Multiple regression

Introduction to multiple regression

Often more than one variable is required to explain variation in a dependent variable.

- Soil infiltration rates depend on rainfall intensity and soil moisture content
- Carbon emissions depend on the number of automobiles and the number of trees in a location

Introduction to multiple regression

Often more than one variable is required to explain variation in a dependent variable.

- Soil infiltration rates depend on rainfall intensity and soil moisture content
- Carbon emissions depend on the number of automobiles and the number of trees in a location

Multiple regression can be used to differentiate between the effects of variables and to ascertain which factors are most important.

Introduction to multiple regression

Multiple regression models the change in the dependent variable y when more than one independent variable has a significant influence.

$$y = b_0 + b_1 x.$$

Introduction to multiple regression

Multiple regression models the change in the dependent variable y when more than one independent variable has a significant influence.

$$y = b_0 + b_1 x.$$

In the simple linear regression above,

- Predict y from the intercept (b_0), one independent variable (x) and a slope (b_1).
- Change in x corresponds to a change of b_1 units in y.

Introduction to multiple regression

Multiple regression predicts y from the following:

- Intercept (b_0)
- Multiple independent variables (x_1, x_2, etc.) and coefficients (b_1, b_2, etc.)

Multiple regression coefficients describe the effects of each of these independent variables on the depend variable, while holding the rest of the independent variables constant.

Introduction to multiple regression

Equation for multiple regression with two independent variables,

$$y = b_0 + b_1 x_1 + b_2 x_2.$$

Introduction to multiple regression

Equation for multiple regression with two independent variables,

$$y = b_0 + b_1 x_1 + b_2 x_2.$$

Extended to any number of k independent variables,

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_k x_k.$$

Adjusted coefficient of determination

The proportion of variation in y explained by a model can only increase with independent variables.

- New independent variables cannot explain *negative* variation

Adjusted coefficient of determination

The proportion of variation in y explained by a model can only increase with independent variables.

- New independent variables cannot explain *negative* variation
- Random variables will explain some variation by chance

Adjusted coefficient of determination

The proportion of variation in y explained by a model can only increase with independent variables.

- New independent variables cannot explain *negative* variation
- Random variables will explain some variation by chance
- Need to account for spurious improvement of fit

Adjusted coefficient of determination

The proportion of variation in y explained by a model can only increase with independent variables.

- New independent variables cannot explain *negative* variation
- Random variables will explain some variation by chance
- Need to account for spurious improvement of fit
- Want to retain only relevant predictors in a model

Adjusted coefficient of determination

The proportion of variation in y explained by a model can only increase with independent variables.

- New independent variables cannot explain *negative* variation
- Random variables will explain some variation by chance
- Need to account for spurious improvement of fit
- Want to retain only relevant predictors in a model

To help account for spurious improvement of fit, we can use an adjusted R squared.

Adjusted coefficient of determination

An adjusted R squared value,

$$R^2_{adj} = 1 - \left(1 - R^2\right)\left(\frac{n-1}{n-k-1}\right).$$

Adjusted coefficient of determination

An adjusted R squared value,

$$R^2_{adj} = 1 - \left(1 - R^2\right)\left(\frac{n-1}{n-k-1}\right).$$

- R^2 is the coefficient of determination
- k is the number of independent variables in the model
- n is the number of samples

Multiple regression in Jamovi

In 1993, monthly amounts of nitrate (as tonnes of nitrogen) were measured in the River Thames, together with mean daily evaporation and rainfall (in mm) for 12 months of the year.

We wish to examine if evaporation and rainfall can be used to predict nitrate amounts in the river.

[1] **Image:** Public Domain

Multiple regression in Jamov

	Nitrogen_(tonnes)	Evaporation_(mm)	Rainfall_(mm)
Jan	22	0.3	2.1
Feb	26	0.4	1.5
Mar	27	0.9	2.0
Apr	18	1.9	1.5
May	12	2.7	1.7
Jun	9	3.3	1.9
Jul	7	3.4	1.5
Aug	6	3.4	1.4
Sep	7	2.3	1.7
Oct	8	1.2	1.9
Nov	12	0.5	2.4
Dec	21	0.3	2.2

Fitting the full model

Include both evaporation and rainfall as independent variables.

$$Nitrogen = b_0 + b_1 Evaporation + b_2 Rainfall.$$

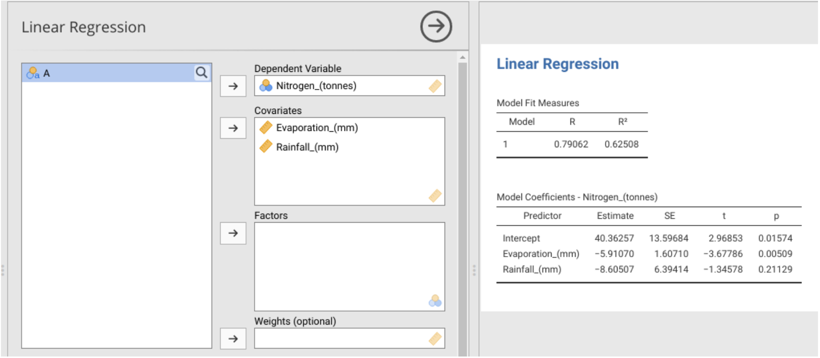
Fitting the full model

Include both evaporation and rainfall as independent variables.

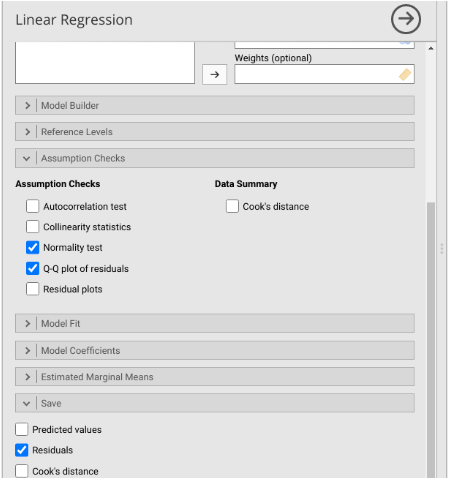
$$Nitrogen = b_0 + b_1 Evaporation + b_2 Rainfall.$$

- Select 'Regress', then 'Linear regression' from pulldown menu.

Fitting the full model



Check model assumptions

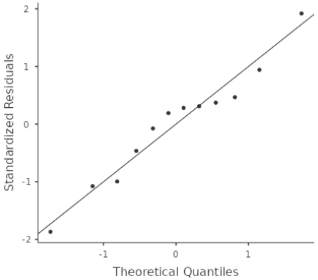


Assumption Checks

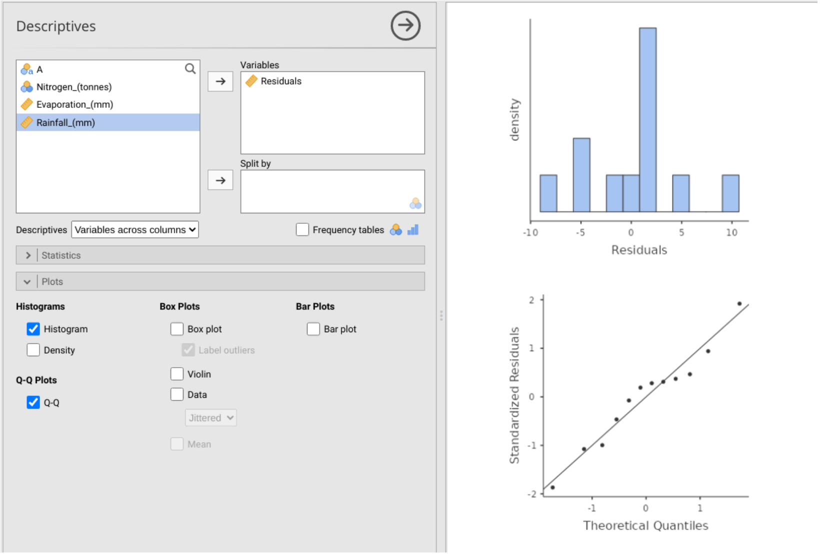
Normality Test (Shapiro-Wilk)

Statistic	p
0.96538	0.85691

Q-Q Plot



Check model assumptions



Model output

Linear Regression

Model Fit Measures

			Overall Model Test			
Model	R	Adjusted R^2	F	df1	df2	p
1	0.79062	0.54177	7.50271	2	9	0.01210

Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
Evaporation_(mm)	376.92330	1	376.92330	13.52665	0.00509
Rainfall_(mm)	50.46705	1	50.46705	1.81111	0.21129
Residuals	250.78715	9	27.86524		

Note. Type 3 sum of squares

[3]

Model Coefficients - Nitrogen_(tonnes)

Predictor	Estimate	SE	t	p
Intercept	40.36257	13.59684	2.96853	0.01574
Evaporation_(mm)	−5.91070	1.60710	−3.67786	0.00509
Rainfall_(mm)	−8.60507	6.39414	−1.34578	0.21129

Model output: Overall and adjusted R-squared

Model Fit Measures

			Overall Model Test			
Model	R	Adjusted R^2	F	df1	df2	p
1	0.79062	0.54177	7.50271	2	9	0.01210

Model output: Omnibus

Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
Evaporation_(mm)	376.92330	1	376.92330	13.52665	0.00509
Rainfall_(mm)	50.46705	1	50.46705	1.81111	0.21129
Residuals	250.78715	9	27.86524		

Note. Type 3 sum of squares

Model output: coefficients

Model Coefficients - Nitrogen_(tonnes)

Predictor	Estimate	SE	t	p
Intercept	40.36257	13.59684	2.96853	0.01574
Evaporation_(mm)	−5.91070	1.60710	−3.67786	0.00509
Rainfall_(mm)	−8.60507	6.39414	−1.34578	0.21129

Conclusions from the multiple regression model

According to the multiple regression

- Overall model is significant
- High R-squared value
- Evaporation has a significant effect
- Rainfall does not have a significant effect

There is a relationship between evaporation and nitrogen, but not rainfall and nitrogen.