



Parkinson's Disease Dataset Analysis

The purpose of this research paper is to provide a comprehensive, step-by-step guide for conducting **ANOVA analyses** using the Isalos Analytics Platform. These statistical methods are widely used in healthcare research to identify the effects of demographic and lifestyle factors on disease severity, cognitive performance, and functional impairment.

Isalos version used: 2.1.5

The dataset used in this analysis is the **Parkinson's Disease Dataset Analysis**, available on Kaggle.

Source: [Parkinson's Disease Dataset Analysis](#)

This dataset simulates real-world healthcare records and includes comprehensive information for **2,105** patients diagnosed with Parkinson's Disease. It contains demographic details (age, gender, ethnicity, education), lifestyle factors (smoking, alcohol consumption, physical activity, diet quality), and clinical measurements (UPDRS, MoCA, Functional Assessment). It is designed to support predictive modelling and risk assessment for Parkinson's disease progression and cognitive decline.

Step 1: Import Data from file

Firstly, launch the Isalos Analytics Platform (version 2.1.5). Then, in the left spreadsheet panel, create a new tab by clicking **File** and choose **Open** to load the Parkinson's Disease dataset (or simply copy-paste the data from your spreadsheet).

Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (D)	Col9 (D)	Col10 (D)	
User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption	PhysicalActivity	DietQuality	
3058	85	0	3	1	19.61987796	0	5.108240607	1.380659917	3.893969135	
3059	75	0	0	2	16.24733916	1	6.027648029	8.40980405	8.51342825	
3060	70	1	0	0	15.36823871	0	2.242135331	0.213274591	6.498804606	
3061	52	0	0	0	15.45455733	0	5.997787563	1.375045164	6.715033333	
3062	87	0	0	1	18.61604177	0	9.775242923	1.188607062	4.657572037	
3063	68	1	2	1	39.42331141	1	13.5968889	7.796704004	7.070238878	
3064	78	1	0	0	30.54200329	1	2.011281313	9.028536304	9.838445926	
3065	70	1	0	0	36.75828161	1	19.98886597	3.891748622	3.421960006	
3066	80	0	2	1	22.3805865	1	7.293287715	2.595670177	4.784827139	
3067	71	0	3	2	23.72708628	1	17.78290985	7.344890316	3.393018462	
3068	70	0	0	3	38.48254474	0	6.639761993	7.872186697	9.225710184	
3069	53	0	0	1	35.8967387	1	5.212906263	7.185203018	7.918912222	
3070	74	0	2	2	30.22551208	0	3.762918066	4.316651243	5.112520425	
3071	87	1	1	2	38.29830655	0	12.61599522	9.299289996	6.71557911	
3072	58	0	0	3	34.96532287	1	11.70859735	4.392463152	5.182038871	
3073	56	1	0	0	18.95878125	1	2.04712047	9.432830343	1.720277262	
3074	54	1	0	0	28.04958007	1	7.260733832	4.311617423	2.509936233	

Step 2: Access the ANCOVA function

Navigate to the ANCOVA analysis module through the main menu by choosing:

Statistics → Analysis of (Co)Variance → ANCOVA

The ANCOVA configuration window will appear:

Step 3: Configure ANCOVA parameters

The ANCOVA model we will estimate is:

$$\text{BMI} = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{Gender}) + \beta_3(\text{Smoking}) + \beta_4(\text{Gender} \times \text{Smoking}) + \varepsilon$$

This model predicts BMI (continuous) based on the covariate Age, the two categorical factors (Gender and Smoking), and their **interaction**.

In the ANCOVA configuration window, fill in the fields exactly as follows:

ANCOVA

Confidence Level (%) 95

Dependent Variable Col7 -- BMI

Sum of Squares for Tests Adjusted (Type III)

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

Excluded Columns

- Col2 -- PatientID
- Col6 -- EducationLevel
- Col9 -- AlcoholConsumption
- Col10 -- PhysicalActivity
- Col11 -- DietQuality
- Col5 -- Ethnicity

> <

Factors

- Col4 -- Gender
- Col8 -- Smoking

Specify Reference Levels

Covariates

- Col3 -- Age

Formula Type: ☐ Custom ☐ Main Effects ☒ Full Factorial

Specify Custom Model Formula

Formula: Age + Gender + Smoking + Age:Gender + Age:Smoking + Gender:Smoking +

Execute Cancel

After configuring all parameters:

- * Review all selections to ensure accuracy.
- * Click the **Execute** button to run the descriptive analysis.
- * Wait for the computation to complete (processing time depends on dataset size and the number of dependent variables).

Step 4: Results

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

Source	DF	Adj SS	Adj MS	F-Value	P-Value	Significant
Age	1	124.0047029	124.0047029	2.3957645	0.1218152	No
Gender	1	1.5833659	1.5833659	0.0305905	0.8611739	No
Smoking	1	1.9316053	1.9316053	0.0373185	0.8468367	No
Age*Gender	1	8.1105965	8.1105965	0.1566963	0.6922568	No
Age*Smoking	1	1.4292327	1.4292327	0.0276127	0.8680386	No
Gender*Smoking	1	3.3016479	3.3016479	0.0637877	0.8006313	No
Age*Gender*Smoking	1	9.2546383	9.2546383	0.1787991	0.6724506	No
Error	2097	108540.6608756	51.7599718			
Total	2104	109316.8762629				

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:



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
 ☐ Residuals vs Covariates

Save

Cancel

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

Residual Analysis & Assumption's Checks

?

X

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
- ☒ Residuals vs Covariates

Save

Cancel

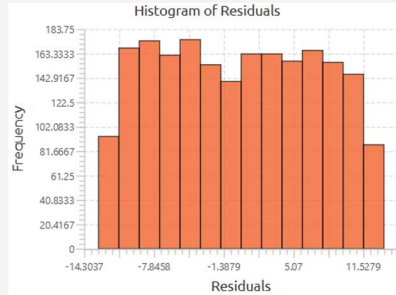
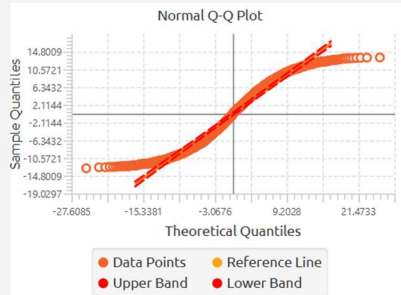
After that, click execute and you get the results (some of them below):

ANCOVA

Normality of Residuals

Statistical Tests

Normality Test	Statistic	Prob	P-value	Decision
Shapiro Wilk	0.957	1.000	0.000	Reject Null Hypothesis
Anderson Darling	23.840	1.000	0.000	Reject Null Hypothesis
Kolmogorov Smirnov	0.065	1.000	0.000	Reject Null Hypothesis



Homogeneity of Variance

Statistical Test (Univariate)

Test	F-statistic	df1	df2	P-value	Decision
Levene	0.198	3	2101	0.898	Fail to Reject Null Hypothesis
Brown - Forsythe	0.227	3	2101	0.878	Fail to Reject Null Hypothesis
Bartlett	0.245	3		0.970	Fail to Reject Null Hypothesis

