



Two-Way ANOVA analysis on individual medical costs with age and sex as predictors

This dataset contains individual medical costs billed by health insurance for individuals of various ages. We will analyze the effect of select factors on the medical cost as well as potential interactions between them.

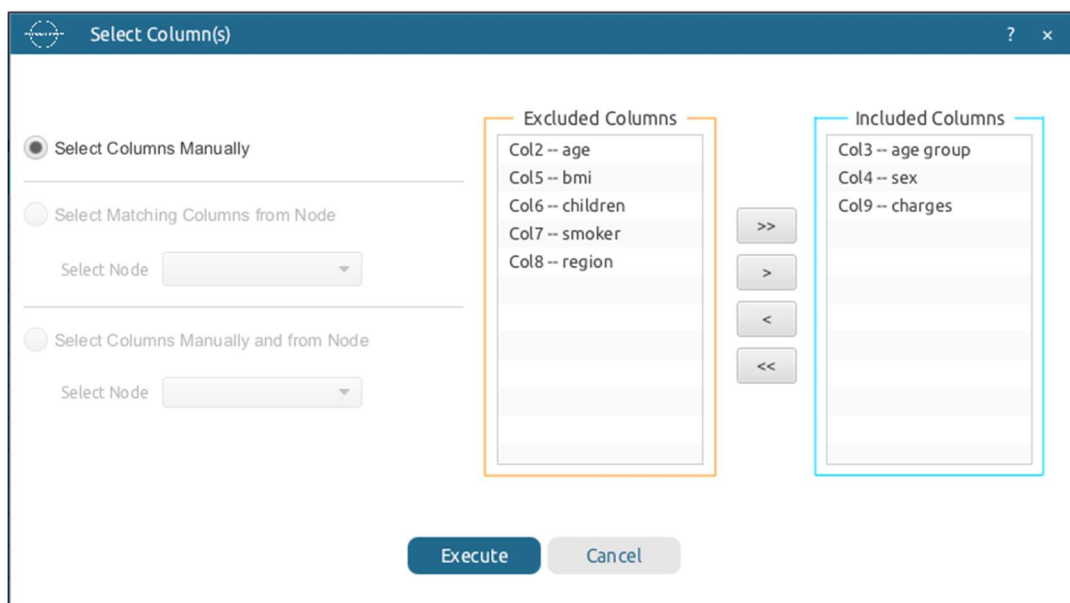
The variables examined are Charges as a dependent variable and Sex, Age as factors. The dataset is modified to be used for Two-Way ANOVA

Isalos version used: 2.1.2

Dataset: <https://www.kaggle.com/datasets/mirichoi0218/insurance>

Step 1: Data preparation

Import dataset and rename the tab to “Data”. Go to Data Transformation → Data Manipulation → Select Column(s), transfer “Col3 – age group”, “Col4 -- sex”, “Col9 -- charges” to the “Included Columns” and click “Execute”



	Col1	Col2 (S)	Col3 (S)	Col4 (D)
User Header	User Row ID	age group	sex	charges
1		18-24	male	1725.5523
2		18-24	male	1163.4627
3		18-24	female	14283.4594
4		18-24	male	11272.33139
5		18-24	male	1720.3537
6		18-24	male	1727.54
7		18-24	female	21344.8467
8		18-24	male	1708.92575
9		18-24	female	38792.6856
10		18-24	female	1621.8827
11		18-24	male	1136.3994
12		18-24	male	1131.5066
13		18-24	male	36307.7983
14		18-24	male	1146.7966
15		18-24	female	2217.6012
16		18-24	male	1712.227
17		18-24	male	34617.84065
18		18-24	female	1634.5734
19		18-24	female	2801.2588
20		18-24	male	33732.6867
21		18-24	male	2304.0022
22		18-24	male	1121.8739

Step 2a: Full Two-Way ANOVA

Create a new tab named “Interactions” and import the spreadsheet from the “Data” tab. To perform the analysis, go to Statistics → Analysis of (Co)variance → Two-Way ANOVA and select “age group” as Factor A, “sex” as Factor B and “charges” as Response Variable.

Step 2b: Residual Analysis & Assumption's Checks

Click on Residual Analysis & Assumption's Checks and check the Anderson-Darling Test box under “Statistical Tests”, Normal Q-Q Plots and Histogram of Residuals boxes under “Plots”. Check the Levene's Test box under “Statistical Tests” and Residuals vs Fitted Values 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

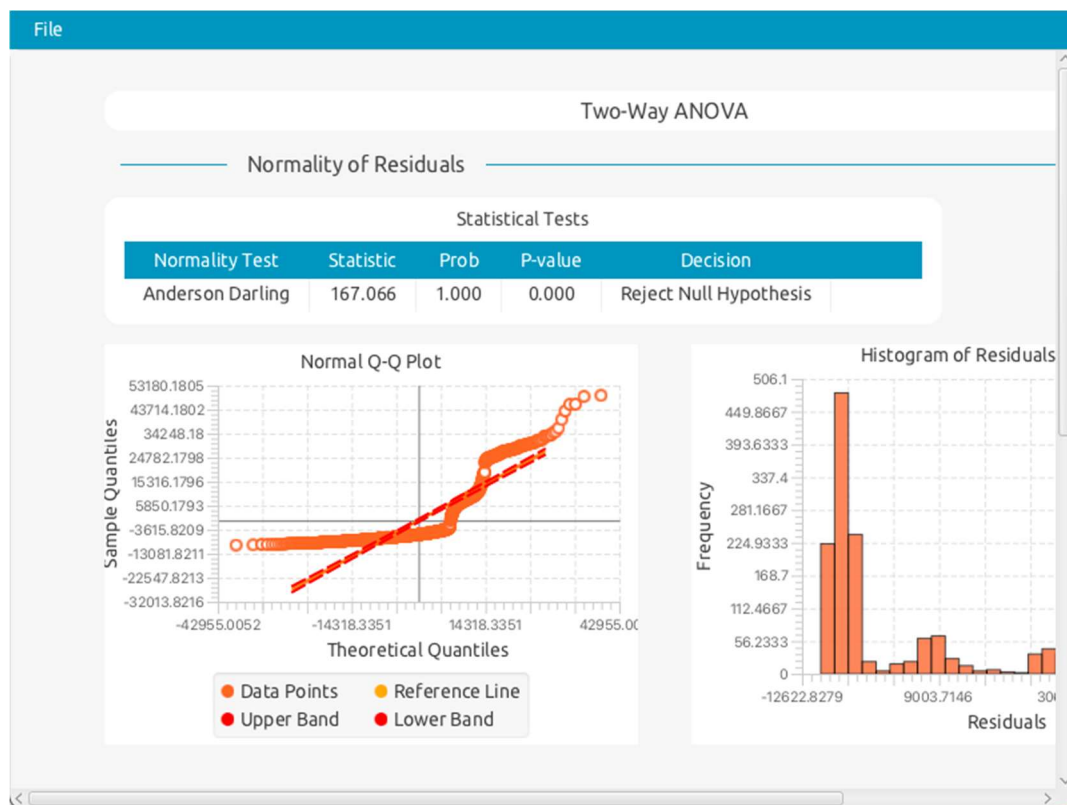
Model Adequacy

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)
User Header	User Row ID	Source	DF	Adj SS	Adj MS	F-value	P-value
1		Factor A: age group	4	1584194507 4.401525	396048626 8.6003814	29.3024468	0.0
2		Factor B: sex	1	660149998. 1106205	660149998. 1106205	4.8842513	0.0272729
3		Interaction: age group * sex	4	80897813.4 513512	20224453.3 628378	0.1496346	0.9631909
4		Error	1327	1793558506 18.06378	135158892.7 038913		
5		Total	1336	1959388435 04.02728			

Step 3: Main Effects Two-Way 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 (%)**, “charges” 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 -- age group
Col3 -- sex

Specify Reference Le...

Formula Type: ☐ Custom ☒ Main Effects ☐ Full Factorial

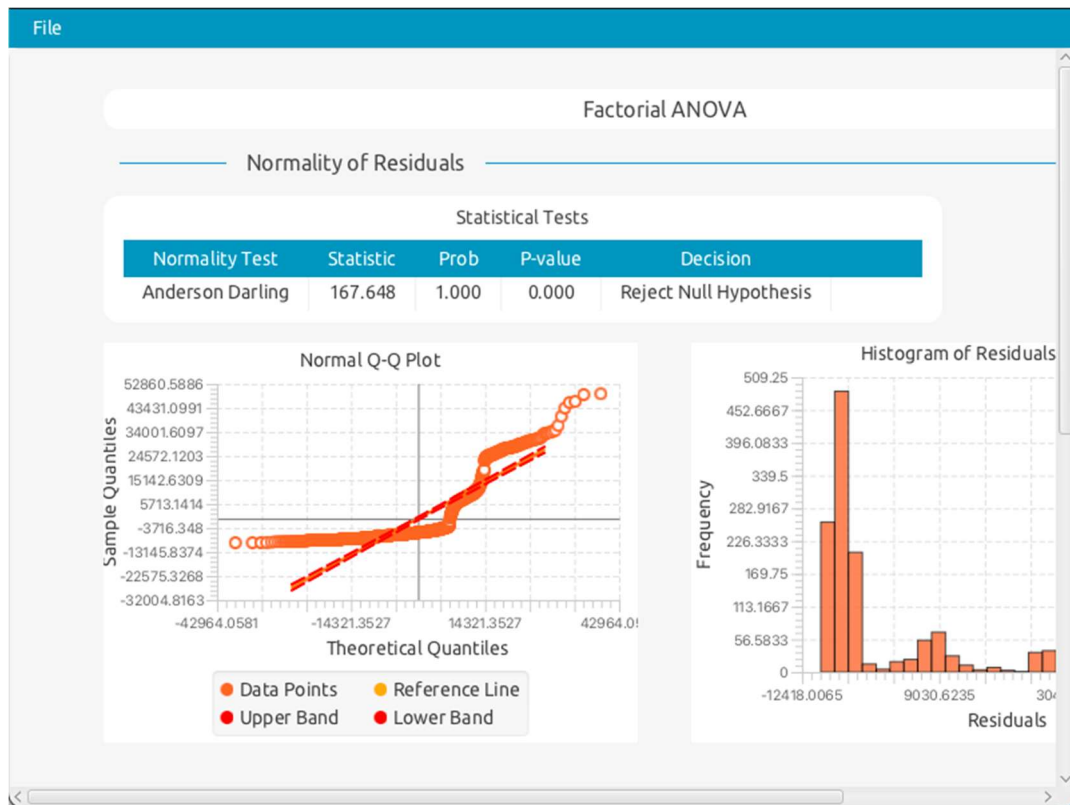
Specify Custom Model Formula

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		age group	4	1596317302 1.27826	399079325 5.319565	29.6223428	0.0	Yes
2		sex	1	781377944. 9877014	781377944. 9877014	5.7999109	0.0161622	Yes
3		Error	1331	1793155204 84.63818	134722404. 5714787			
4		Total	1336	1959388435 04.0271				

Final Isalos Workflow:

