



Health Care - Predicting Chronic Disease Risk

The purpose of this research paper is to provide a comprehensive, step-by-step guide for conducting a **Multivariate Analysis of Covariance (MANCOVA)** using the Isalos Analytics Platform. MANCOVA extends ANCOVA by simultaneously modelling **multiple continuous dependent variables**, allowing analysts to assess how categorical factors and continuous covariates jointly influence a set of interrelated health outcomes. This methodology is particularly valuable in healthcare research, where physiological markers are often correlated and should be analysed together to avoid inflating Type I error.

MANCOVA combines the principles of multivariate ANOVA (MANOVA) and linear regression. By including continuous covariates, the procedure adjusts the dependent variables for the effects of confounding variables (e.g., age), thereby increasing statistical power and providing more accurate estimates of group differences while controlling for individual variability.

Isalos version used: 2.1.5

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

Source: [HealthCare](#)

This dataset simulates real-world healthcare records and includes patient-level information such as demographic characteristics, lifestyle factors (smoking status, physical activity level), and clinical measurements (cholesterol, triglycerides, blood pressure). It is designed to support predictive modelling and risk assessment for chronic diseases.

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

➤ **Dependent Variables (Responses):**

- *Cholesterol*
- *Triglycerides*

- *Blood Pressure*
— these are continuous outcome measures to be predicted simultaneously.

➤ **Factors (Categorical Predictors):**

- *Smoking Status* (e.g., Smoker, Non-smoker)
- *Physical Activity* (e.g., Low, Moderate, High)
— these categorical variables represent lifestyle groups whose multivariate effect on the health outcomes is of primary interest.

➤ **Covariates (Continuous Predictors):**

- *Age* (in years) — this continuous variable is controlled for to remove its confounding influence on the dependent variables.

Step 1: Import Data from file

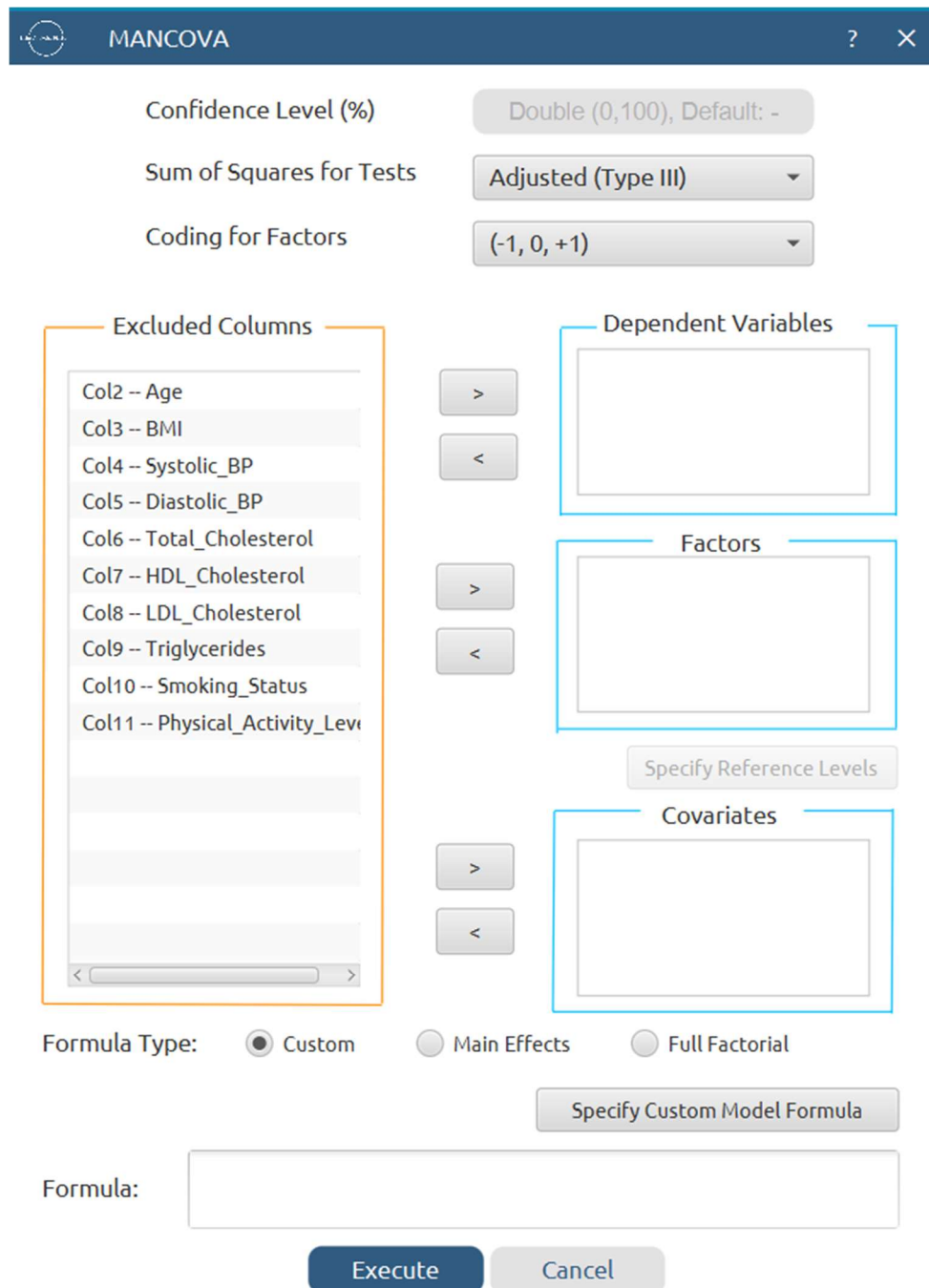
Firstly, launch the Isalos Analytics Platform (version 2.1.2). Then, in the left spreadsheet panel, create a new tab by clicking **File** and choose **Open** to load the healthcare dataset (or simply copy-paste the data from your spreadsheet).

	Col1	Col2 (I)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (S)	Col11 (S)
User Header	User Row ID	Age	BMI	Systolic_BP	Diastolic_BP	Total_Cholesterol	HDL_Cholesterol	LDL_Cholesterol	Triglycerides	Smoking_Status	Physical_Activity_Level
1	0	69	20.71605856	107.1484378	58.25150312	168.9185111	38.41767411	183.8350619	243.3415266	non-smoker	moderate
2	1	32	26.72023069	102.9483265	86.64159761	203.9120143	39.61157136	209.0845952	266.3754245	smoker	high
3	2	89	25.97163564	116.3270326	90.282522	213.5648167	39.40622256	245.8848312	256.9745251	non-smoker	moderate
4	3	78	29.58078818	104.2738	85.58239162	204.3019143	33.02688134	207.4936471	257.7178209	non-smoker	moderate
5	4	38	28.35313009	104.052584	81.32626963	195.7671433	34.5588417	213.8267368	246.5134004	non-smoker	moderate
6	5	41	27.99637206	108.0406024	76.96239823	209.9205489	30.07611828	242.8432338	247.8182192	smoker	moderate
7	6	20	31.8345508	111.3742138	84.41643937	190.2644671	30.70454858	219.3706415	247.0353627	non-smoker	low
8	7	39	24.85020864	101.2783479	79.69000733	173.177954	40.84672701	227.8676114	233.3656432	smoker	high
9	8	70	29.6936226	103.8973874	90.32200171	207.0352856	41.72652826	225.1478171	243.5505893	non-smoker	high
10	9	19	28.4043326	94.79826619	81.77350065	212.8369979	37.48954601	219.5703095	271.2561166	non-smoker	high
11	10	47	27.56972725	108.9766879	75.17545959	174.1897904	35.20735554	205.9701063	273.6107119	smoker	high
12	11	55	26.67147594	104.6732216	84.28814902	176.7064396	39.17127736	199.0488471	268.1764309	smoker	low
13	12	19	20.64359728	118.8595887	76.46472812	194.6331945	40.42664249	220.0423365	259.4321243	non-smoker	moderate
14	13	81	25.73043682	108.7268169	88.15288356	218.8406267	33.38691287	225.1802907	243.5895551	non-smoker	high
15	14	77	26.05103067	111.7757652	82.06919921	205.0012397	33.49923567	237.5538942	257.6523818	non-smoker	moderate
16	15	38	23.38637655	114.4239333	94.35251788	187.3657695	38.53334565	204.3101446	241.4443639	non-smoker	low
17	16	50	20.21057494	115.2734137	81.27894994	179.5394275	29.04048094	215.7118832	240.4004145	smoker	moderate

Step 2: Access the MANCOVA function

Navigate to the MANCOVA analysis module through the main menu by choosing the following path: **Statistics → Analysis of (Co)Variance → MANCOVA**

Then, you will see the MANCOVA configuration window:



The MANCOVA configuration window is a dialog box with a dark blue header bar containing the title "MANCOVA", a question mark icon, and a close button (X). The main area is white and contains several configuration options:

- Confidence Level (%)**: A text label followed by a button labeled "Double (0,100), Default: -".
- Sum of Squares for Tests**: A text label followed by a dropdown menu showing "Adjusted (Type III)".
- Coding for Factors**: A text label followed by a dropdown menu showing "(-1, 0, +1)".
- Excluded Columns**: A list box containing the following items:
 - Col2 -- Age
 - Col3 -- BMI
 - Col4 -- Systolic_BP
 - Col5 -- Diastolic_BP
 - Col6 -- Total_Cholesterol
 - Col7 -- HDL_Cholesterol
 - Col8 -- LDL_Cholesterol
 - Col9 -- Triglycerides
 - Col10 -- Smoking_Status
 - Col11 -- Physical_Activity_Level
- Dependent Variables**: A text label above an empty rectangular box.
- Factors**: A text label above an empty rectangular box.
- Covariates**: A text label above an empty rectangular box.
- Formula Type**: Three radio buttons labeled "Custom" (selected), "Main Effects", and "Full Factorial".
- Specify Custom Model Formula**: A button located below the radio buttons.
- Formula:**: A text label followed by a large empty text input field.
- Execute** and **Cancel**: Two buttons at the bottom of the window.

Step 3: Configure MANCOVA parameters

The MANCOVA model can be specified with varying levels of complexity. The model is expressed as:

$$[\text{Cholesterol, Triglycerides, Blood Pressure}] = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{Smoking Status}) + \beta_3(\text{Physical Activity}) + \varepsilon$$

This is a multivariate linear model where the three dependent variables are regressed onto the covariate (Age) and the two categorical factors.

Important: The inclusion of interaction terms (e.g., Smoking Status $\times$ Physical Activity) should be guided by theoretical considerations and exploratory data analysis. MANCOVA assumes homogeneity of regression slopes across groups—if this assumption is violated, interaction terms may be necessary. For this basic demonstration, we focus on the **main-effects model**.

Steps in Isalos Analytics Platform:

In the MANCOVA configuration window, set the following parameters:

- * Confidence Level (%) : 95
- * **Dependent Variables:** Select
 - Col3 – BMI,
 - Col4 — Systolic_BP,
 - Col5 – Diastolic_BP ,
 - Col6 – Total_Cholesterol,
 - Col7 – HDL_Cholesterol,
 - Col8 – LDL_Cholesterol,
 - Col9– Triglycerides,

and move them to the **Dependent Variables** field.

- * **Factors (Categorical Predictors):** Move
 - Col10 – Smoking Status and
 - Col11 – Physical Activity Level

to the **Factors** field.

- * **Covariates (Continuous Predictors):** Move

- Col2 – Age

to the **Covariates** field.

Then, for **Formula Type**, choose **Full Factorial**. The formula preview will display the multivariate model as shown below:

MANCOVA

Confidence Level (%) 95

Sum of Squares for Tests Adjusted (Type III)

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

Excluded Columns

Dependent Variables

- Col3 -- BMI
- Col4 -- Systolic_BP
- Col5 -- Diastolic_BP
- Col6 -- Total_Cholesterol

Factors

- Col10 -- Smoking_Status
- Col11 -- Physical_Activity_Level

Specify Reference Levels

Covariates

- Col2 -- Age

Formula Type: ☐ Custom ☐ Main Effects ☒ Full Factorial

Specify Custom Model Formula

Formula: Age + Smoking_Status + Physical_Activity_Level + Age:Smoking_Status + A

Execute Cancel

After configuring all parameters:

- * Review all selections to ensure accuracy.
- * Click the **Execute** button to run the MANCOVA 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	Dependent Variable	DF	Adj SS	Adj MS	F-Value	P-Value	Significant
Age	BMI	1	10.7802015	10.7802015	0.4276238	0.5131632	No
	Systolic_BP	1	77567.6089904	77567.6089904	738.4615588	0.0	Yes
	Diastolic_BP	1	306.3900798	306.3900798	4.3747993	0.0364842	Yes
	Total_Cholesterol	1	348406.1373341	348406.1373341	1342.9886634	0.0	Yes
	HDL_Cholesterol	1	99.0999997	99.0999997	3.9164866	0.0478257	Yes
	LDL_Cholesterol	1	129051.0153109	129051.0153109	667.9874627	0.0	Yes
	Triglycerides	1	44649.6031603	44649.6031603	94.4668874	0.0	Yes
Smoking_Status	BMI	1	3.1627076	3.1627076	0.1254568	0.7231929	No
	Systolic_BP	1	106.6053960	106.6053960	1.0149080	0.3137397	No
	Diastolic_BP	1	0.4750033	0.4750033	0.0067823	0.9343651	No
	Total_Cholesterol	1	0.0028816	0.0028816	0.0000111	0.9973408	No
	HDL_Cholesterol	1	30.5045377	30.5045377	1.2055562	0.2722246	No
	LDL_Cholesterol	1	0.1855756	0.1855756	0.0009606	0.9752754	No
	Triglycerides	1	202.7272946	202.7272946	0.4289180	0.5125264	No
Physical_Activity_Level	BMI	2	41.2424269	20.6212134	0.8179923	0.4413286	No
	Systolic_BP	2	51.9798137	25.9899068	0.2474299	0.7808069	No
	Diastolic_BP	2	45.7049602	22.8524801	0.3262998	0.7215919	No
	Total_Cholesterol	2	215.9180589	107.9590294	0.4161458	0.6595887	No
	HDL_Cholesterol	2	6.4349293	3.2174647	0.1271560	0.8805969	No
	LDL_Cholesterol	2	86.2978714	43.1489357	0.2233454	0.7998401	No
	Triglycerides	2	1164.3834053	582.1917026	1.2317654	0.2917947	No
Age*Smoking_Status	BMI	1	0.1835336	0.1835336	0.0072803	0.9320039	No
	Systolic_BP	1	139.7215804	139.7215804	1.3301817	0.2487839	No
	Diastolic_BP	1	0.3013452	0.3013452	0.0043028	0.9477004	No
	Total_Cholesterol	1	13.6882156	13.6882156	0.0527635	0.8183244	No
	HDL_Cholesterol	1	62.2834691	62.2834691	2.4614770	0.1166817	No
	LDL_Cholesterol	1	5.6750532	5.6750532	0.0293749	0.8639176	No
	Triglycerides	1	186.0591877	186.0591877	0.3936526	0.5303914	No

Age*Smoking_Status*Physical_Activity_Level	BMI	2	12.0203372	6.0101686	0.2384085	0.7878826	No
	Systolic_BP	2	265.9061197	132.9530598	1.2657438	0.2820475	No
	Diastolic_BP	2	29.6047250	14.8023625	0.2113559	0.8094873	No
	Total_Cholesterol	2	335.0411962	167.5205981	0.6457357	0.5242855	No
	HDL_Cholesterol	2	6.3848520	3.1924260	0.1261664	0.8814687	No
	LDL_Cholesterol	2	44.1989353	22.0994676	0.1143902	0.8919104	No
	Triglycerides	2	1301.5700047	650.7850024	1.3768909	0.2523811	No
Corrected Model	BMI	11	119.3007059	10.8455187	0.4302148	0.9434145	No
	Systolic_BP	11	110823.3202365	10074.8472942	95.9148739	0.0	Yes
	Diastolic_BP	11	796.8040777	72.4367343	1.0342899	0.4122694	No
	Total_Cholesterol	11	477111.6632387	43373.7875672	167.1913860	0.0	Yes
	HDL_Cholesterol	11	266.6111711	24.2373792	0.9578746	0.4828869	No
	LDL_Cholesterol	11	174934.0582334	15903.0962030	82.3168176	0.0	Yes
	Triglycerides	11	64833.7772432	5893.9797494	12.4701203	0.0	Yes
Age*Physical_Activity_Level	BMI	2	50.0403267	25.0201634	0.9924877	0.3706681	No
	Systolic_BP	2	5.1908675	2.5954338	0.0247091	0.9755937	No
	Diastolic_BP	2	63.0215225	31.5107612	0.4499273	0.6376797	No
	Total_Cholesterol	2	542.3529168	271.1764584	1.0452942	0.3516037	No
	HDL_Cholesterol	2	0.0345906	0.0172953	0.0006835	0.9993167	No
	LDL_Cholesterol	2	139.5577643	69.7788822	0.3611860	0.6968530	No
	Triglycerides	2	983.4253845	491.7126923	1.0403355	0.3533514	No
Smoking_Status*Physical_Activity_Level	BMI	2	9.1746303	4.5873152	0.1819674	0.8336296	No
	Systolic_BP	2	374.7227931	187.3613965	1.7837238	0.1680327	No
	Diastolic_BP	2	35.1820095	17.5910048	0.2511737	0.7778892	No
	Total_Cholesterol	2	548.1813321	274.0906661	1.0565275	0.3476765	No
	HDL_Cholesterol	2	34.7360369	17.3680185	0.6863937	0.5033977	No
	LDL_Cholesterol	2	81.0732653	40.5366326	0.2098237	0.8107286	No
	Triglycerides	2	1385.5157324	692.7578662	1.4656945	0.2309374	No

Error	BMI	24988	629936.0804927	25.2095438	
	Systolic_BP	24988	2624726.2166485	105.0394676	
	Diastolic_BP	24988	1750040.3419773	70.0352306	
	Total_Cholesterol	24988	6482536.1502560	259.4259705	
	HDL_Cholesterol	24988	632278.6278431	25.3032907	
	LDL_Cholesterol	24988	4827525.8899728	193.1937686	
	Triglycerides	24988	11810532.9189817	472.6481879	
Corrected Total	BMI	24999	630055.3811986		
	Systolic_BP	24999	2735549.5368851		
	Diastolic_BP	24999	1750837.1460550		
	Total_Cholesterol	24999	6959647.8134946		
	HDL_Cholesterol	24999	632545.2390143		
	LDL_Cholesterol	24999	5002459.9482062		
	Triglycerides	24999	11875366.6962249		

Step 5 : Residual Analysis & Assumption's checks

Choose

Statistics → Analysis of (Co)Variance → MANCOVA

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

Homogeneity of Variance (Homoscedasticity)

Statistical Tests

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

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:

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)
 ☐ Box M's Test

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):

9

