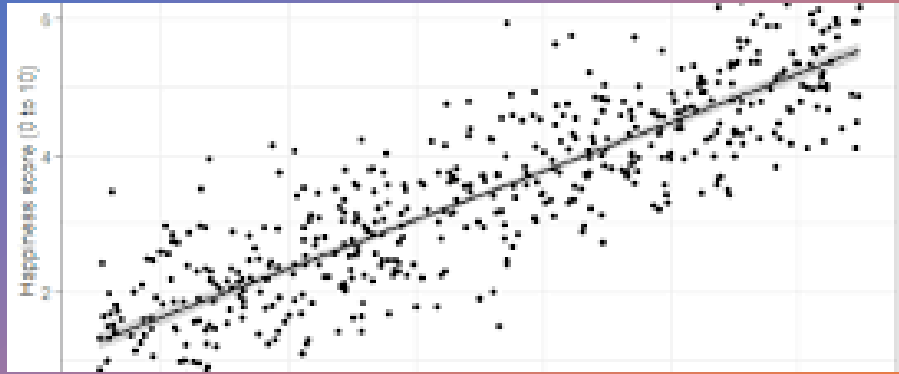
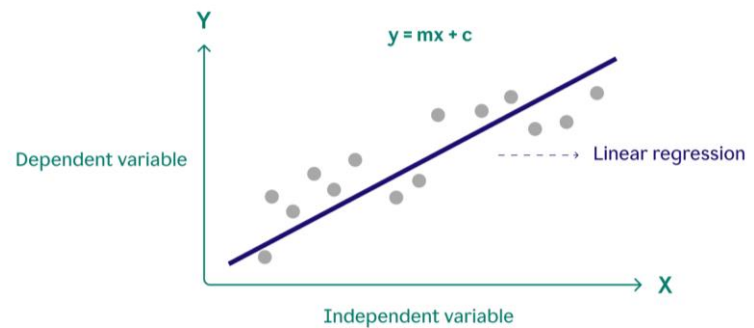




LCDB7011 Quantitative Research Methods



Week 10&11 – Simple Linear Regression
Dr. Helen LEE

By the end of
this session, you
should ...



Factor Analysis



Simple regression model



Analysis and Interpretation

By the end of this session, you should conduct...

Factor
Analysis -

Principal
Components
Analysis

Exploratory
Factor
Analysis

By the end of this session, you should ...

1. Know about simple regression model
2. Understand the Regression Analysis and its interpretation of the SPSS Output

$$Y_i = \beta_0 + \beta_1 X_i$$

The diagram illustrates the components of the simple regression equation $Y_i = \beta_0 + \beta_1 X_i$. It features four labels with arrows pointing to specific parts of the equation: 'Dependent Variable' points to Y_i , 'Constant/Intercept' points to β_0 , 'Slope/Coefficient' points to β_1 , and 'Independent Variable' points to X_i .

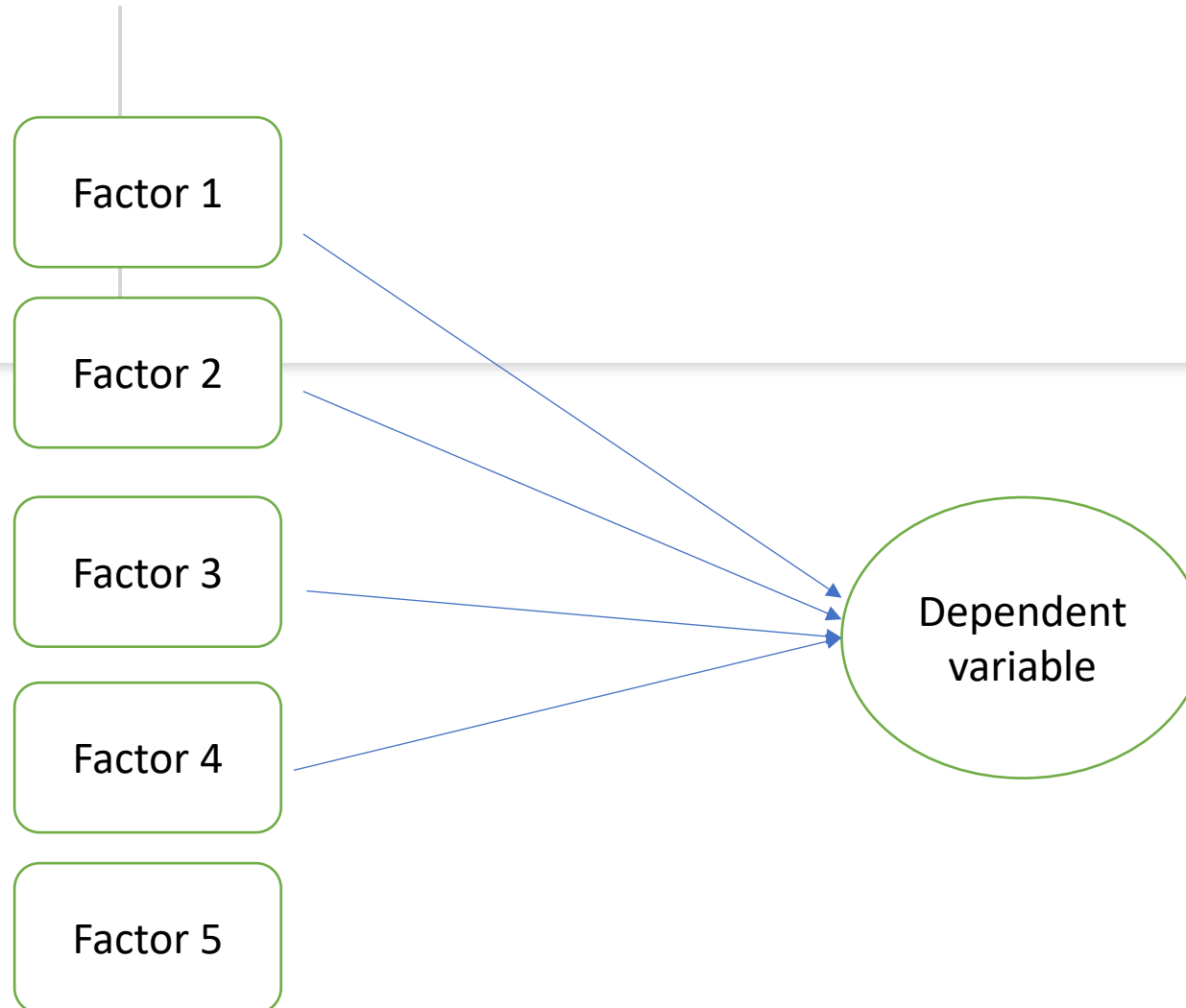
Dependent Variable

Constant/Intercept

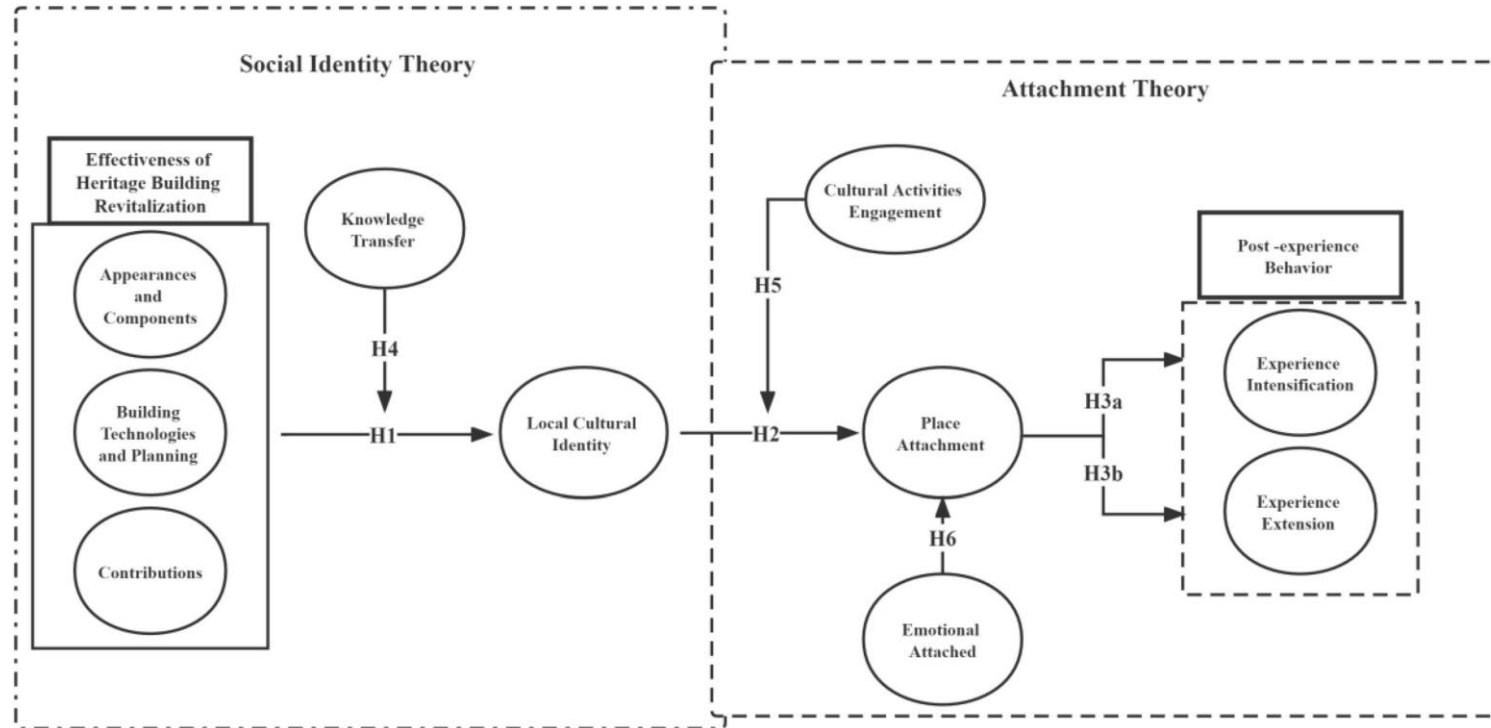
Slope/Coefficient

Independent Variable

Factor Analysis



Factor Analysis



Source : Chan, et al., (2024)

Factor Analysis

Factor analysis attempts to identify underlying variables, or **factors**, that explain the pattern of correlations within a set of observed variables.

It is often used in **data reduction** to identify a small number of factors that explain most of the variance that is observed in a much larger number of manifest variables.

Factor analysis can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis (for example, to identify collinearity prior to performing a linear regression analysis).



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

Factor Analysis - Assumption

Data. The variables should be quantitative at the *interval or ratio level*. Categorical data (such as religion or country of origin) are not suitable for factor analysis.

Assumptions. The data should have a bivariate normal distribution for each pair of variables, and observations should be independent. The factor analysis model specifies that variables are determined by common factors (the factors estimated by the model) and unique factors (which do not overlap between observed variables); the computed estimates are based on the assumption that all unique factors are uncorrelated with each other and with the common factors.



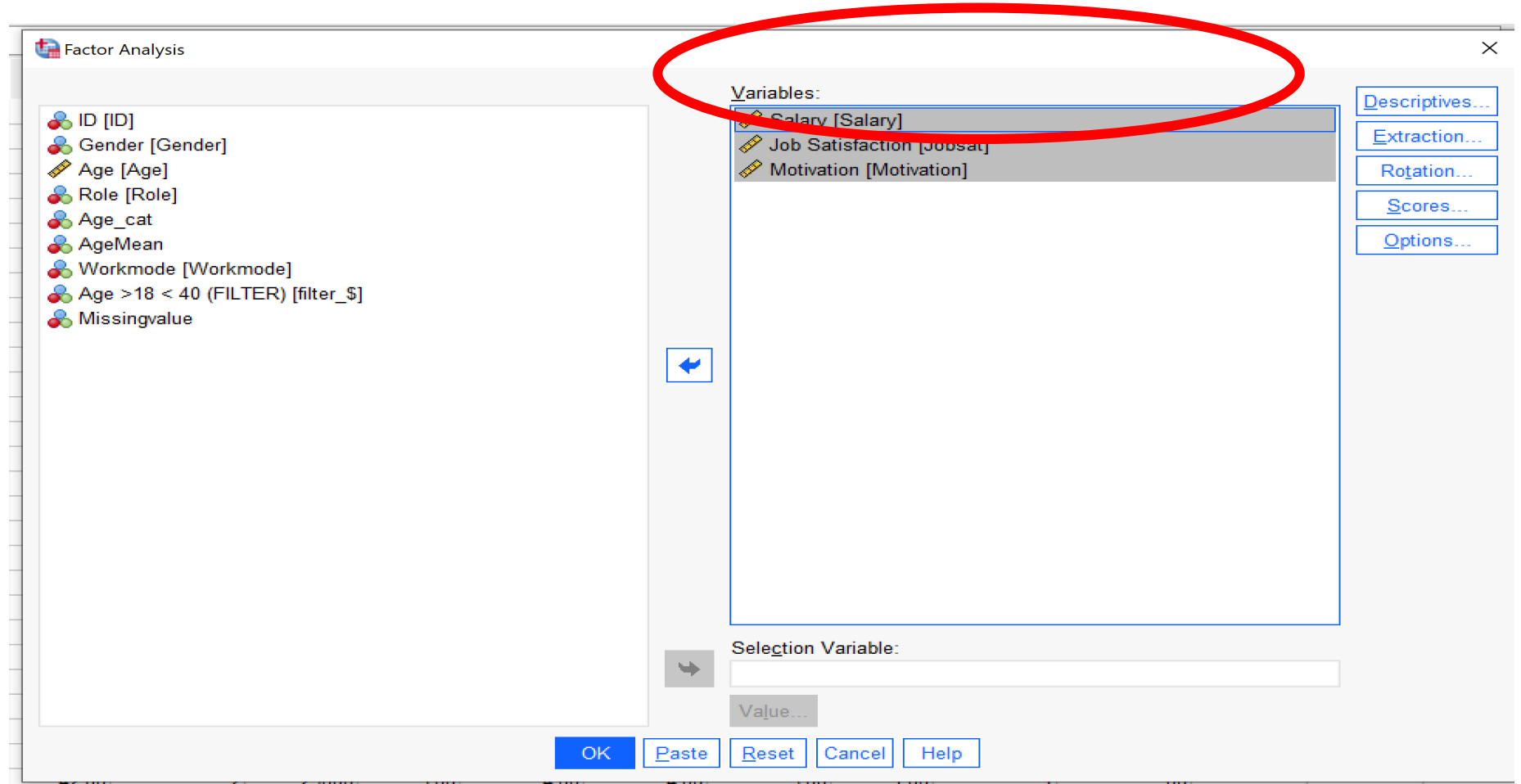
Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Go to Analyze => Dimension
Reduction => Factor



Factor Analysis

Scale factor



Factor Analysis

Go to
descriptives

Factor Analysis: Descriptives

Statistics

- ☒ Univariate descriptives
- ☒ Initial solution

Correlation Matrix

- ☒ Coefficients ☐ Inverse
- ☐ Significance levels ☐ Reproduced
- ☒ Determinant ☐ Anti-image
- ☒ KMO and Bartlett's test of sphericity

☐ Covariance matrix

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



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

Factor Analysis

Go to descriptives
Eigenvalues >1

Factor Analysis: Extraction

Method: Principal components

Analyze

- ☒ Correlation matrix
- ☐ Covariance matrix

Display

- ☒ Unrotated factor solution
- ☒ Scree plot

Extract

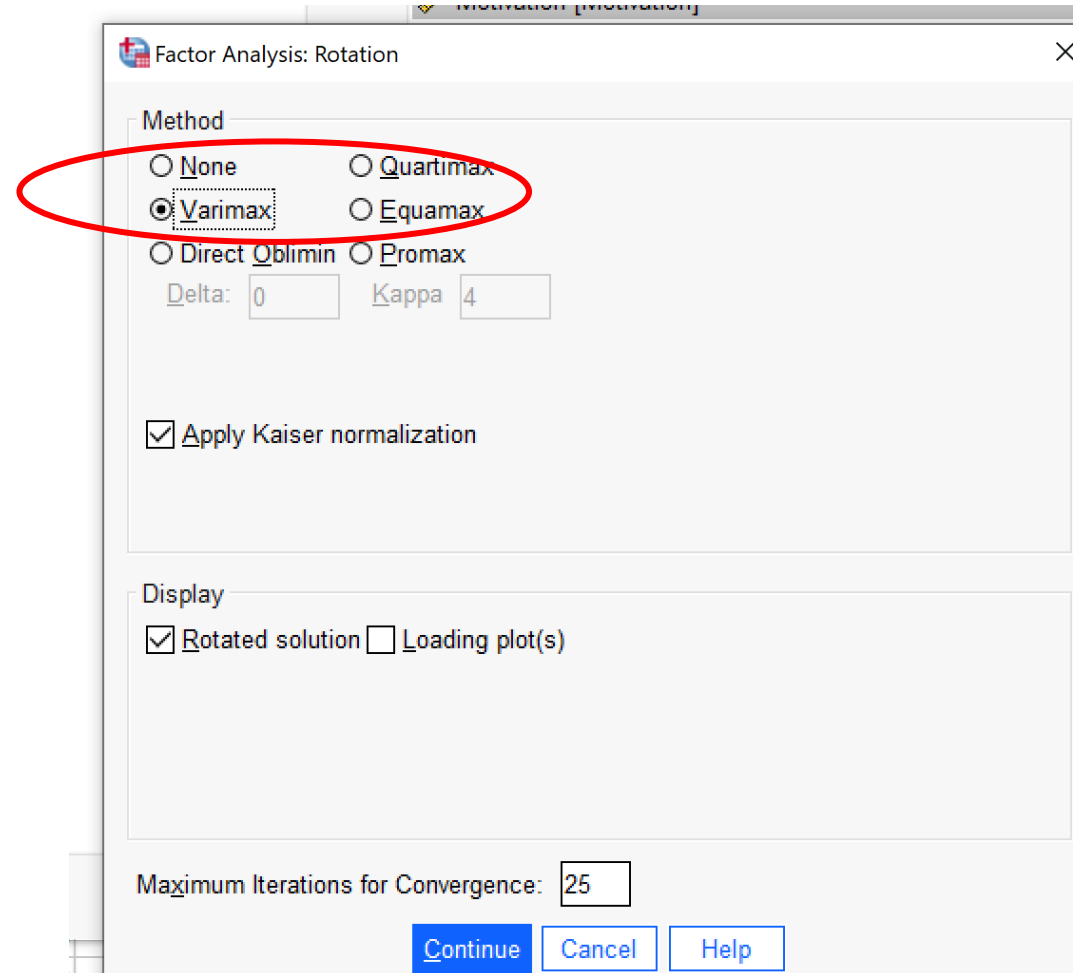
- ☒ Based on Eigenvalue
Eigenvalues greater than: 1
- ☐ Fixed number of factors
Factors to extract:

Maximum Iterations for Convergence: 25

Continue Cancel Help

Factor Analysis

Go to Rotation



Factor Analysis: Rotation

Method

☐ None ☐ Quartimax

☒ Varimax ☐ Equamax

☐ Direct Oblimin ☐ Promax

Delta: 0 Kappa: 4

☒ Apply Kaiser normalization

Display

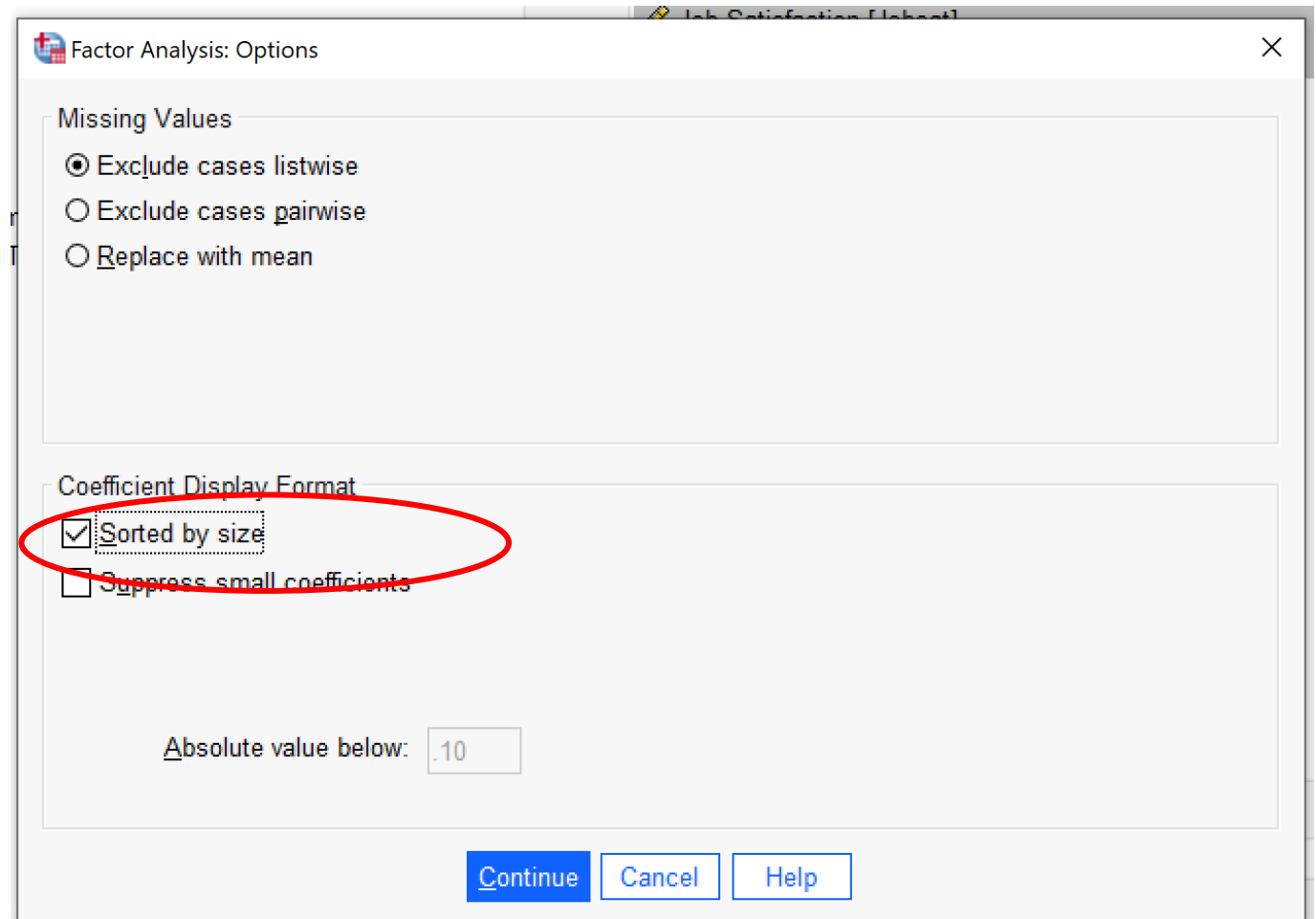
☒ Rotated solution ☐ Loading plot(s)

Maximum Iterations for Convergence: 25

Continue Cancel Help

Factor Analysis

Go to Options



Factor Analysis: Options

Missing Values

- ☒ Exclude cases listwise
- ☐ Exclude cases pairwise
- ☐ Replace with mean

Coefficient Display Format

- ☒ Sorted by size
- ☐ Suppress small coefficients

Absolute value below: .10

Continue Cancel Help

Factor Analysis

SPSS Output

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Salary	23758.62	4365.175	58
Job Satisfaction	2.8621	1.01650	58
Motivation	3.4483	.59743	58



Factor Analysis

SPSS Output
< -0.8 < correlation
matrix value < 0.8

Correlation Matrix^a

		Salary	Job Satisfaction	Motivation
Correlation	Salary	1.000	.643	.079
	Job Satisfaction	.643	1.000	.421
	Motivation	.079	.421	1.000

a. Determinant = .446

Should be >.00001



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

Factor Analysis

SPSS Output

KMO value : 0.8 to 0.9

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.877
Bartlett's Test of Sphericity	Approx. Chi-Square	4354.035
	df	210
	Sig.	<.001



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

Factor Analysis

SPSS Output

Communalities

	Initial	Extraction
JS1	1.000	.652
OC1	1.000	.616
OC2	1.000	.776
EP1	1.000	.633
OC3	1.000	.641
OC4	1.000	.735
EP2	1.000	.731
EP3	1.000	.705
AC1	1.000	.762
EP4	1.000	.748
JS2	1.000	.649
JS3	1.000	.575
AC2	1.000	.759
SI1	1.000	.741
JS4	1.000	.607
SI2	1.000	.793
JS5	1.000	.625
AC3	1.000	.773
SI3	1.000	.685
AC4	1.000	.749
SI4	1.000	.781

Extraction Method: Principal Component Analysis.



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

Factor Analysis

SPSS Output
Eigenvalue > 1
No. of extractions

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.280	29.904	29.904	6.280	29.904	29.904	3.111	14.812	14.812
2	2.947	14.033	43.937	2.947	14.033	43.937	3.083	14.682	29.494
3	2.350	11.191	55.127	2.350	11.191	55.127	3.075	14.642	44.136
4	1.667	7.936	63.063	1.667	7.936	63.063	2.835	13.498	57.633
5	1.491	7.098	70.161	1.491	7.098	70.161	2.631	12.528	70.161
6	.615	2.930	73.091						
7	.563	2.683	75.774						
8	.529	2.519	78.293						
9	.517	2.463	80.757						
10	.482	2.294	83.051						
11	.431	2.055	85.105						
12	.430	2.047	87.152						
13	.398	1.894	89.046						
14	.374	1.779	90.825						
15	.348	1.656	92.481						
16	.317	1.510	93.992						
17	.304	1.445	95.437						
18	.278	1.324	96.761						
19	.241	1.150	97.911						
20	.234	1.114	99.025						
21	.205	.975	100.000						

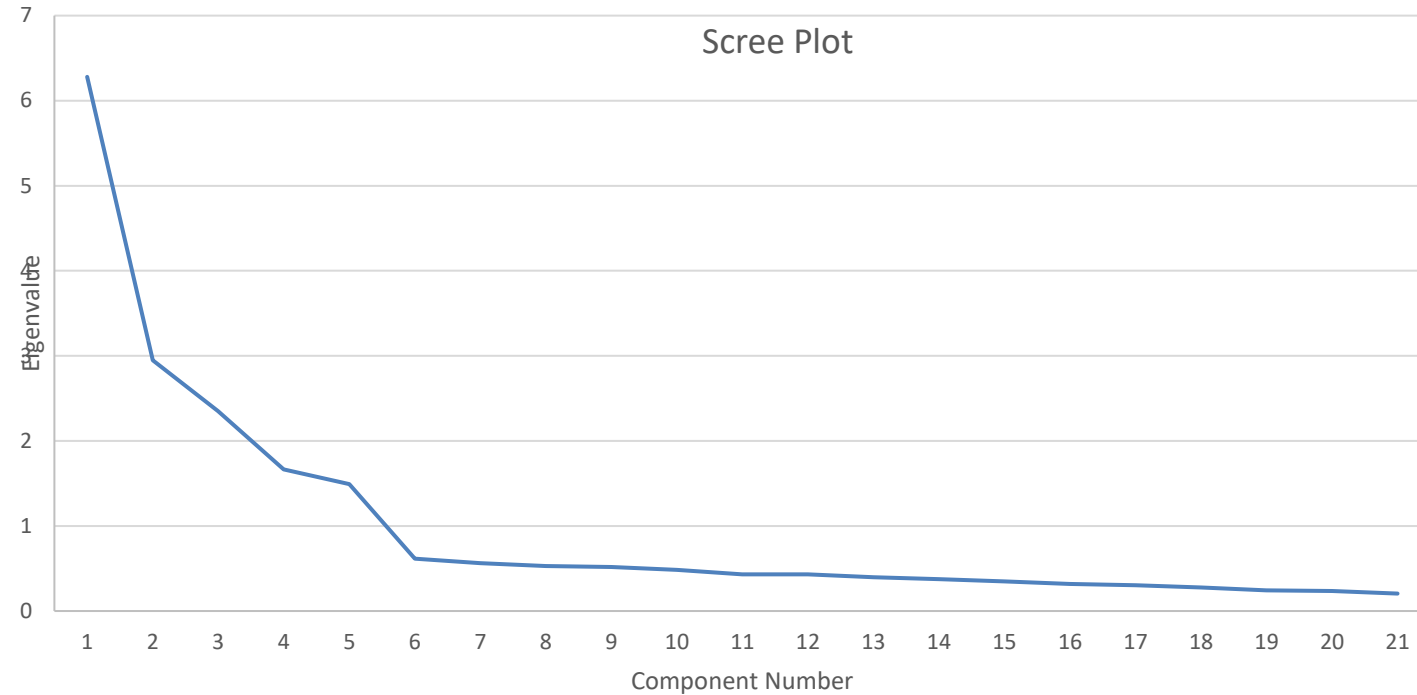
Extraction Method: Principal Component Analysis.



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

Factor Analysis

SPSS Output Screen Plot



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



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.

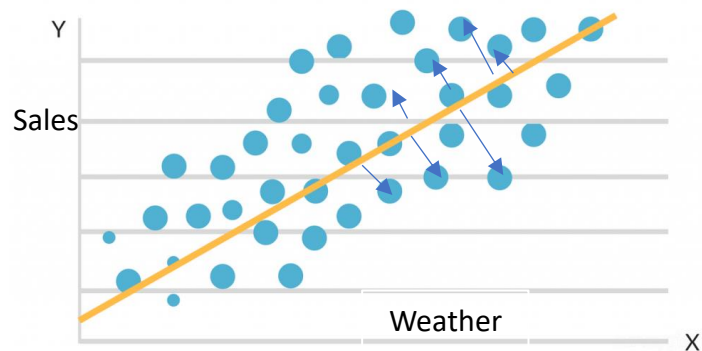
“Best fit” here means that the **error term, the distance from each point to the line**, is minimized. Since the relationship between variables is probably not completely linear and because there are other factors outside the scope of the study the line does not go through all the data points. The distance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

Simple Regression



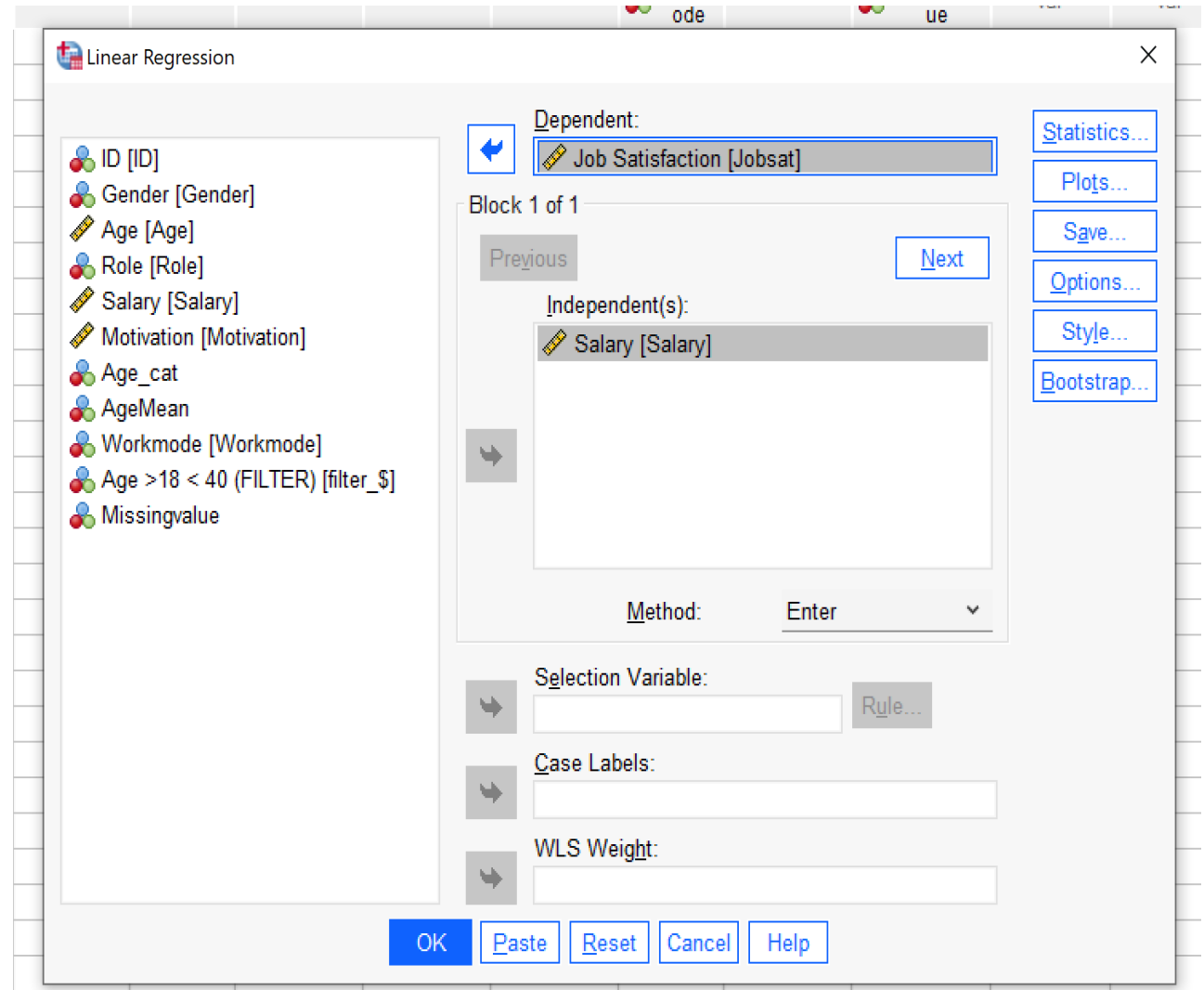
Simple Regression

Go to Analyze
=>Regression =>
Linear

The screenshot shows the SPSS software interface. The 'Analyze' menu is open, and the 'Regression' option is selected, which has opened a sub-menu. In this sub-menu, the 'Linear...' option is highlighted. The background shows a data grid with columns labeled EP2, EP3, AC1, EP4, JS2, JS3, and AC. The 'Linear...' option in the sub-menu is preceded by an 'R' icon with 'LIN' below it.

	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	3	2	
				4	2	1	
				4	5	5	
				4	1	2	
				4	3	1	
				5	4	3	
				6	3	1	
				5	4	2	
				4	3	3	
				6	3	2	

Simple Regression



The image shows the 'Linear Regression' dialog box in SPSS. On the left, a list of variables is shown: ID [ID], Gender [Gender], Age [Age], Role [Role], Salary [Salary], Motivation [Motivation], Age_cat, AgeMean, Workmode [Workmode], Age >18 < 40 (FILTER) [filter_\$], and Missingvalue. In the 'Dependent:' field, 'Job Satisfaction [Jobsat]' is selected. In the 'Independent(s):' field, 'Salary [Salary]' is selected. 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'. The 'Previous' and 'Next' buttons are also visible near the 'Independent(s):' field.

Linear Regression

Dependent: Job Satisfaction [Jobsat]

Block 1 of 1

Previous Next

Independent(s): Salary [Salary]

Method: Enter

Selection Variable: Rule...

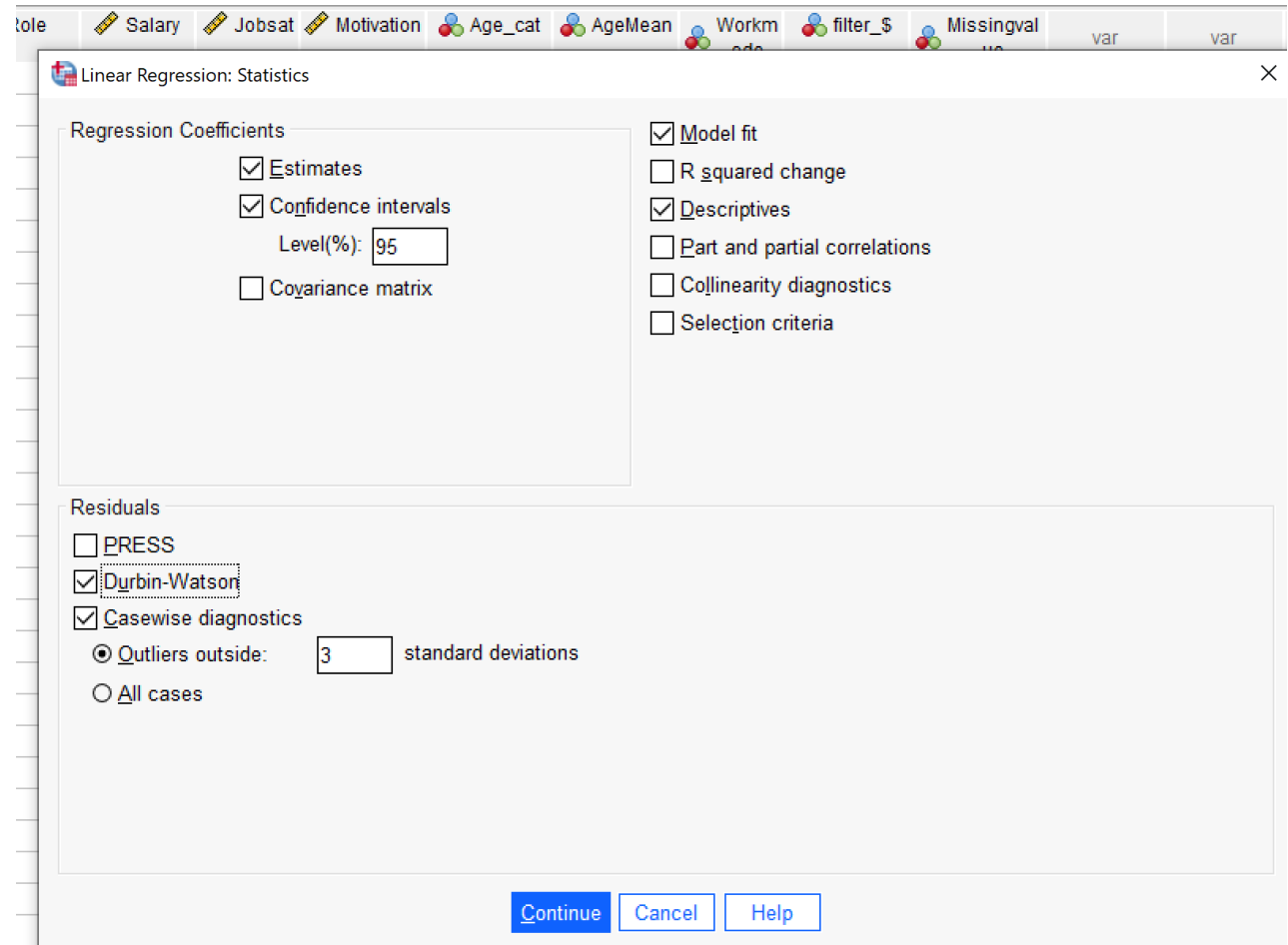
Case Labels:

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. Under 'Casewise diagnostics', '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

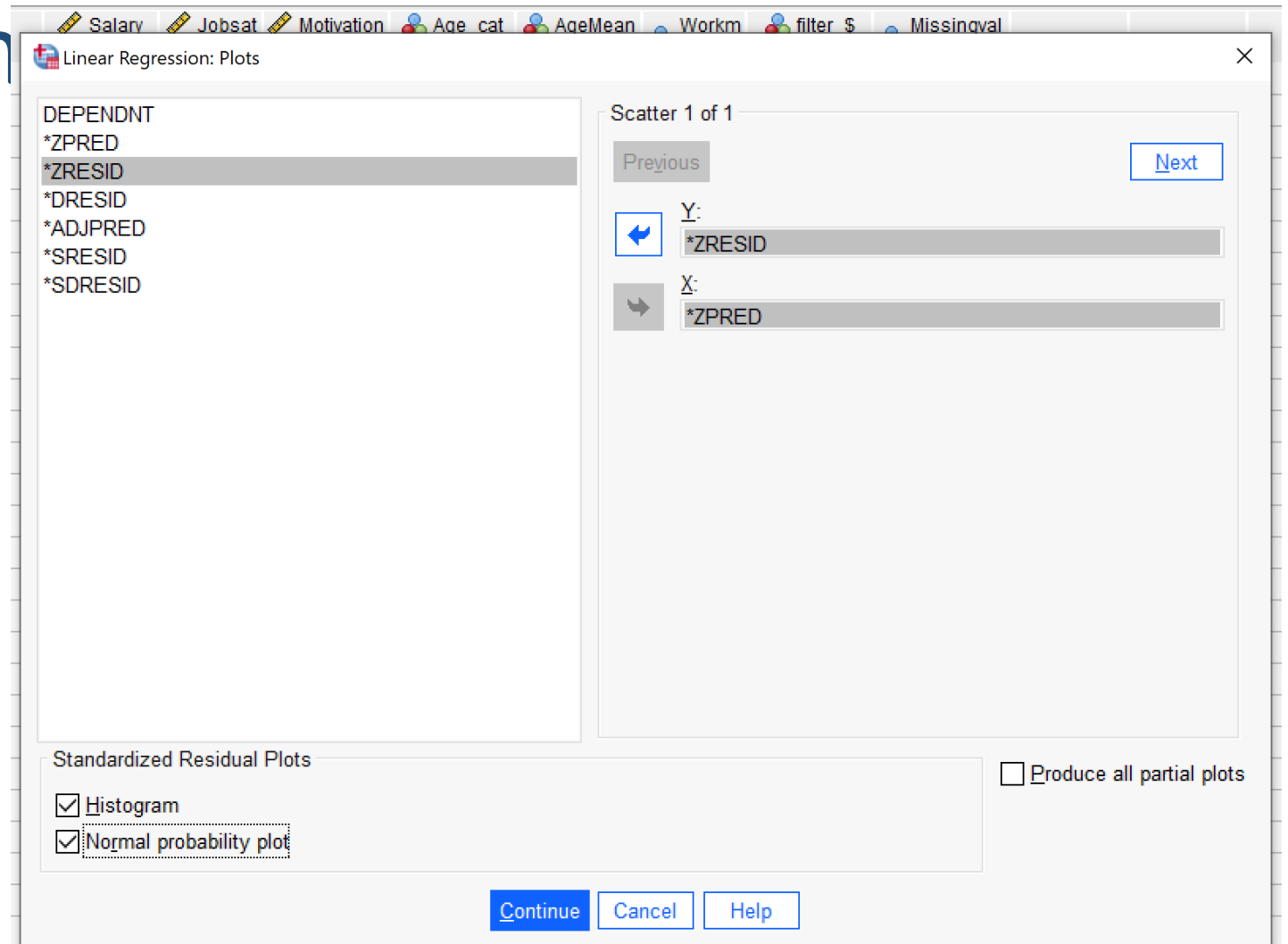
Continue Cancel Help

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

SPSS Output
-3.29 <Residual value<
+3.29

Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.7008	4.8440	2.8621	.65337	58
Residual	-1.89820	1.32632	.00000	.77870	58
Std. Predicted Value	-1.777	3.033	.000	1.000	58
Std. Residual	-2.416	1.688	.000	.991	58

a. Dependent Variable: Job Satisfaction

Simple Regression

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.643 ^a	.413	.403	.78562	1.276

a. Predictors: (Constant), Salary

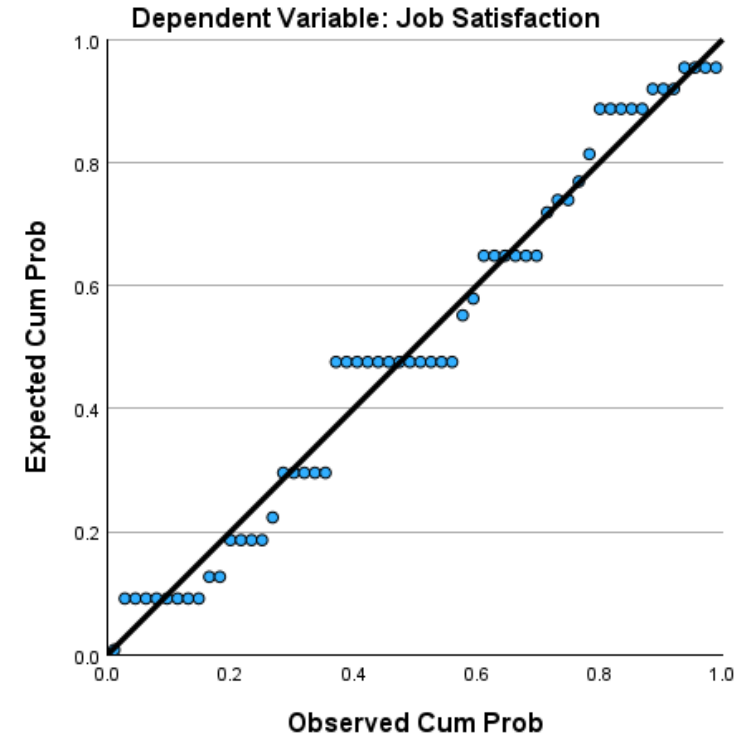
b. Dependent Variable: Job Satisfaction

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

Simple Regression

Assumption of Normality :
QQ Plot : along 45°

Normal P-P Plot of Regression Standardized Residual



Simple Regression

Correlations

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24.333	1	24.333	39.425	<.001 ^b
	Residual	34.563	56	.617		
	Total	58.897	57			

a. Dependent Variable: Job Satisfaction

b. Predictors: (Constant), Salary

Simple Regression

		Unstandardized Coefficients		Coefficients ^a				95.0% Confidence Interval for B	
Model		B	Std. Error	Standardized Coefficients Beta	t	Sig.	Lower Bound	Upper Bound	
1	(Constant)	-.694	.576		-1.206	.233	-1.847	.459	
	Salary	.000	.000	.643	6.279	<.001	.000	.000	

a. Dependent Variable: Job Satisfaction

.643 positively correlated
between salary and job sat.

Simple Regression

Predictor of salary and Job satisfaction

$$Y_{js} = -.694 + X_s$$

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

Wrap Up



By the end of this session, you should be able to understand and conduct...

Factor
Analysis -

Principal
Components
Analysis

Exploratory
Factor
Analysis

Q&A





Assignment
Submission due :
26th May 2026
