



Car Price Prediction with ANCOVA in Isalos

The purpose of this research paper is to provide a comprehensive, step-by-step guide for conducting an Analysis of Covariance (ANCOVA) using the Isalos Analytics Platform to predict used car prices. The methodology outlined herein enables analysts to estimate the value of a vehicle based on its specifications and condition, integrating both categorical factors and continuous covariates within a unified statistical framework.

ANCOVA extends ANOVA by incorporating continuous predictors, which increases statistical power and reduces bias arising from pre-existing group differences. The technique adjusts the dependent variable for the effect of covariates, thereby reducing within-group variance and enhancing the ability to detect significant differences among categorical groups

Isalos version used: 2.1.5

The dataset used in this analysis is the **Car Price Prediction Dataset**, available on Kaggle.

Source: [Car Pricing Regression Dataset](#)

This dataset contains detailed information about used cars, including key features such as make, model, year of manufacture, engine size, mileage, fuel type, transmission, and price. It is designed to support data analysis, visualisation, and machine learning tasks—particularly regression models aimed at predicting car prices.

For the purposes of this ANCOVA analysis, variables are classified as follows:

- **Dependent Variable (Response):** Price — the continuous outcome to be predicted.
- **Factors (Categorical Predictors):** Make, Model, Fuel Type, Transmission — categorical variables whose effect on price is of primary interest.
- **Covariates (Continuous Predictors):** Year, Engine Size, Mileage — continuous variables to be controlled for in the analysis.

Step 1: Import Data from file

Firstly, launch the Isalos Analytics Platform (version 2.1.2). Then, in the left spreadsheet panel, create a new tab by clicking the file and choose open the file `car_price_prediction` (or just copy-paste the data from your spreadsheet).

Step 3: Configure ANCOVA parameters

The ANCOVA model can be specified with varying levels of complexity:

The model :

$$\text{Price} = \beta_0 + \beta_1(\text{Year}) + \beta_2(\text{Engine Size}) + \beta_3(\text{Mileage}) + \beta_4(\text{Make}) + \beta_5(\text{Model}) + \beta_6(\text{Fuel Type}) + \beta_7(\text{Transmission}) + \epsilon$$

Important: The inclusion of interaction terms should be guided by theoretical considerations and exploratory data analysis. ANCOVA assumes homogeneity of regression slopes—if this assumption is violated, interaction terms may be necessary.

Steps in Isalos Analytics Platform:

In the above window choose the following :

- Confidence Level (%) : 95
- Dependent Variable : Col9 – Price
- And then move :
 - Col2 – Make , Col3 – Model, Col7 – Fuel Type, Col8 – Transmission **to Factors**
 - Col4 – Year, Col5 – Engine Size, Col6 – Mileage **to Covariates**.

Then choose Formula Type: Main Effects and you get the Formula as the tab below:

ANCOVA

Confidence Level (%) : 95

Dependent Variable : Col9 -- Price

Sum of Squares for Tests : Adjusted (Type III)

Coding for Factors : (-1, 0, +1)

Excluded Columns

Factors

- Col2 -- Make
- Col3 -- Model
- Col7 -- Fuel Type
- Col8 -- Transmission

Specify Reference Levels

Covariates

- Col4 -- Year
- Col5 -- Engine Size
- Col6 -- Mileage

Formula Type: ☐ Custom ☒ Main Effects ☐ Full Factorial

Specify Custom Model Formula

Formula: Year + Engine Size + Mileage + Make + Model + Fuel Type + Transmission

Execute Cancel

After configuring all parameters:

- * Review all selections to ensure accuracy.
- * Click the [Execute](#) button to run the ANCOVA analysis.
- * Wait for the computation to complete (processing time depends on dataset size).

Step 4: Results

The analysis results will be displayed in the right spreadsheet panel. The output typically includes the following components:

	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		Year	1	10279058375.104025	10279058375.104025	2366.6671230	0.0	Yes
2		Engine Size	1	3887021946.7368703	3887021946.7368703	894.9542567	0.0	Yes
3		Mileage	1	8383325868.409044	8383325868.409044	1930.1905865	0.0	Yes
4		Make	4	8011089.2964120	2002772.3241030	0.4611216	0.7643075	No
5		Model	4	17772017.9355178	4443004.4838794	1.0229646	0.3942902	No
6		Fuel Type	2	978968.7140894	489484.3570447	0.1126997	0.8934305	No
7		Transmission	1	3194.6423974	3194.6423974	0.0007355	0.9783689	No
8		Error	985	4278114315.752788	4343263.2647236			
9		Total	999	26820073213.693928				
10								

Step 5 : Residual Analysis & Assumption's checks

Choose

Statistics → Analysis of (Co)Variance → ANCOVA

and choose

Residual Analysis & Assumption's Checks.

You get the following window:

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

Residual Diagnostics

Plots

☐ Residuals vs Observation Order

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

Model Adequacy

Plots

☐ Residuals vs Fitted values
 ☐ Residuals vs Each Factors
 ☐ Residuals vs Covariates

Save

Cancel

then choose the tests you want. For example, choose the following:

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

Residual Diagnostics

Plots

☐ Residuals vs Observation Order

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

Model Adequacy

Plots

☐ Residuals vs Fitted values
 ☐ Residuals vs Each Factors
 ☐ Residuals vs Covariates

Save

Cancel

After that, click execute and you get the results (some of them below):

5

