



LCDB7011 Quantitative Research Methods

Week 10&11 – Simple Linear
Regression

Dr. Helen LEE



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

By the end of this session, you should be able to
understand the essentials of

1. Simple regression model

2. Analysis and Interpretation



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

What is Simple Regression ?

- It is a way of **predicting** the value of one variable from another
- Therefore, it is a hypothetical model of the relationship between two variables relationship between two variables
- The model is a linear one
- Therefore, we describe the relationship using the equation of a straight line



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Regression

The purpose of a regression analysis is to find out if one or more independent variables influence one dependant variable

The basis of a regression analysis is to plot data on a chart and to draw a line of best fit through the middle of this data points.

This allows the researcher to estimate the expected values of the dependent variable when the independent variables take on a given set of values.

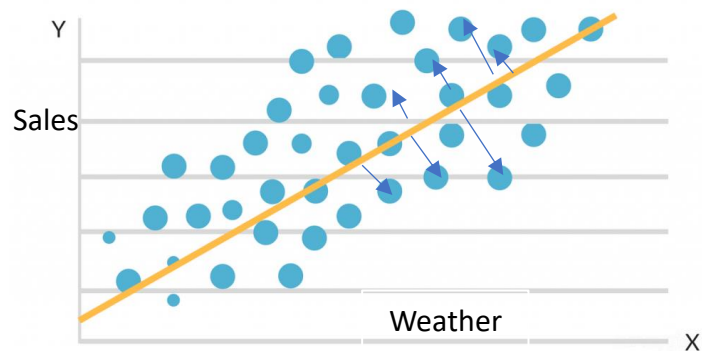
“Best fit” here means that the **error term, the distance from each point to the line**, is minimized. Since the relationship between variables is probably not completely linear and because there are other factors outside the scope of the study the line does not go through all the data points. The distance between each point and the linear graph (shown as arrows on the graph) is the error term. Errors are also called residuals.

Regression analysis tests a model, such as:

H_0 = Weather does not influence Sales

H_1 = Sales are influenced by Sales

or H_1 = Sales increase with poor weather



The dependant variable is always plotted on the Y axis

Weather

Sales

Assumptions of regression

- There must be a linear relationship between the outcome variable and the independent variables. Scatterplots can show whether there is a linear or curvilinear relationship. Use the chart builder
- Multivariate Normality—Multiple regression assumes that the residuals are normally distributed. (Check in Descriptives, Explore or look at QQ plots or run 1 Sample Kolmogorov-Smirnow test, must be not significant)
- No Multicollinearity—Multiple regression assumes that the independent variables are not highly correlated with each other. This assumption is tested using Variance Inflation Factor (VIF) values. VIF values must be under 10.
- Homoscedasticity—This assumption states that the variance of error terms are similar across the values of the independent variables. A plot of standardized residuals versus predicted values can show whether points are equally distributed across all values of the independent variables. (Linear Regression → Plots, enter the standardized residual plot (ZPRED on x-axis and ZRESID on y-axis), which allows us to see homoscedasticity and normality of residuals.

What is simple regression ?

Introduction to Regression Analysis (3:50)

https://www.google.com/search?q=what+is+simple+regression+analysis&sca_esv=9967b14e20c7abd4&sca_upv=1&rlz=1C1GCHA_enGB1126GB1126&tbm=vid&sxsrf=ADLYWILtaLgwBKbows7jrTrw4psnNSBhhw:1726645193405&ei=yYPqZpGzGN7OhbIPqf7zyQs&start=20&sa=N&ved=2ahUKEwjRu5_L_sulAxVeZ0EAHSn_PLk4ChDy0wN6BAgUEAc&biw=1642&bih=925&dpr=1#fpstate=ive&vld=cid:c19bbbe2,vid:vPde9bYrr80,st:0

Simple Linear Regression



Simple Regression

Graph => Chart Builder

Chi_square_58cases_v1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Chart Builder...

Graphboard Template Chooser...

Relationship Map...

Weibull Plot...

Compare Subgroups

Regression Variable Plots

Bar...

3-D Bar...

Line...

Area...

Pie...

High-Low...

Boxplot...

Error Bar...

Population Pyramid...

Scatter/Dot...

Histogram...

	ID	Gender	Age	Role
4	7	1	31.00	
5	8	1	38.00	
6	10	1	20.00	
7	11	1	58.00	
8	12	1	18.00	
9	13	1	44.00	
10	14	1	22.00	
11	16	1	60.00	
12	17	1	53.00	
13	18	1	48.00	
14	20	1	28.00	
15	21	1	35.00	
16	22	1	22.00	
17	23	1	40.00	
18	23	1	40.00	
19	24	1	62.00	
20	24	1	62.00	
21	26	1	31.00	
22	26	1	31.00	

Age_cat

3.00

3.00

2.00

5.00

1.00

4.00

2.00

5.00

4.00

4.00

3.00

3.00

2.00

3.00

3.00

5.00

5.00

3.00

3.00

Simple Regression

Select Plot and Drag the plot diagram onto the chart

Chi_square_58cases_v1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Chart Builder

Variables: Chart preview uses example data

Drag a Gallery chart here to use it as your starting point

OR

Click on the Basic Elements tab to build a chart element by element

Filter by:

Category 1
Category 2

Gallery Basic Elements Groups/Point ID Titles/Footnotes

Choose from:

Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes

Scatter Plot

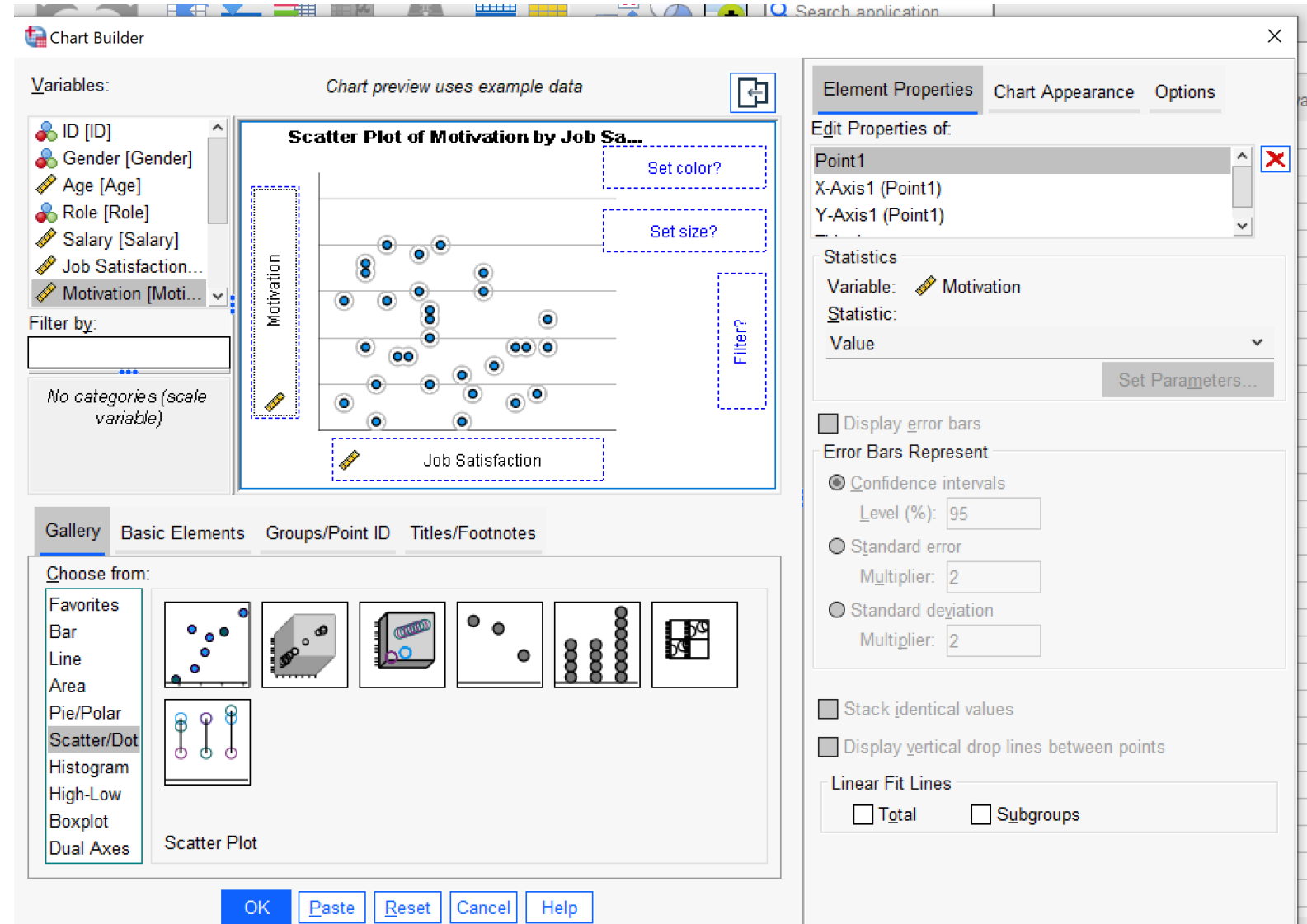
Summary Point Plot

OK Paste Reset Cancel Help

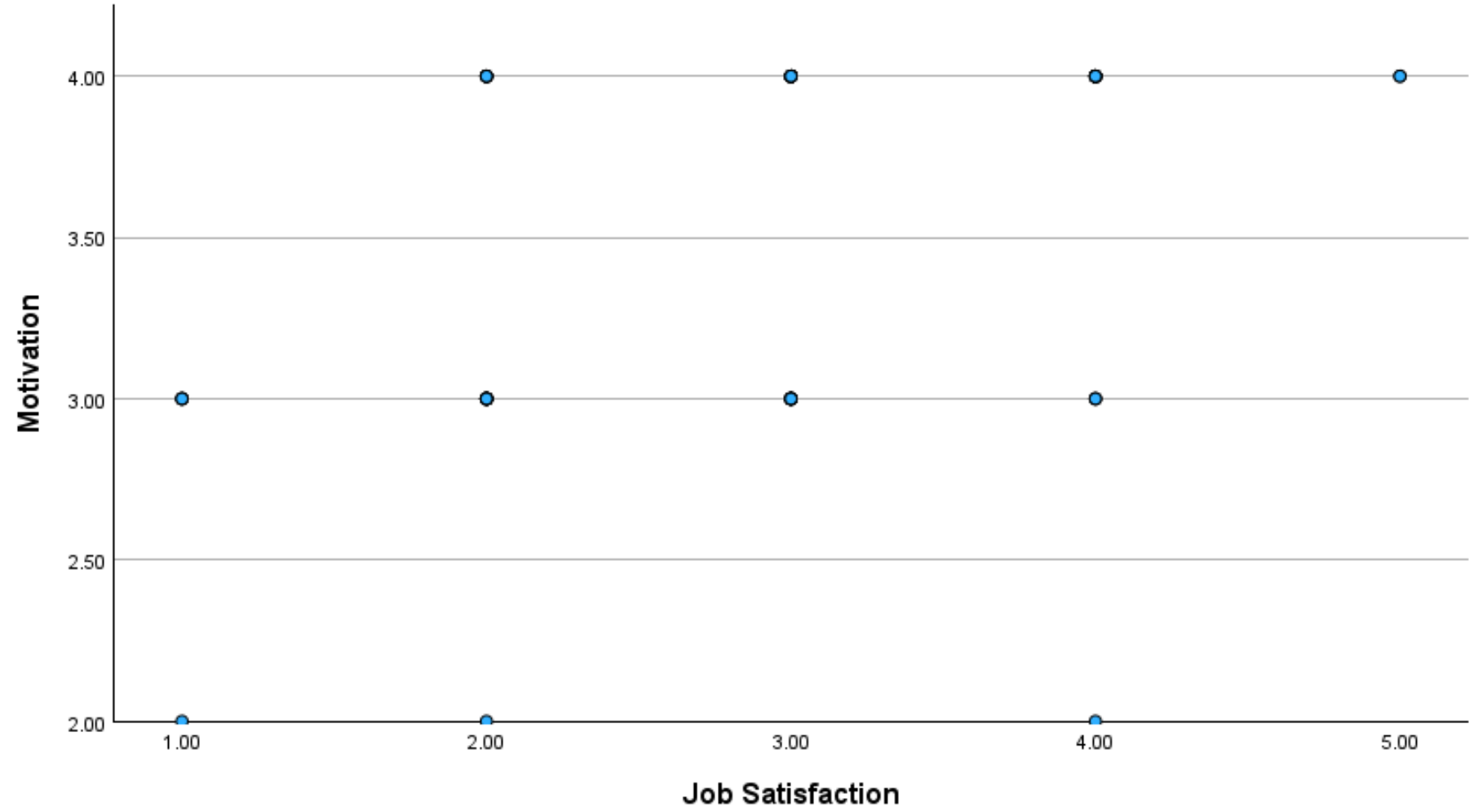
	ID	Gender
4	7	1
5	8	1
6	10	1
7	11	1
8	12	1
9	13	1
10	14	1
11	16	1
12	17	1
13	18	1
14	20	1
15	21	1
16	22	1
17	23	1
18	23	1
19	24	1
20	24	1
21	26	1
22	26	1
23	27	1
24	27	1
25	28	1
26	28	1
27	28	1
28	31	1
29	31	1
30	31	1
31	32	1
32	32	1

Simple Regression

X axis = Job satisfaction (Predictor)
Y axis = Motivation (Dependent variable)



Scatter Plot of Motivation by Job Satisfaction



Simple Regression

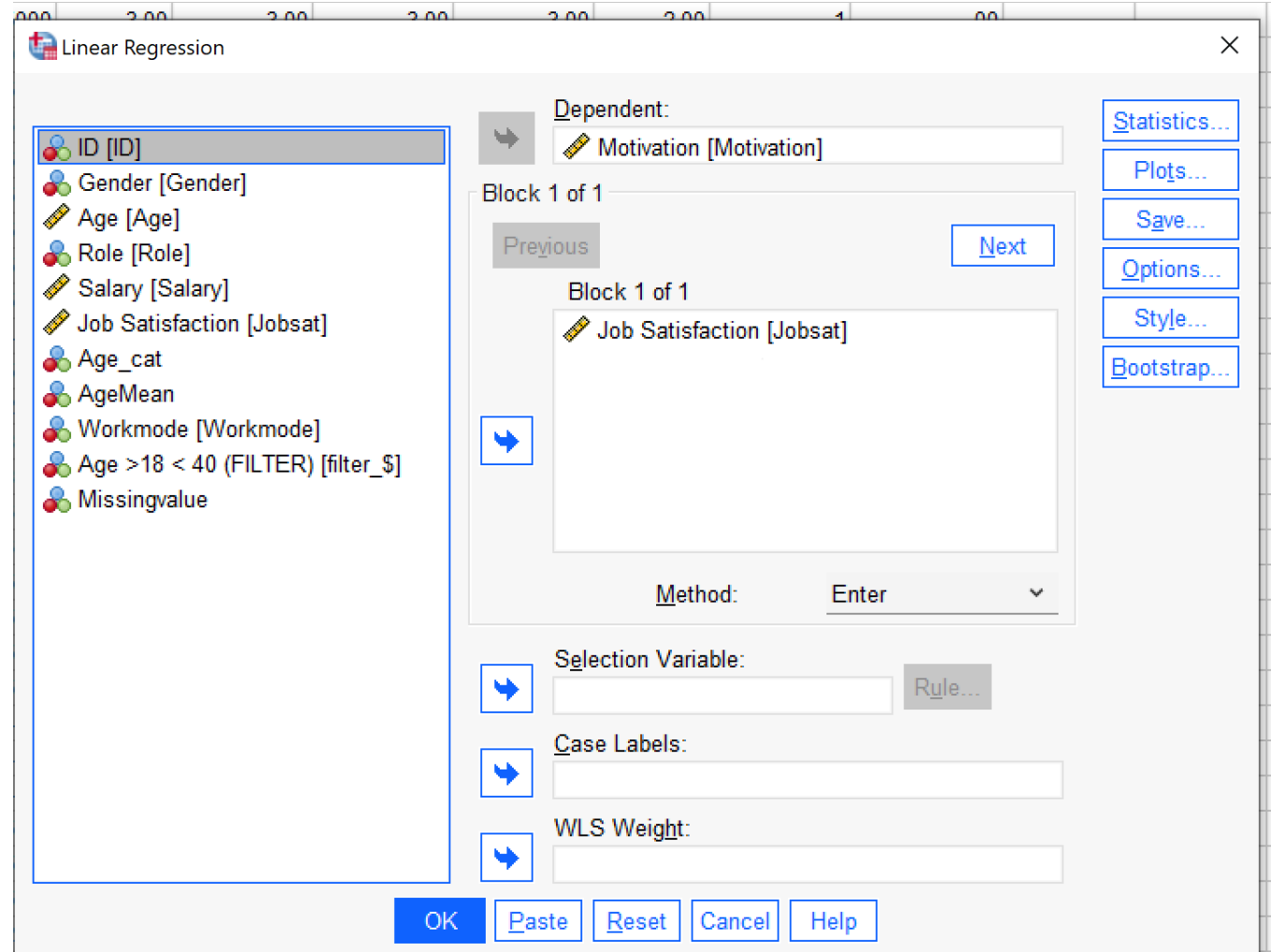
Go to Analyze
=>Regression =>
Linear

The screenshot shows the SPSS software interface. The 'Analyze' menu is open, and the 'Regression' option is selected. The 'Linear' option is highlighted under the 'Regression' submenu. The background shows a data grid with columns labeled EP2, EP3, AC1, EP4, JS2, JS3, and AC. The data values are as follows:

	EP2	EP3	AC1	EP4	JS2	JS3	AC
0	8	2	7	3	2	4	
0	9	4	7	7	5	4	
0	9	4	7	4	4	5	
9	8	1	5	4	3	2	
3	8	1	5	4	3	1	
9	9	2	6	4	4	5	
0	10	4	7	4	5	2	
0	10	3	6	4	2	3	
0	10	2	7	6	3	3	
0	0	4	7	2	4	2	
4	4	3	2	4	2	1	
4	4	5	5	4	1	2	
4	4	3	1	5	4	3	
6	6	3	1	5	4	2	
5	4	3	3	6	3	2	

Simple Regression

X axis = Job satisfaction (Predictor)
Y axis = Motivation (Dependent variable)



The image shows the 'Linear Regression' dialog box in SPSS. On the left, a list of variables includes 'ID [ID]', 'Gender [Gender]', 'Age [Age]', 'Role [Role]', 'Salary [Salary]', 'Job Satisfaction [Jobsat]', 'Age_cat', 'AgeMean', 'Workmode [Workmode]', 'Age >18 < 40 (FILTER) [filter_\$]', and 'Missingvalue'. 'Motivation [Motivation]' is entered in the 'Dependent:' field. 'Job Satisfaction [Jobsat]' is entered in the 'Block 1 of 1' list. The 'Method:' dropdown is set to 'Enter'. On the right, there are buttons for 'Statistics...', 'Plots...', 'Save...', 'Options...', 'Style...', and 'Bootstrap...'. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

Linear Regression

Dependent: Motivation [Motivation]

Block 1 of 1

Job Satisfaction [Jobsat]

Method: Enter

Selection Variable: Rule...

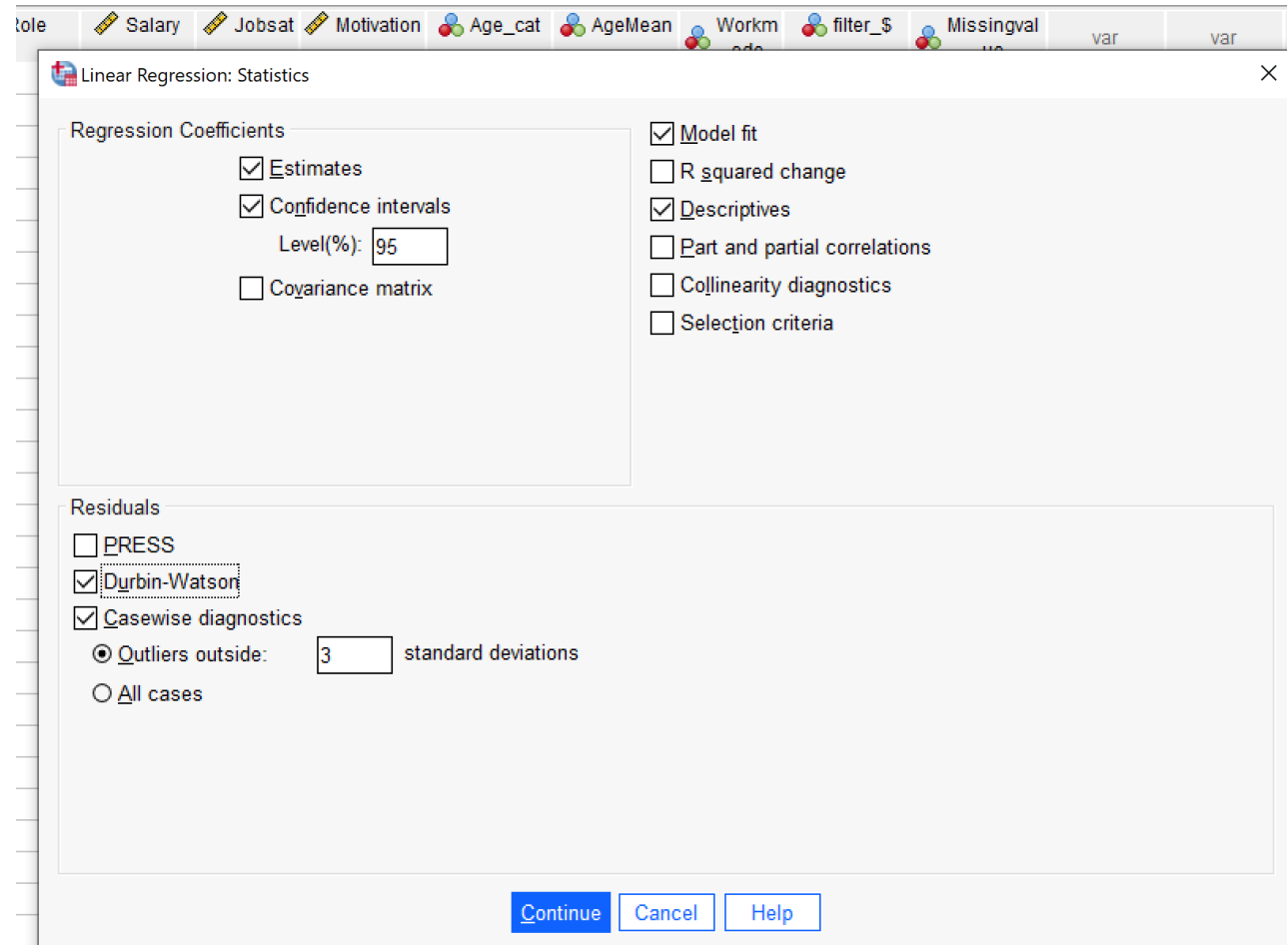
Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Simple Regression

Go to Statistics



The image shows the 'Linear Regression: Statistics' dialog box in SPSS. The 'Regression Coefficients' section has 'Estimates' and 'Confidence intervals' checked, with 'Level(%)' set to 95. 'Model fit' is also checked. The 'Residuals' section has 'Durbin-Watson' and 'Casewise diagnostics' checked. 'Outliers outside' is selected with a value of 3 standard deviations. The 'Continue' button is highlighted in blue.

Linear Regression: Statistics

Regression Coefficients

- ☒ Estimates
- ☒ Confidence intervals
Level(%): 95
- ☐ Covariance matrix
- ☒ Model fit
- ☐ R squared change
- ☒ Descriptives
- ☐ Part and partial correlations
- ☐ Collinearity diagnostics
- ☐ Selection criteria

Residuals

- ☐ PRESS
- ☒ Durbin-Watson
- ☒ Casewise diagnostics
 - ☒ Outliers outside: 3 standard deviations
 - ☐ All cases

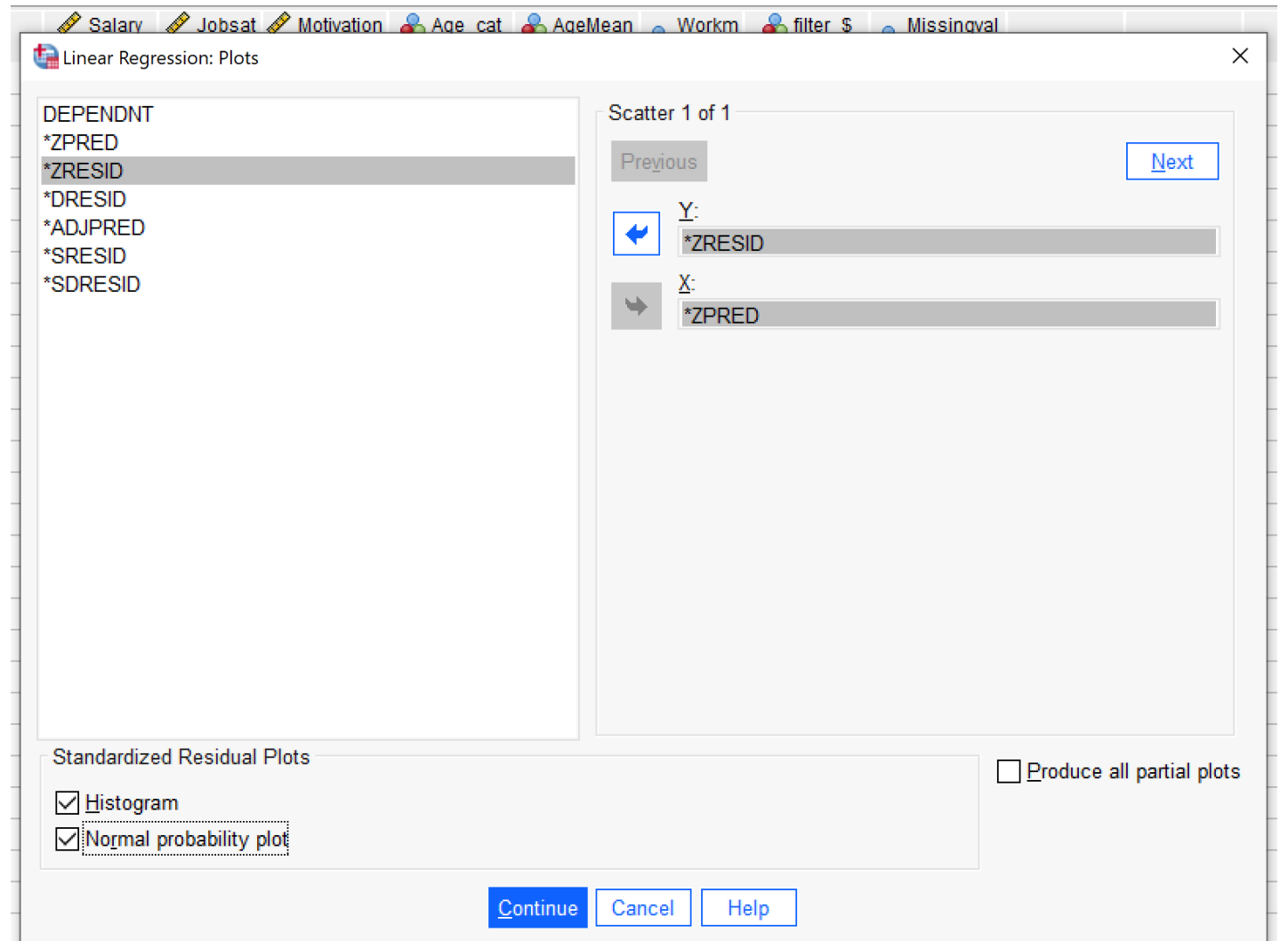
[Continue](#) [Cancel](#) [Help](#)

Simple Regression

Go to Plot

X : ZPRED

Y: ZRESID



The image shows the 'Linear Regression: Plots' dialog box in SPSS. The 'DEPENDNT' list on the left contains several variables, with '*ZRESID' selected. The 'Scatter 1 of 1' section on the right has 'Y:' set to '*ZRESID' and 'X:' set to '*ZPRED'. At the bottom, under 'Standardized Residual Plots', both 'Histogram' and 'Normal probability plot' are checked. There are 'Previous', 'Next', 'Continue', 'Cancel', and 'Help' buttons.

Salary Jobsat Motivation Age_cat AgeMean Workm filter \$ Missingval

Linear Regression: Plots

DEPENDNT

- *ZPRED
- *ZRESID
- *DRESID
- *ADJPRED
- *SRESID
- *SDRESID

Scatter 1 of 1

Previous Next

Y: *ZRESID

X: *ZPRED

Standardized Residual Plots

- ☒ Histogram
- ☒ Normal probability plot

☐ Produce all partial plots

Continue Cancel Help

Simple Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Motivation	3.4483	.59743	58
Job Satisfaction	2.8621	1.01650	58

Simple Regression

Correlations

		Motivation	Job Satisfaction
Pearson Correlation	Motivation	1.000	.421
	Job Satisfaction	.421	1.000
Sig. (1-tailed)	Motivation	.	<.001
	Job Satisfaction	.000	.
N	Motivation	58	58
	Job Satisfaction	58	58

Simple Regression

1. SPSS Output (Outlier checking)

$-3.29 < \text{Std. residual value} < +3.29$

Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.9871	3.9778	3.4483	.25174	58
Residual	-1.73009	.76522	.00000	.54180	58
Std. Predicted Value	-1.832	2.103	.000	1.000	58
Std. Residual	-3.165	1.400	.000	.991	58

a. Dependent Variable: Motivation

Simple Regression

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics				Durbin-Watson
						F Change	df1	df2	Sig. F Change	
1	.421 ^a	.178	.163	.54662	.178	12.090	1	56	<.001	1.931

a. Predictors: (Constant), Job Satisfaction

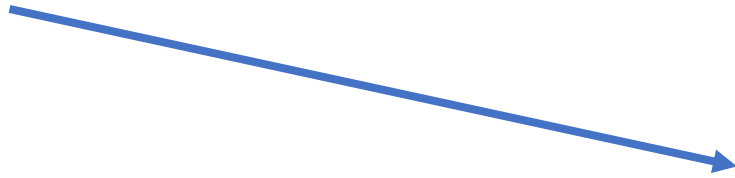
b. Dependent Variable: Motivation

2. Assumption of independent observations
 $+1 < \text{DW value} < +3$

Simple Regression

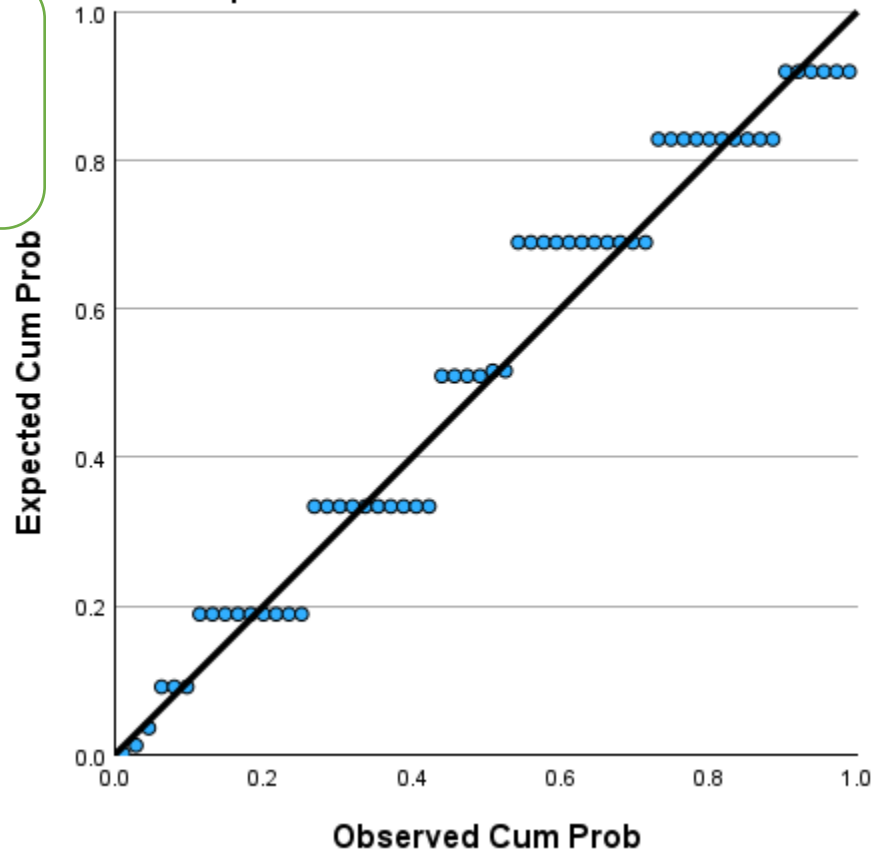
3. Assumption : Normality of Residuals

PP Plot : along 45%

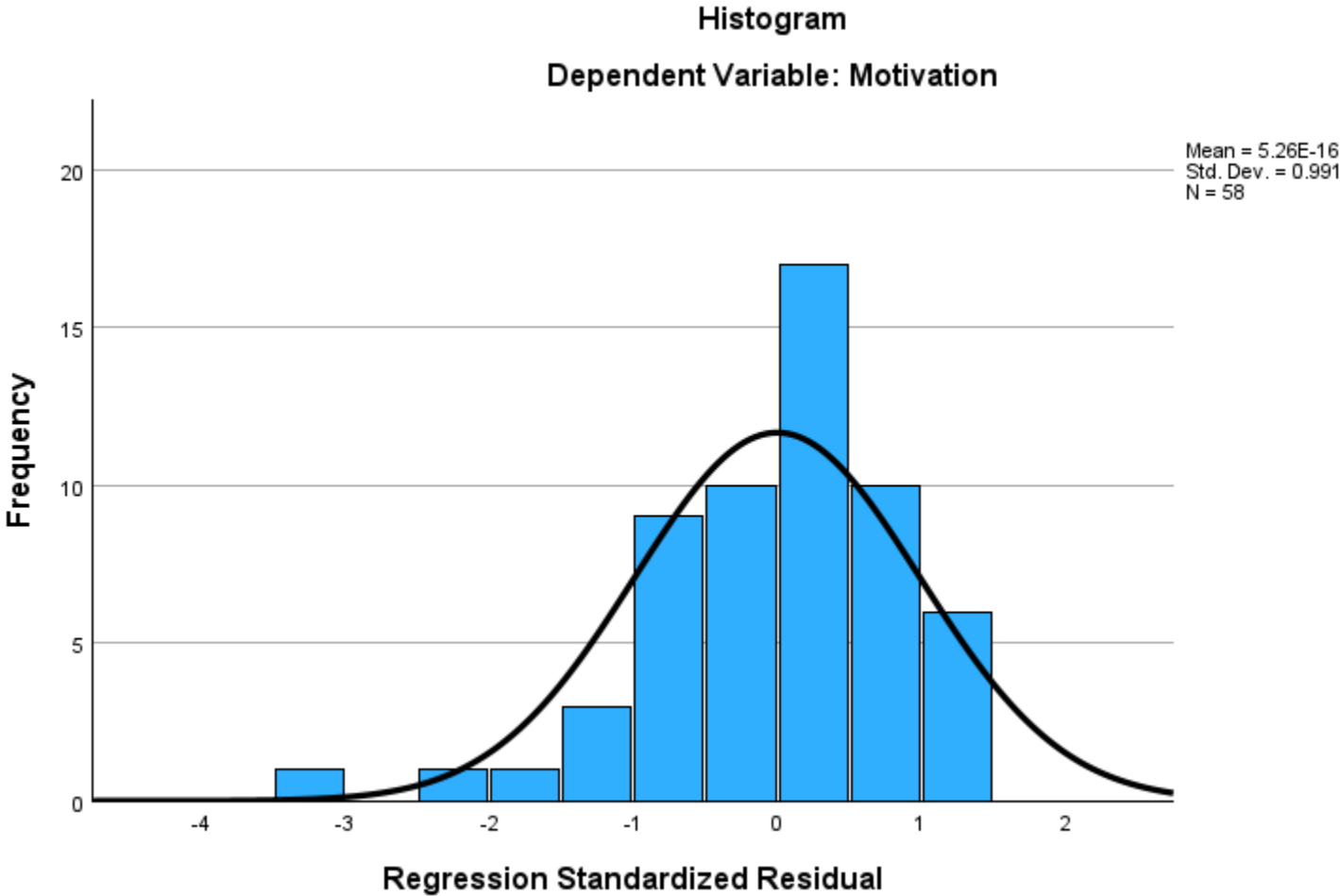


Normal P-P Plot of Regression Standardized Residual

Dependent Variable: Motivation



Simple Regression



Simple Regression

Correlation

Correlations

		Motivation	Job Satisfaction
Pearson Correlation	Motivation	1.000	.421
	Job Satisfaction	.421	1.000
Sig. (1-tailed)	Motivation	.	<.001
	Job Satisfaction	.000	.
N	Motivation	58	58
	Job Satisfaction	58	58

Simple Regression

The table : Independent variable(s) and dependent variable listed

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Job Satisfaction ^b	.	Enter

a. Dependent Variable: Motivation

b. All requested variables entered.

Simple Regression

R Square = proportion of variance of dependent variable (Job satisfaction)

Model Summary^b

Model			Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics				Sig. F Change	Durbin-Watson
	R	R Square				F Change	df1	df2			
1	.421 ^a	.178	.163	.54662	.178	12.090	1	56		<.001	1.931

a. Predictors: (Constant), Job Satisfaction

b. Dependent Variable: Motivation

Simple Regression

- $<.05$ means that the predictor of job satisfaction works better than simply predicting using the mean
- $<.001$ means prediction with job satisfaction is better than prediction without the job satisfaction in the model.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.612	1	3.612	12.090	$<.001^b$
	Residual	16.732	56	.299		
	Total	20.345	57			

a. Dependent Variable: Motivation

b. Predictors: (Constant), Job Satisfaction

Simple Regression

		Coefficients ^a					
Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound Upper Bound
1	(Constant)	2.739	.216		12.676	<.001	2.307 3.172
	Job Satisfaction	.248	.071	.421	3.477	<.001	.105 .390

a. Dependent Variable: Motivation

.248 positively correlated between
job sat and motivation.

Simple Regression Formula

Predictor of Job satisfaction and motivation

$$Y_{mo} = 2.739 + 0.248X_{js}$$

Simple Regression

20 more employees are measured for high job satisfaction. What do you predict their motivation will be ?

These number represent X. add them into the formula

- | | |
|--------|---|
| a. 100 | $Y_{mo} = 2.739 + 0.248 (100) \Rightarrow 27.539$ |
| b. 120 | $Y_{mo} = 2.739 + 0.248 (120) \Rightarrow 32.499$ |
| c. 110 | $Y_{mo} = 2.739 + 0.248 (110) \Rightarrow 30.019$ |
| d. 150 | $Y_{mo} = 2.739 + 0.248 (150) \Rightarrow 39.939$ |

Interpreting results

Unstandardised Beta coefficients – the amount of units the dependant variable changes in relation to one unit change in the predictor variable. If multiple variables are used, this is calculated with other variables held constant. Unstandardised Beta coefficients have associated p values that indicate significance.

Standardised Beta coefficients – compares the strength of the effect of each individual independent variable to the dependent variable. The higher the absolute value of the **beta coefficient**, the stronger the effect. Standardised Beta coefficients have associated p values that indicate significance.

The t-statistic measures how many standard errors the coefficient is away from zero. Generally, any **t-value** greater than +2 or less than – 2 is acceptable. The higher the **t-value**, the greater the confidence that the coefficient as a predictor.

Model fit

- R – correlation coefficient. Shows the strength of the relationship in a model
- R –squared – shows the proportion of variance explained by the predictor variables
- Adjusted R square – adjusts the model for the number of terms(variables) and avoids ‘overfitting’ adding more and more predictor variables because the R square increases with more variables – but this does not mean that the model fit is really better!

In-class activity



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

In-class activity

- What is the value of the correlation ?

$r = +0.421$

In the regression model, What is ANOVA significant ?

Yes (<.001)

In-class activity

What is the r- squared value of predictor variable?

➤ $r^2 = +0.178$

What does the r-squared value imply about the dependent variable ?

➤ 17.8% of the variability in the motivation is due to job satisfaction.

In-class activity

- What is the value of (a) and (b) in simple regression ?

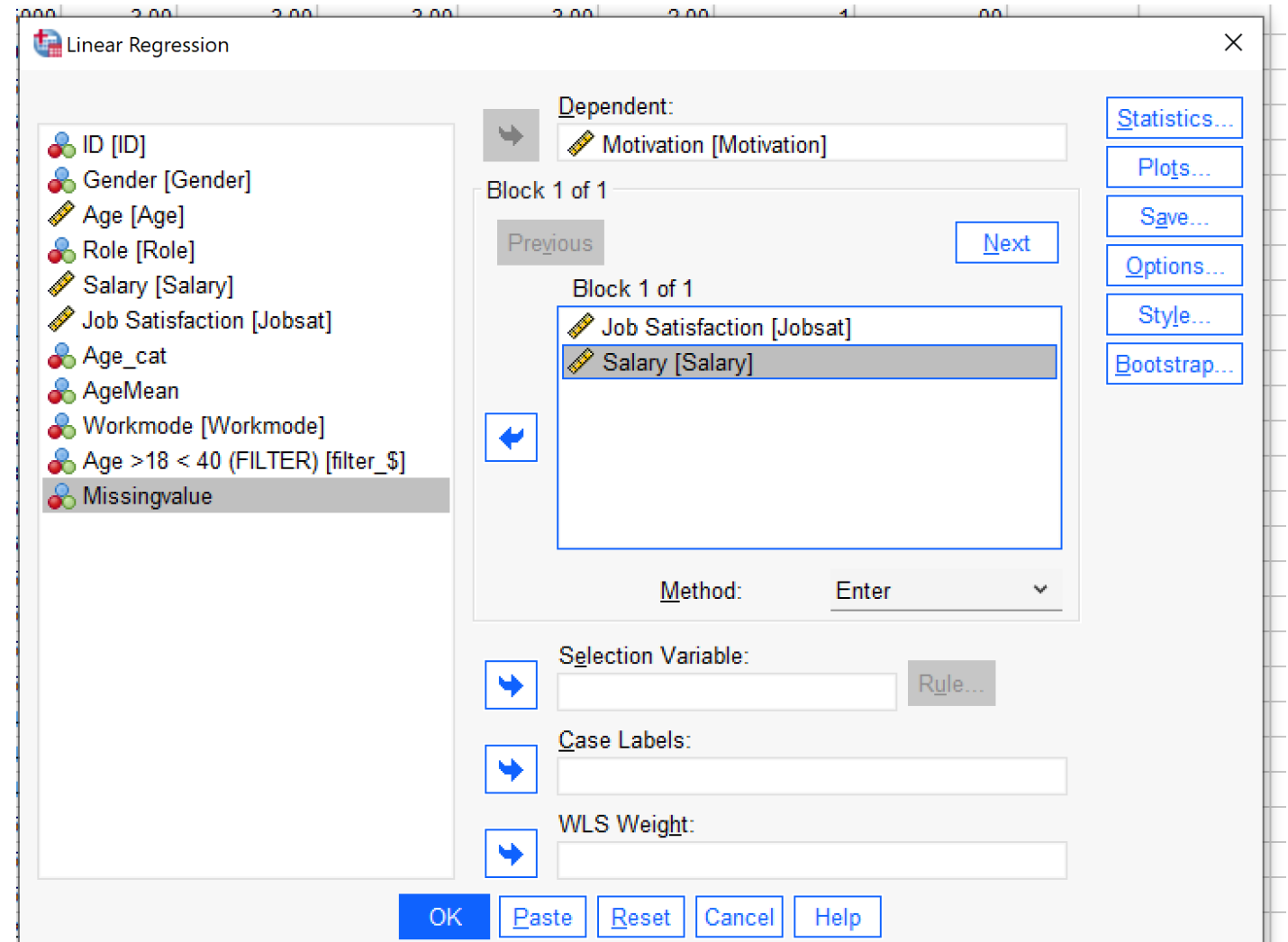
a (constant) = 2.739

b = .248

Differences Between Simple Linear Regression Model and ...

https://www.google.com/search?q=simple+regression+vs+multiple+regression&sca_esv=9967b14e20c7abd4&sca_upv=1&rlz=1C1GCHA_enGB1126GB1126&biw=1642&bih=925&tbm=vid&sxsrf=ADLYWILv42X9G1eesdPFScu9PJFtTv9Axw%3A1726645614840&ei=boXqZov-MpC0hblPu-iTsA0&ved=0ahUKEwiL7ZmUgMyIAxUQWkEAHTv0BNY4ChDh1QMIDQ&uact=5&oq=simple+regression+vs+multiple+regression&gs_lp=Eg1nd3Mtd2l6LXZpZGVvlihzaW1wbGUgcmVncmVzc2lvbiB2cyBtdWx0aXBsZSByZWdyZXNzaW9uMgQQIlggnMgYQABglGB4yBhAAGAgYHjIGEAAyCBgeMgYQABglGB4yCxAAGIAEGiYDGIoFMggQABiABBiBDIIEAAyGAQYogRIjhBQ6AZY9gdwAHgAkAEAmAFUoAG_AaoBATO4AQPIAQD4AQGYAgOgAsYBwglFEAAyGATCAgcQIxiwAhgnwglGEAAyBxgewglIEAAyBxglGB6YAwCIBgGSBwEzoAeOlw&sclient=gws-wiz-video#fpstate=ive&vld=cid:34b273b1,vid:NU8OlwEnEk4,st:0

Multivariable (Multiple) Linear Regression



The image shows the 'Linear Regression' dialog box in SPSS. On the left, a list of variables includes ID [ID], Gender [Gender], Age [Age], Role [Role], Salary [Salary], Job Satisfaction [Jobsat], Age_cat, AgeMean, Workmode [Workmode], Age >18 < 40 (FILTER) [filter_\$], and Missingvalue. The 'Dependent' variable is 'Motivation [Motivation]'. The 'Block 1 of 1' list contains 'Job Satisfaction [Jobsat]' and 'Salary [Salary]'. The 'Method' is set to 'Enter'. On the right, there are buttons for 'Statistics...', 'Plots...', 'Save...', 'Options...', 'Style...', and 'Bootstrap...'. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

Linear Regression

Dependent: Motivation [Motivation]

Block 1 of 1

Previous Next

Block 1 of 1

Job Satisfaction [Jobsat]
Salary [Salary]

Method: Enter

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Multivariable (Multiple) Linear Regression

Salary	Jobsat	Motivation	Age_cat	AgeMean	Workmode	filter_\$	Missingvalue	var	var	var
25000	3.00	3.00	3.00	3.00	2.00	1	.00			

Linear Regression: Statistics

Regression Coefficients

☒ Estimates
☒ Confidence intervals
Level(%): 95
☐ Covariance matrix

☒ Model fit
☒ R squared change
☒ Descriptives
☐ Part and partial correlations
☒ Collinearity diagnostics
☐ Selection criteria

Residuals

☐ PRESS
☒ Durbin-Watson
☒ Casewise diagnostics

☒ Outliers outside: 3 standard deviations
☐ All cases

Continue

Cancel

Help

Multivariable (Multiple) Linear Regression

Salary	Jobsat	Motivation	Age_cat	AgeMean	Workmode	filter_\$	Missingvalue	var	var	var
25000	2.00	2.00	2.00	2.00	2.00	1	00			

Linear Regression: Plots

DEPENDNT

*ZPRED

*ZRESID

*DRESID

*ADJPRED

*SRESID

*SDRESID

Scatter 1 of 1

Previous

Next

Y:

*ZRESID

X:

*ZPRED

Standardized Residual Plots

☒ Histogram

☒ Normal probability plot

☐ Produce all partial plots

Continue

Cancel

Help

Multivariable (Multiple) Linear Regression

Correlations

		Motivation	Job Satisfaction	Salary
Pearson Correlation	Motivation	1.000	.421	.079
	Job Satisfaction	.421	1.000	.643
	Salary	.079	.643	1.000
Sig. (1-tailed)	Motivation	.	<.001	.277
	Job Satisfaction	.000	.	.000
	Salary	.277	.000	.
N	Motivation	58	58	58
	Job Satisfaction	58	58	58
	Salary	58	58	58

Multivariable (Multiple) Linear Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Salary, Job Satisfaction ^b	.	Enter

a. Dependent Variable: Motivation

b. All requested variables entered.

Two predictors : Job satisfaction and Salary

Multivariable (Multiple) Linear Regression

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics				Durbin-Watson
						F Change	df1	df2	Sig. F Change	
1	.490 ^a	.240	.213	.53017	.240	8.691	2	55	<.001	2.027

a. Predictors: (Constant), Salary, Job Satisfaction

b. Dependent Variable: Motivation

Multivariable (Multiple) Linear Regression

P < .05

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	VIF
		B	Std. Error	Beta			Lower Bound	Upper Bound		
1	(Constant)	3.448	.394		8.763	<.001	2.660	4.237		
	Job Satisfaction	.371	.090	.631	4.114	<.001	.190	.552	.587	1.704
	Salary	-4.469E-5	.000	-.327	-2.128	.038	.000	.000	.587	1.704

a. Dependent Variable: Motivation

Tolerance value = 0.587
Statistics VIF = 1.704.

- End -



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David