UNIVERSITY
OF MIAMI

Multiple Regression Using SPSS

Presented by Nasser Hasan - Statistical Supporting Unit

6/3/2020

nasser.hasan@miami.edu

UNIVERSITY
OF MIAMI

Overview

- Brief introduction of Multiple Linear Regression.
  - Model specification
  - Assumptions
- Multiple Linear Regression Analysis Using SPSS.
- Variable Selection.

Overview

Simple Linear Regression

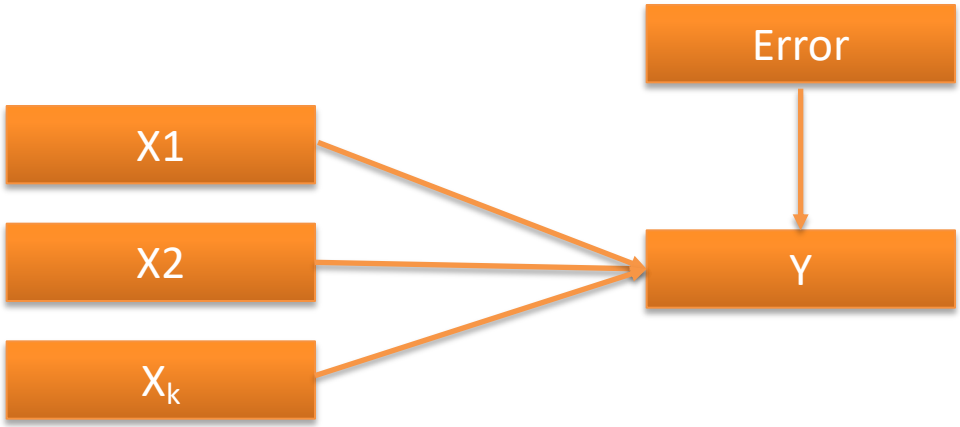
A simple linear regression is carried out to estimate the relationship between a dependent variable, Y, and a single predictor variable, X.

Overview

Multiple Linear Regression

A multiple linear regression analysis is carried out to predict the values of a dependent variable, Y, given a set of k^{th} predictor variables (X1, X2, …, X_k).

We also use it when we want to determine which variables are better predictors than others. (Variables Selection)



Overview

Assumptions

1) Normality: residuals/error terms should be normally distributed with a mean of 0 and constant variance.

$$e_i \sim N(0, \sigma^2)$$

2) Homoscedasticity: the variance of residual terms should be the same at each level of the predictor variable(s).

Overview

Assumptions

3) Independence: all of the observations are independent.

4) Independent errors: for any two observations, residual terms are not related, $r_{e_i,e_j} = 0$. Violation of this assumption will cause **autocorrelation** problem.

Overview

Assumptions

5) Linearity: relationship between DV and IVs should be linear.

6) No perfect collinearity.

Performing the Analysis With SPSS

Dataset:

- This data contains four variables from 20 students that are related to students' performance on exam.
  - n=20
  - 4 variables:
    - Exam score (ratio)
    - Hours spent revising (ratio)
    - Anxiety score (ratio)
    - A-level entry points (ratio)

Please download the dataset using this link:
http://staff.bath.ac.uk/pssiw/stats2/examrevision.sav

All thanks to Ian Walker – University of Bath

Performing the Analysis With SPSS

Example 1:

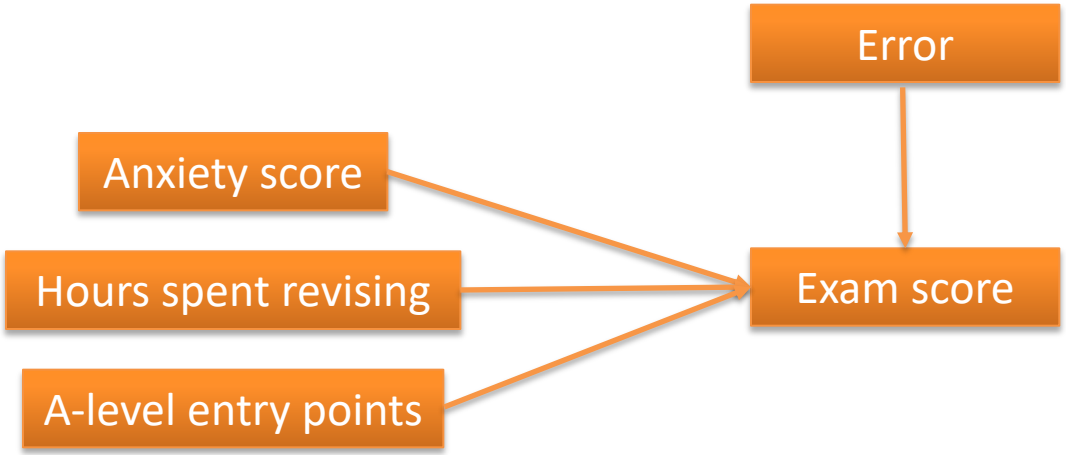
- We want to determine whether hours spent revising, anxiety scores, and A-level entry points have effect on exam scores for participants.

Dependent variable: exam score

Predictors: hours spent revising, anxiety scores, and A-level entry points.

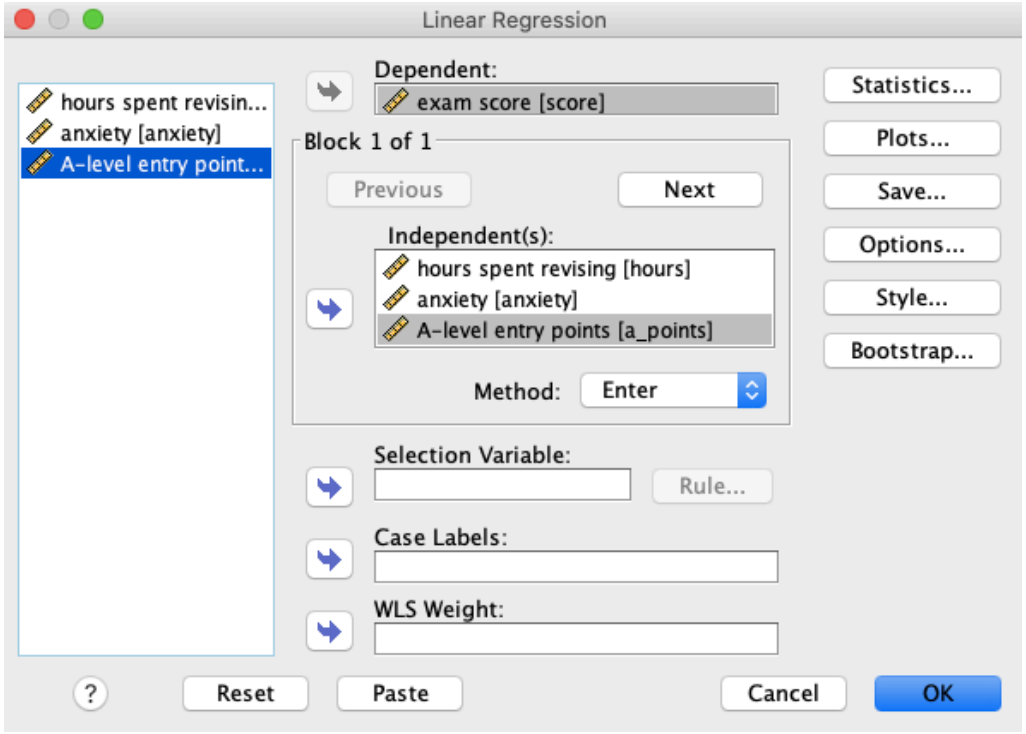
Performing the Analysis With SPSS

Regression Model



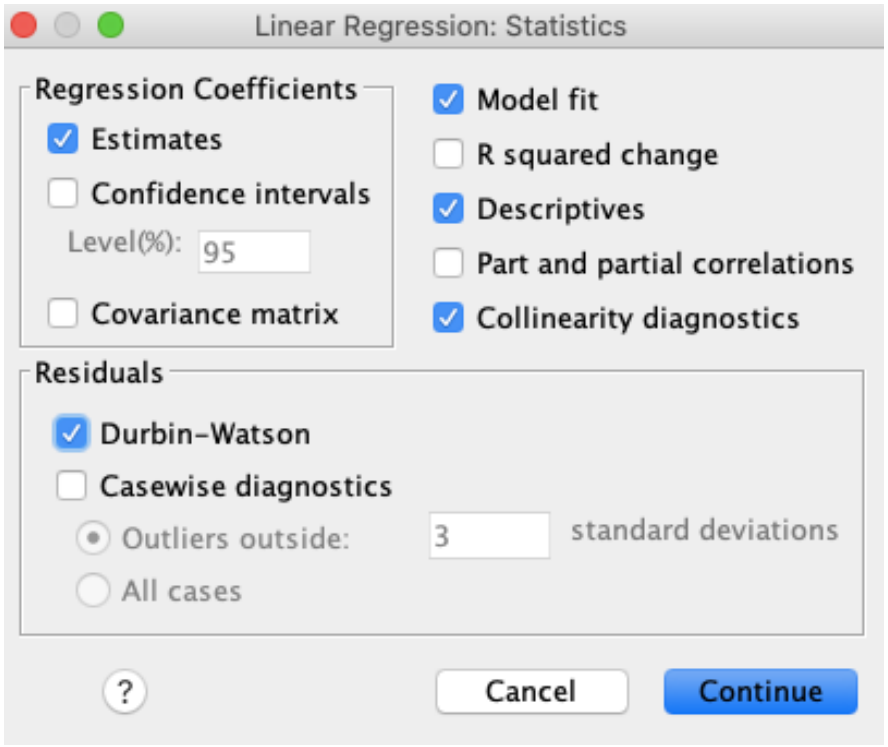
Performing the Analysis With SPSS

Analyze > Regression > Linear > Move score to Dependent, all other variables to Independent(s)



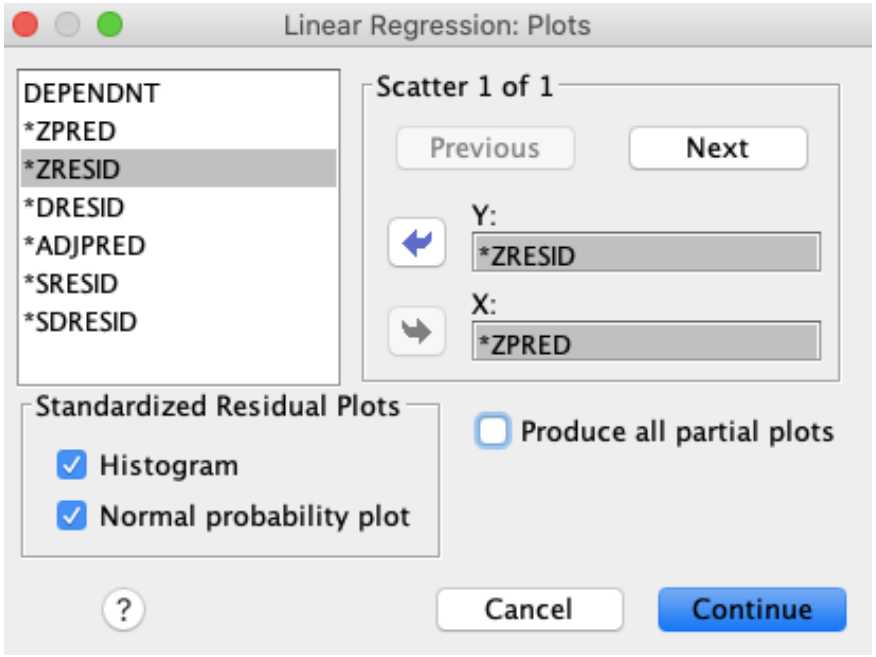
Performing the Analysis With SPSS

Click on Statistics > Check on the following options > Continue



Performing the Analysis With SPSS

Click on Plots > Check Histogram and Normal probability plot under Standardized Residual Plots > move *ZRESID to Y and move *ZPRED to X > Continue



SPSS Output

Descriptive Statistics

	Mean	Std. Deviation	N
exam score	61.0000	10.96406	20
hours spent revising	39.1500	9.07440	20
anxiety	49.3000	18.91004	20
A-level entry points	23.2000	3.20526	20

SPSS Output

Correlations

		exam score	hours spent revising	anxiety	A-level entry points
Pearson Correlation	exam score	1.000	.821	-.118	.872
	hours spent revising	.821	1.000	-.340	.732
	anxiety	-.118	-.340	1.000	-.244
	A-level entry points	.872	.732	-.244	1.000
Sig. (1-tailed)	exam score	.	.000	.310	.000
	hours spent revising	.000	.	.072	.000
	anxiety	.310	.072	.	.150
	A-level entry points	.000	.000	.150	.
N	exam score	20	20	20	20
	hours spent revising	20	20	20	20
	anxiety	20	20	20	20
	A-level entry points	20	20	20	20

SPSS Output

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	A-level entry points, anxiety, hours spent revising[b]	.	Enter

a. Dependent Variable: exam score

b. All requested variables entered.

SPSS Output – Model Summery

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.927[a]	.860	.834	4.46756	3.078

a. Predictors: (Constant), A-level entry points, anxiety, hours spent revising

b. Dependent Variable: exam score

R: multiple correlation coefficient= .927.
R^2: coefficient of determination= .860.

The model explains 86.0% of the variation in the dependent variable.

Durbin-Watson (to assess autocorrelation) – Residuals are negatively correlated

SPSS Output – ANOVA table

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1964.654	3	654.885	32.811	.000[b]
	Residual	319.346	16	19.959		
	Total	2284.000	19			

a. Dependent Variable: exam score

b. Predictors: (Constant), A-level entry points, anxiety, hours spent revising

The overall model is significantly useful in explaining exam score, F (3, 16) = 32.81, p < .05.

SPSS Output – Collinearity Assumption

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-11.823	8.806		-1.343	.198		
	hours spent revising	.551	.171	.456	3.226	.005	.437	2.288
	anxiety	.104	.058	.179	1.796	.091	.885	1.130
	A-level entry points	1.989	.469	.581	4.239	.001	.464	2.153

a. Dependent Variable: exam score

A value larger than 10 indicates collinearity between predictors.

SPSS Output – The Significance of the Effect

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-11.823	8.806		-1.343	.198		
	hours spent revising	.551	.171	.456	3.226	.005	.437	2.288
	anxiety	.104	.058	.179	1.796	.091	.885	1.130
	A-level entry points	1.989	.469	.581	4.239	.001	.464	2.153

a. Dependent Variable: exam score

- Hours has significant effect on exam score, $t(16)=3.23$, $p < .05$.
- Anxiety does not have a significant effect on exam score, $t(16)=1.80$, $p = .09$.
- A-level has significant effect on exam score, $t(16)=4.24$, $p < .05$.

SPSS Output – Interpreting Coefficiants

Coefficients[a]

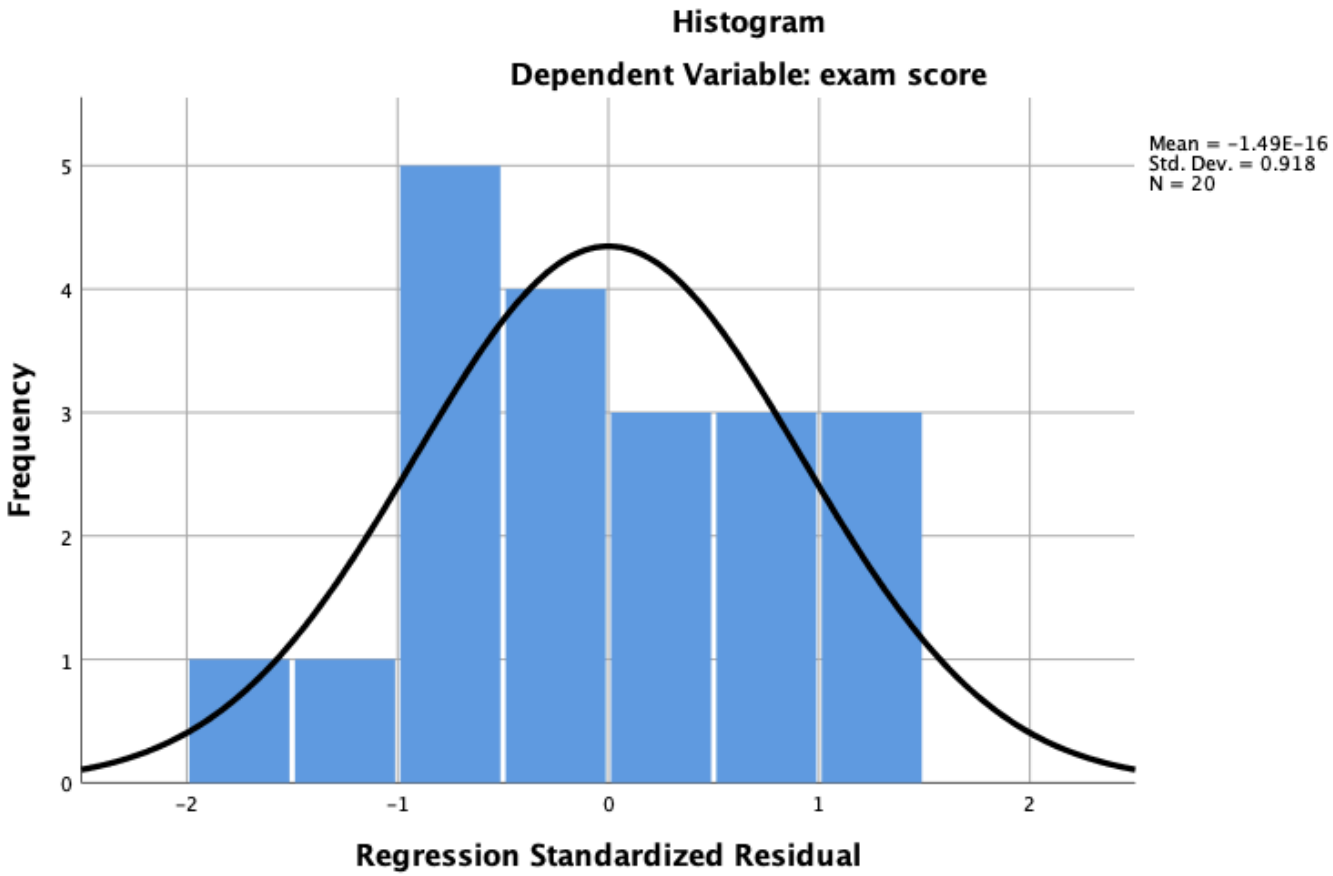
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-11.823	8.806		-1.343	.198		
	hours spent revising	.551	.171	.456	3.226	.005	.437	2.288
	anxiety	.104	.058	.179	1.796	.091	.885	1.130
	A-level entry points	1.989	.469	.581	4.239	.001	.464	2.153

a. Dependent Variable: exam score

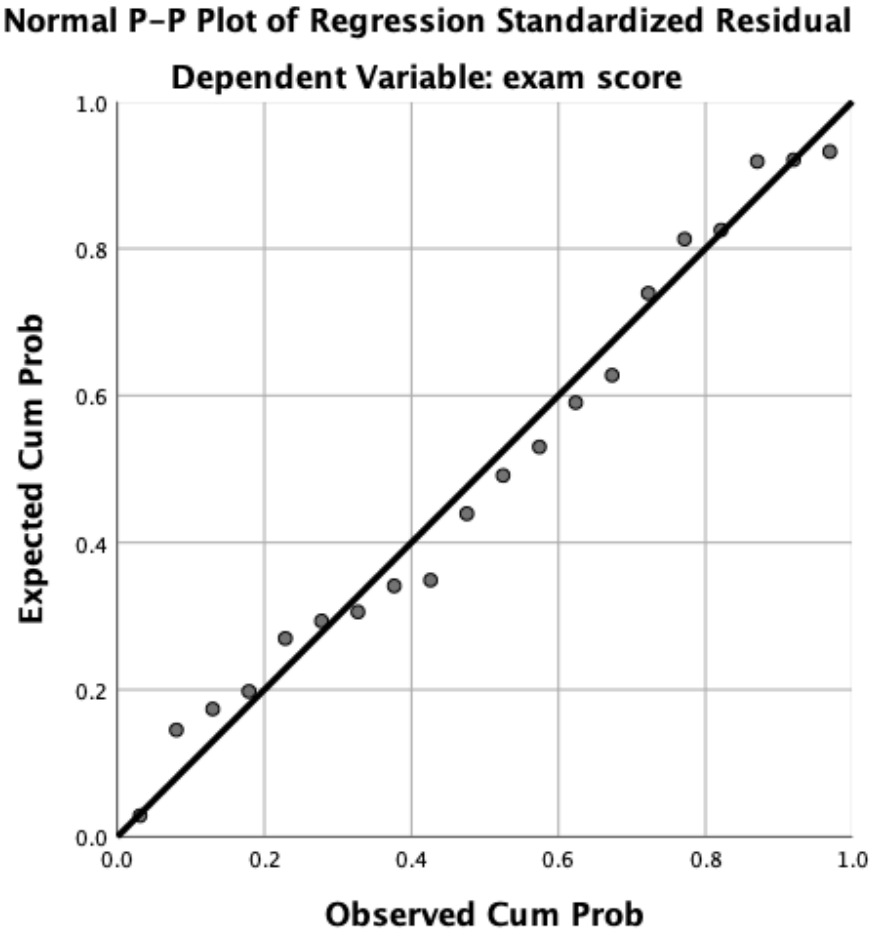
- With one-unit increase in hours, the exam score increases by .55.
- With one-unit increase in anxiety, the exam score increases by .10.
- With one-unit increase in A-level, the exam score increases by .1.99.

Exam score = -11.82 + (.55*hours) + (.10*anxiety) + (1.99*A-level)

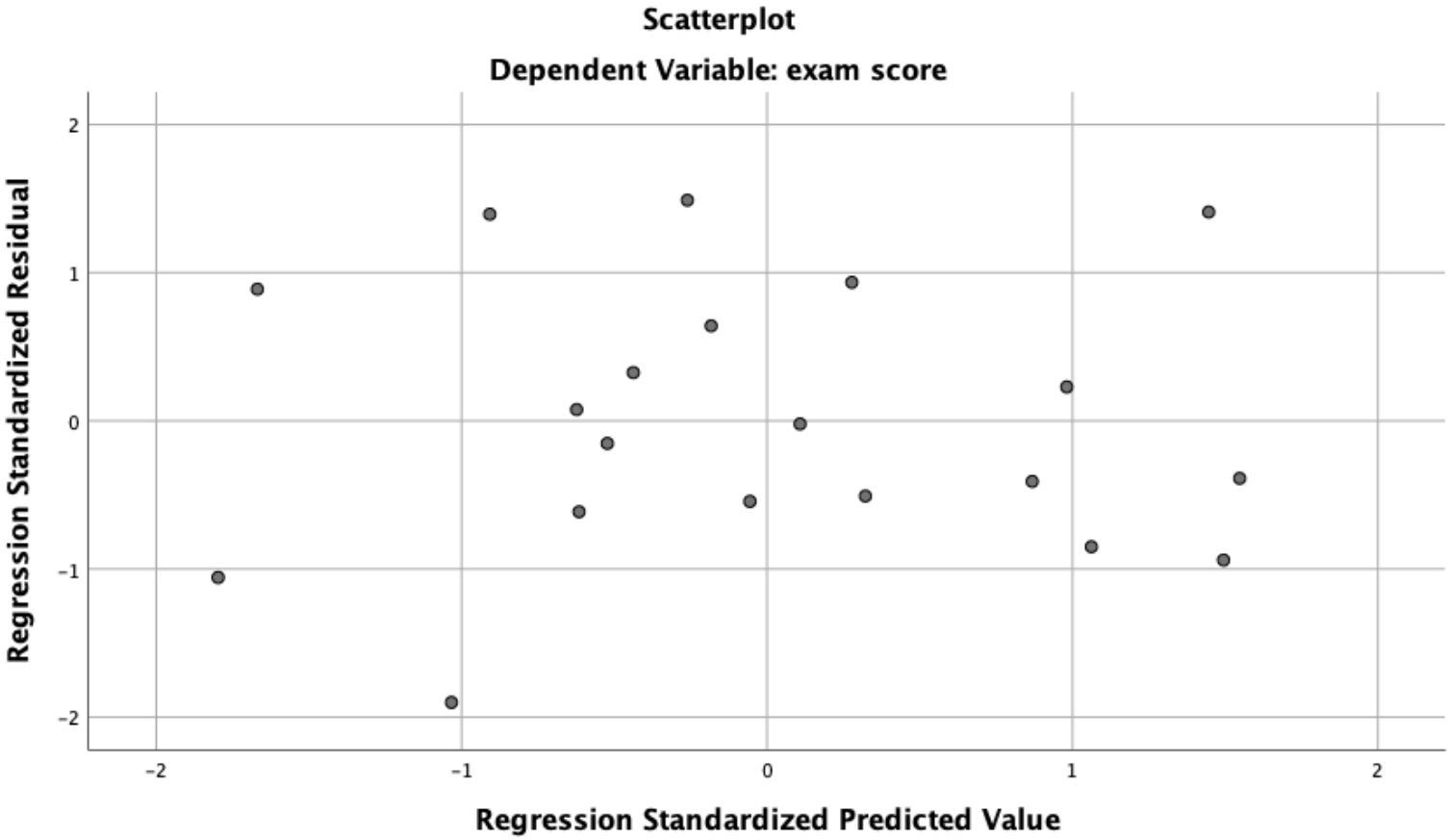
SPSS Output - Residuals Normality Assumption



SPSS Output - Residuals Normality Assumption



SPSS Output – Homoscedasticity Assumption



APA Format Write-up

A multiple linear regression was fitted to explain exam score based on hours spent revising, anxiety score, and A-Level entry points. All of the assumptions were met except the autocorrelation assumption between residuals. The overall model explains 86.0% variation of exam score, and it is significantly useful in explaining exam score, F (3, 16) = 32.81, p < .05.

With one-unit increase in hours, the exam score increases by .55, which was found to be a significant change, t(16)=3.23, p < .05. With one-unit increase in anxiety, the exam score increases by .10, which was not found to be a significant change, t(16)=1.80, p = .09. With one-unit increase in A-level, the exam score increases by .1.99, which was found to be a significant change, t(16)=4.24, p < .05.

Performing the Analysis With SPSS

Example:

- We want to determine whether hours spent revising, anxiety scores, and A-level entry points have effect on exam scores for participants.

What if we are interested in knowing the best predictors to include in our model?

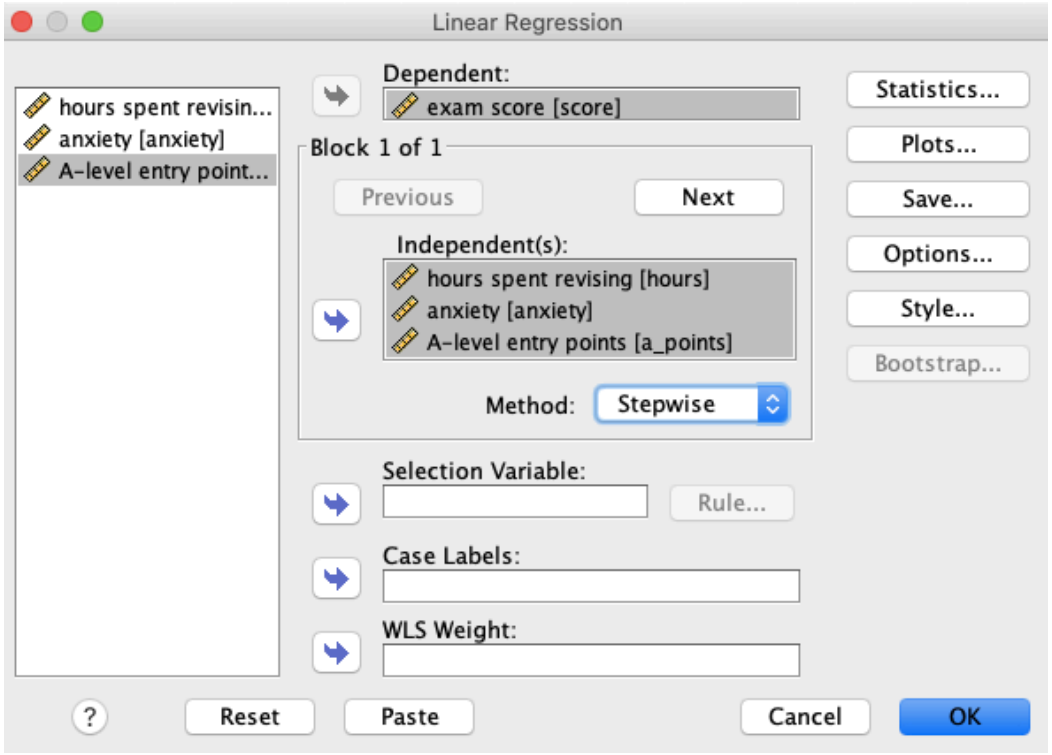
Performing the Analysis With SPSS

Variable Selection Method

- Forward.
- Backward.
- Stepwise.

Performing the Analysis With SPSS

Analyze > Regression > Linear > Move score to Dependent, all other variables to Independent(s) > Click on Method > Change Enter to Stepwise



SPSS Output

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	A-level entry points	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
2	hours spent revising	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: exam score

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.872[a]	.760	.746	5.52134
2	.912[b]	.832	.812	4.75122

a. Predictors: (Constant), A-level entry points

b. Predictors: (Constant), A-level entry points, hours spent revising

Performing the Analysis With SPSS

Exercise:

- We want to determine whether GRE verbal and quantitative have effect on GPA.

Dataset:
Please download it from here:
http://core.ecu.edu/psyc/wuenschk/SPSS/MultReg.sav

© Retrieved from this website:
http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Data.htm

Thanks for Listening and Attending!

Any Questions?

Can you please give us a minute to fill this survey as it will help us to evaluate our performance and take your feedback into consideration for future webinars:
https://umiami.qualtrics.com/jfe/form/SV_a9N5Xta6OlybEeV