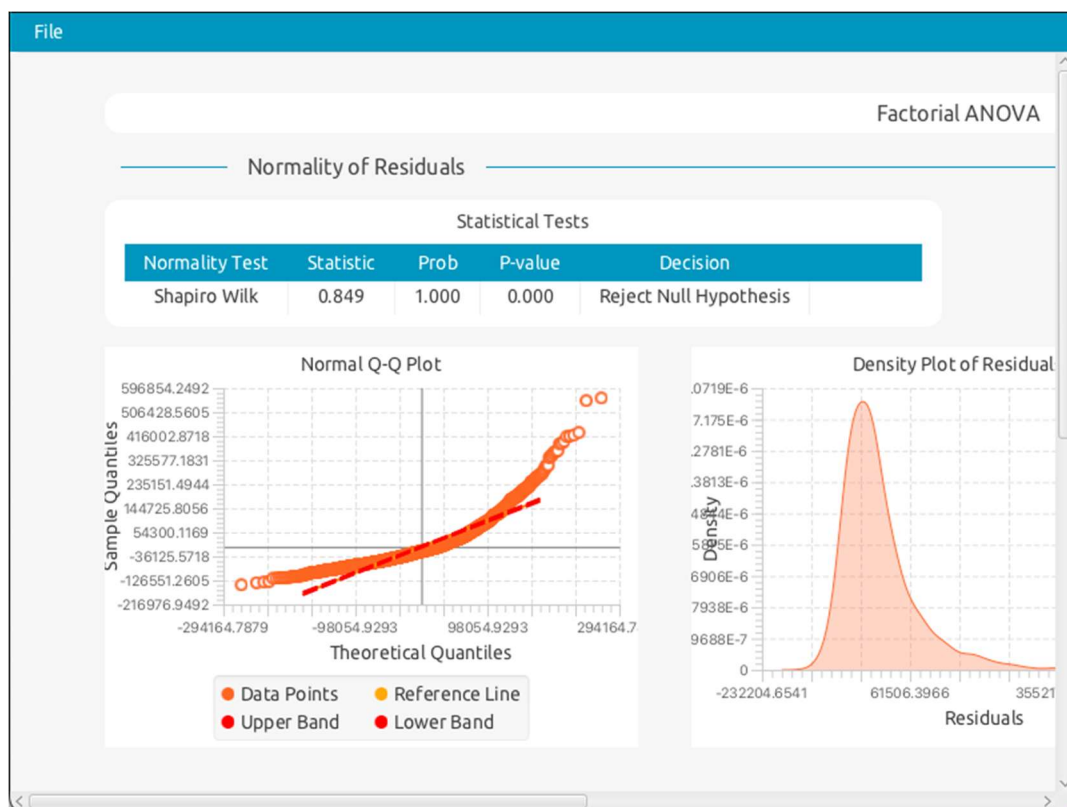
Specify Custom Model Formula

Formula:

Residual Analysis & Assumption's Checks

Execute Cancel

Test results and plots appear in a new window, analysis results are shown in right spreadsheet:



|             | Col1        | Col2 (S)   | Col3 (I) | Col4 (D)               | Col5 (D)               | Col6 (D)        | Col7 (D)  | Col8 (S)    |
|-------------|-------------|------------|----------|------------------------|------------------------|-----------------|-----------|-------------|
| User Header | User Row ID | Source     | DF       | Adj SS                 | Adj MS                 | F-Value         | P-Value   | Significant |
| 1           |             | CentralAir | 1        | 646835070<br>282.0293  | 646835070<br>282.0293  | 114.7436912     | 0.0       | Yes         |
| 2           |             | HomeType   | 1        | 2825250791<br>4.558594 | 2825250791<br>4.558594 | 5.0117831       | 0.0252556 | Yes         |
| 3           |             | MSZoning   | 1        | 1069497417<br>969.9062 | 1069497417<br>969.9062 | 189.720822<br>4 | 0.0       | Yes         |
| 4           |             | Error      | 2731     | 1539523923<br>1039.35  | 5637216855<br>.012578  |                 |           |             |
| 5           |             | Total      | 2734     | 1768169007<br>4617.58  |                        |                 |           |             |

## Final Isalos Workflow:

