



LCDB7011 Quantitative Research Methods

Week 10&11 – Simple Linear
Regression

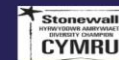
Dr. Badir Miftah



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

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

1. Factor Analysis
2. Simple regression model
3. Analysis and Interpretation



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

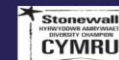
By the end of this session, you should ...

1. Factor Analysis

Principal Components Analysis

Exploratory Factor Analysis

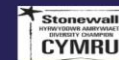
Factor Analysis Using SPSS



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

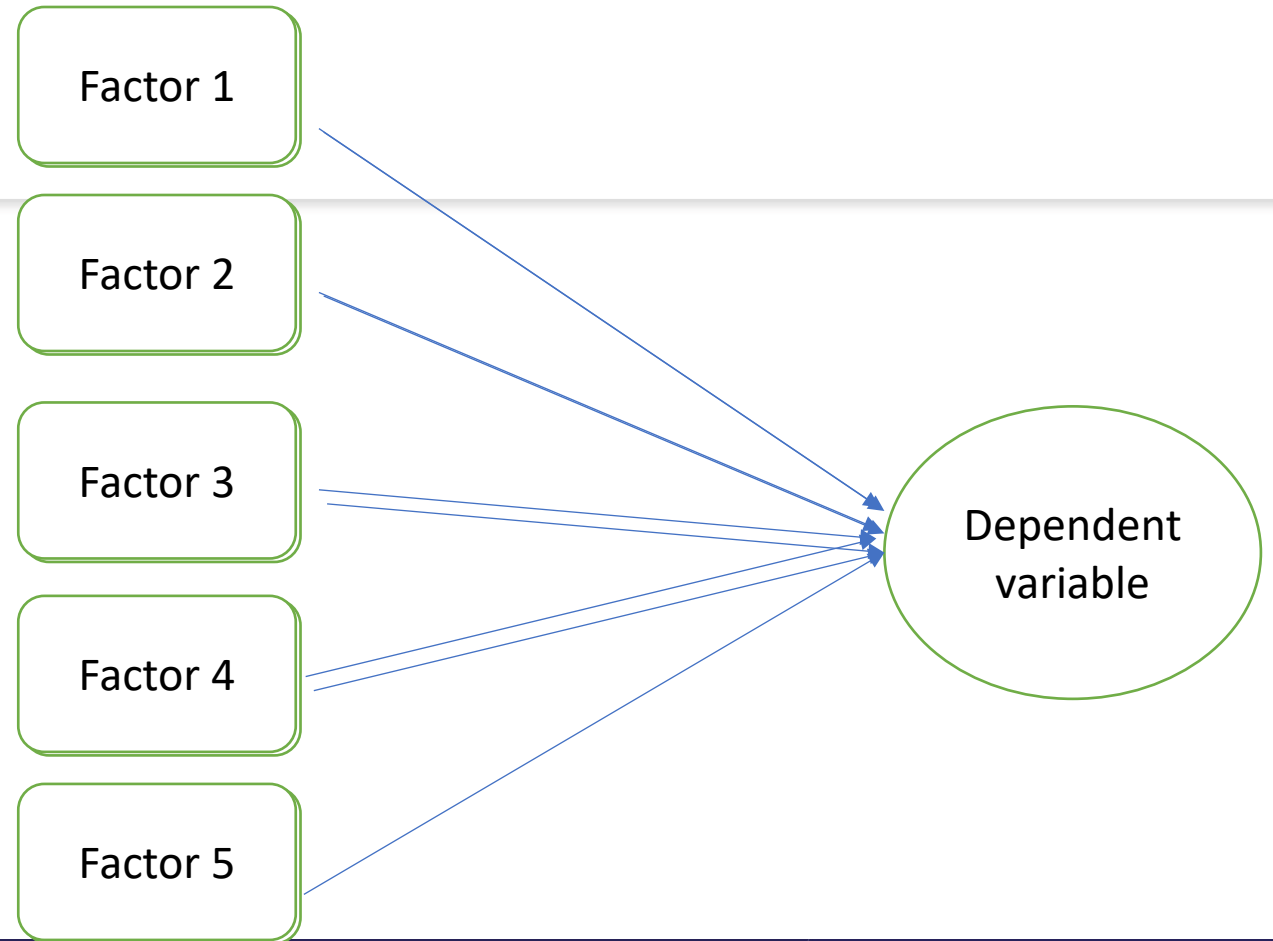
By the end of this session, you should ...

1. Know about simple regression model
2. Analysis and Interpretation

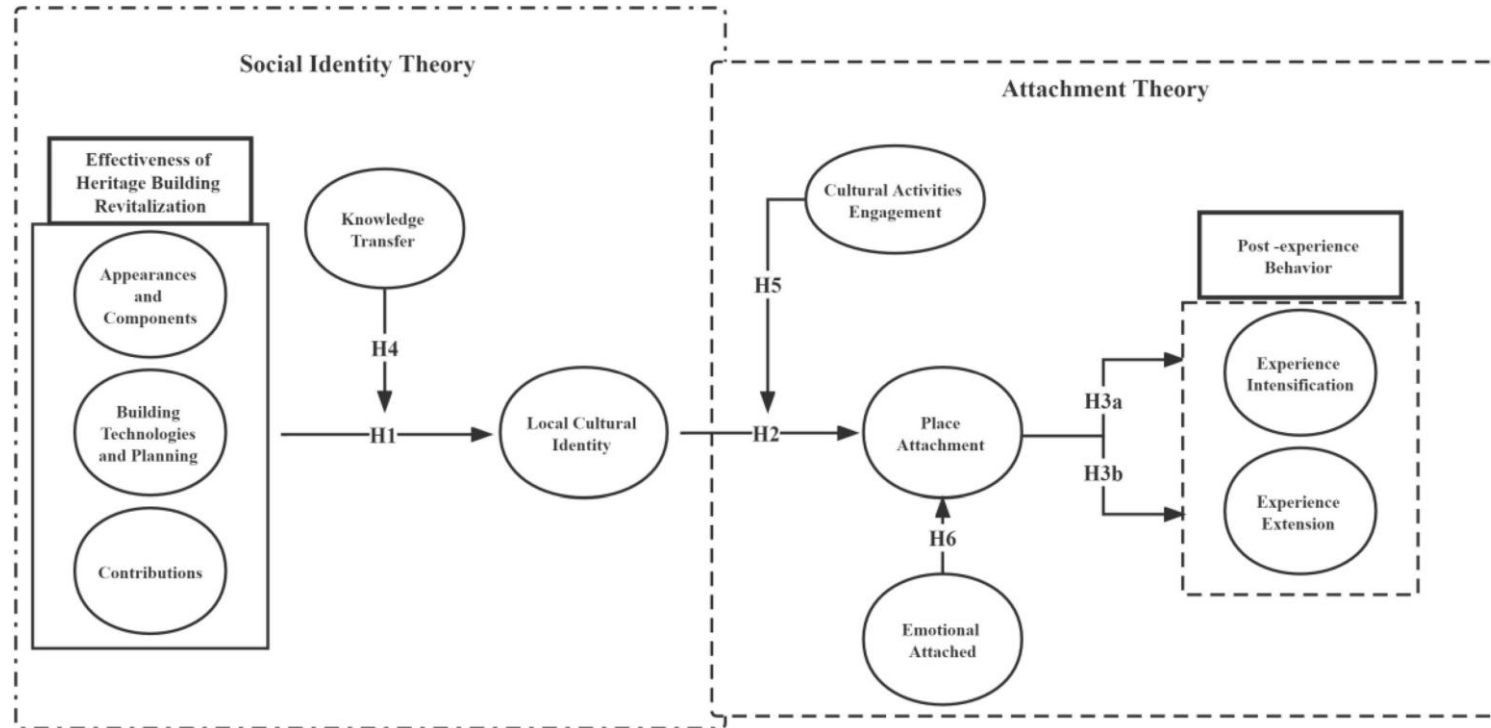


Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Factor Analysis



Factor Analysis



Source : Chan, et al., (2024)



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

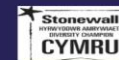
Factor Analysis

- 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).



Factor Analysis - Assumption

- **Type of Data:** Use **numerical data** (interval or ratio).
Not suitable for categories like religion or country.
- **Normality:** Data should follow a **normal distribution** between variable pairs.
- **Independence:** Observations must be **independent** of each other.
- **Model Idea:**
 - Variables are influenced by **common factors** (shared)
 - and **unique factors** (individual).
 - These **unique factors should not be related** to each other or the common ones.

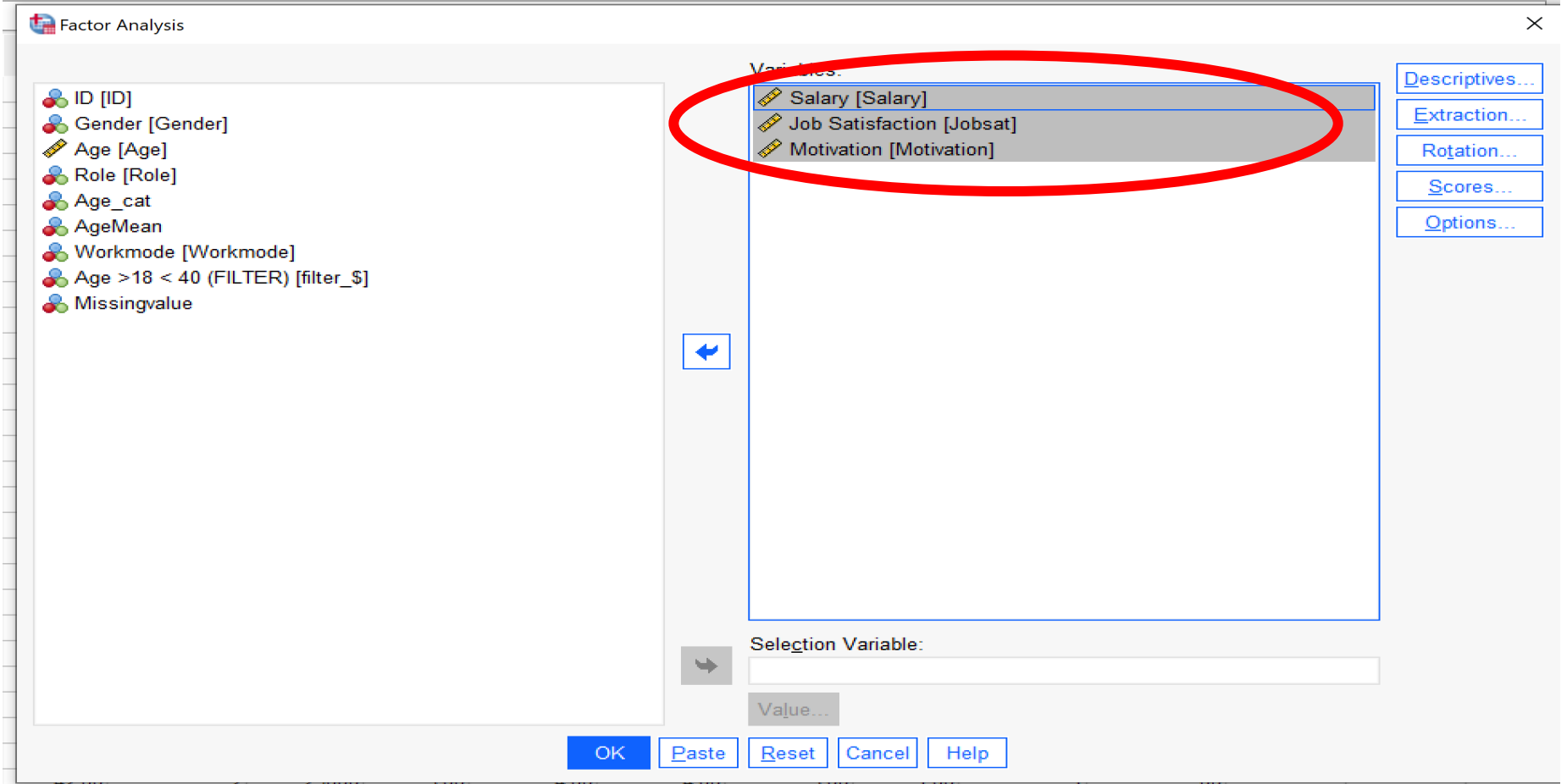


Go to Analyze => Dimension
Reduction => Factor



Factor Analysis

Scale factor



Factor Analysis

Go to
descriptives

Factor Analysis: Descriptives

Statistics

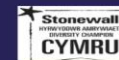
- ☒ Univariate descriptives
- ☒ Initial solution

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

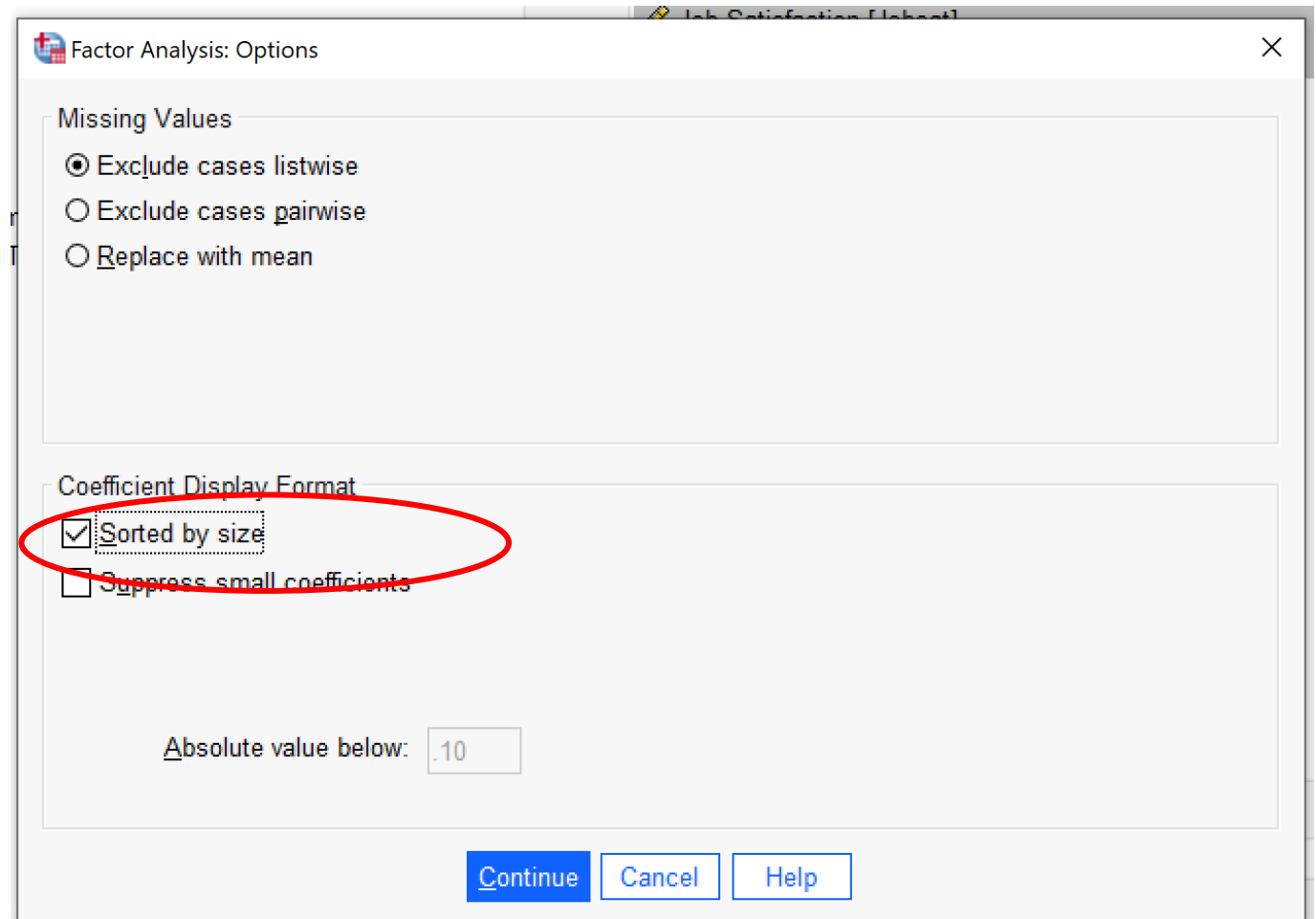
☒ Rotated solution ☐ Loading plot(s)

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						

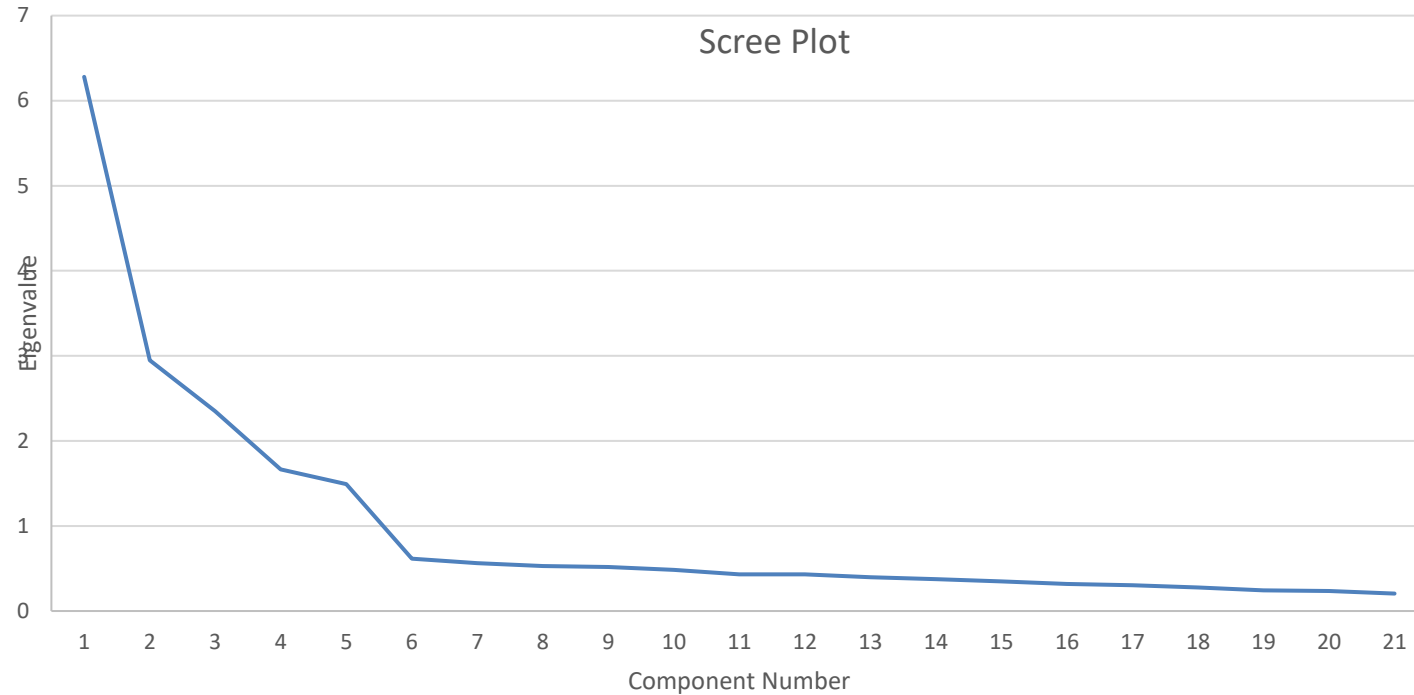
Extraction Method: Principal Component Analysis.



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

Factor Analysis

SPSS Output Screen Plot



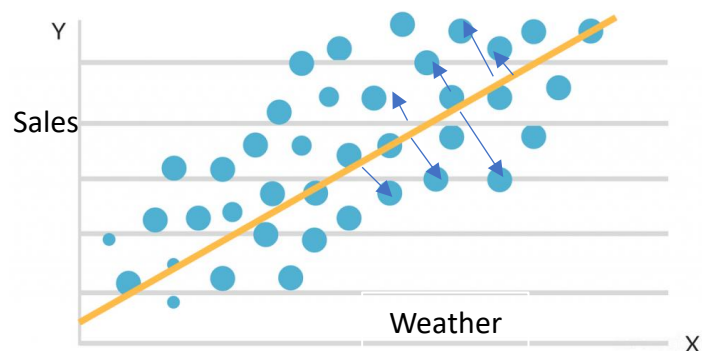
Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

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. [Linear Regression in SPSS](#).



Regression



The dependant variable is always plotted on the Y axis

- **Purpose:** To see if one or more things (independent variables) affect something else (a dependent variable).

Example: Does weather affect sales?

- **How it works:**

- Plot data points on a graph.
- Draw a **line of best fit** through them.
- This line helps us **predict** values.

- **Best Fit Line:**

- It doesn't go through every point.
- The gaps between points and the line are called **errors** (or **residuals**).
- The goal is to make these errors as small as possible.

- **Example Hypotheses:**

- H0: Weather has no effect on sales
- H1: Poor weather increases sales

Weather



Sales

Assumptions of regression

- **Linear Relationship**

- The dependent and independent variables should follow a **straight-line pattern**.
- Use **scatterplots** to check this.

- **Normality of Residuals**

- The **errors (residuals)** should follow a **normal distribution**.
- Check using **QQ plots**, **Descriptives**, or **Kolmogorov-Smirnov test** (should *not* be significant).

- **No Multicollinearity**

- The independent variables should **not be too similar** to each other.
- Check **VIF values** – they should be **below 10**.

- **Homoscedasticity**

- The **spread of errors** should be **even** across all values.
- Use a **residual plot** (ZPRED vs ZRESID) to check this.

Key Terms

- **ZPRED:** Stands for *Standardized Predicted Value*.
→ It shows what the model *predicts* for each case, after converting values into a standard scale.
- **ZRESID:** Stands for *Standardized Residual*.
→ This shows how *far off* the model's prediction was, also on a standard scale.
- **Residual:** The difference between the **actual** value and the **predicted** value. It's the model's error for each data point.
- **VIF (Variance Inflation Factor):** A number that shows if your independent variables are **too similar** to each other (multicollinearity).
 - If $VIF > 10$, that's a problem.
- **Multicollinearity:** When two or more independent variables are **strongly related** to each other, which can confuse the model.
- **Cook's Distance** is a test used to find out **if one data point is having too much influence** on your regression model.
- **R-squared (R^2)** tells you **how much of the change** in your dependent variable (Y) is **explained by your independent variable(s) (X)**.

Key Terms

- **Homoscedasticity:**
→ The **errors** (residuals) should spread out **evenly** across all values, not bunch up or fan out.
- **Normality of Residuals:**
→ The model's errors should follow a **normal distribution** (bell-shaped curve).
- **Scatterplot:**
→ A graph that shows whether there's a straight-line relationship between two variables.

Tolerance measures how much a variable is explained by the other independent variables.

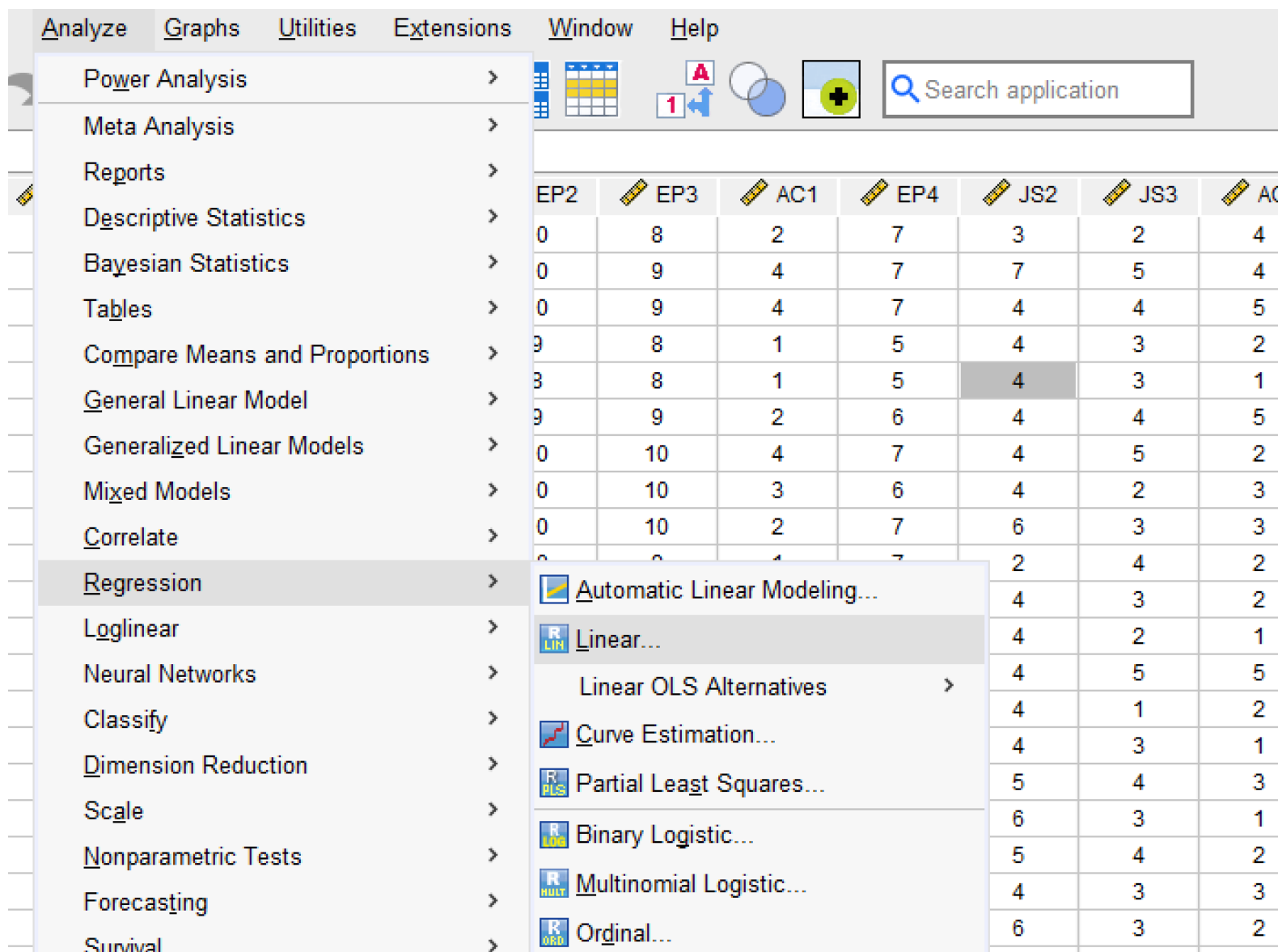
- **Low Tolerance (close to 0)** → High multicollinearity (problematic)
- **High Tolerance (close to 1)** → Low multicollinearity (good)
- **Standardized Residual** is the difference between the actual value and the predicted value of your dependent variable, **scaled by standard deviation**.

Key Terms

- **Constant:** The Constant is the intercept of the regression equation. It is the value of the dependent variable (Y) when all independent variables (X) are zero.
- What would Y be if all Xs = 0?, It is the starting point of your regression line.
- Example: If the Constant = 5.2, it means: When all your predictors (X values) are zero, your outcome (Y) is 5.2.
- Regression Equation Format: $Y = \text{Constant} + (B1 \times X1) + (B2 \times X2) + \dots$ The Constant is the first number in this formula.

Simple Regression

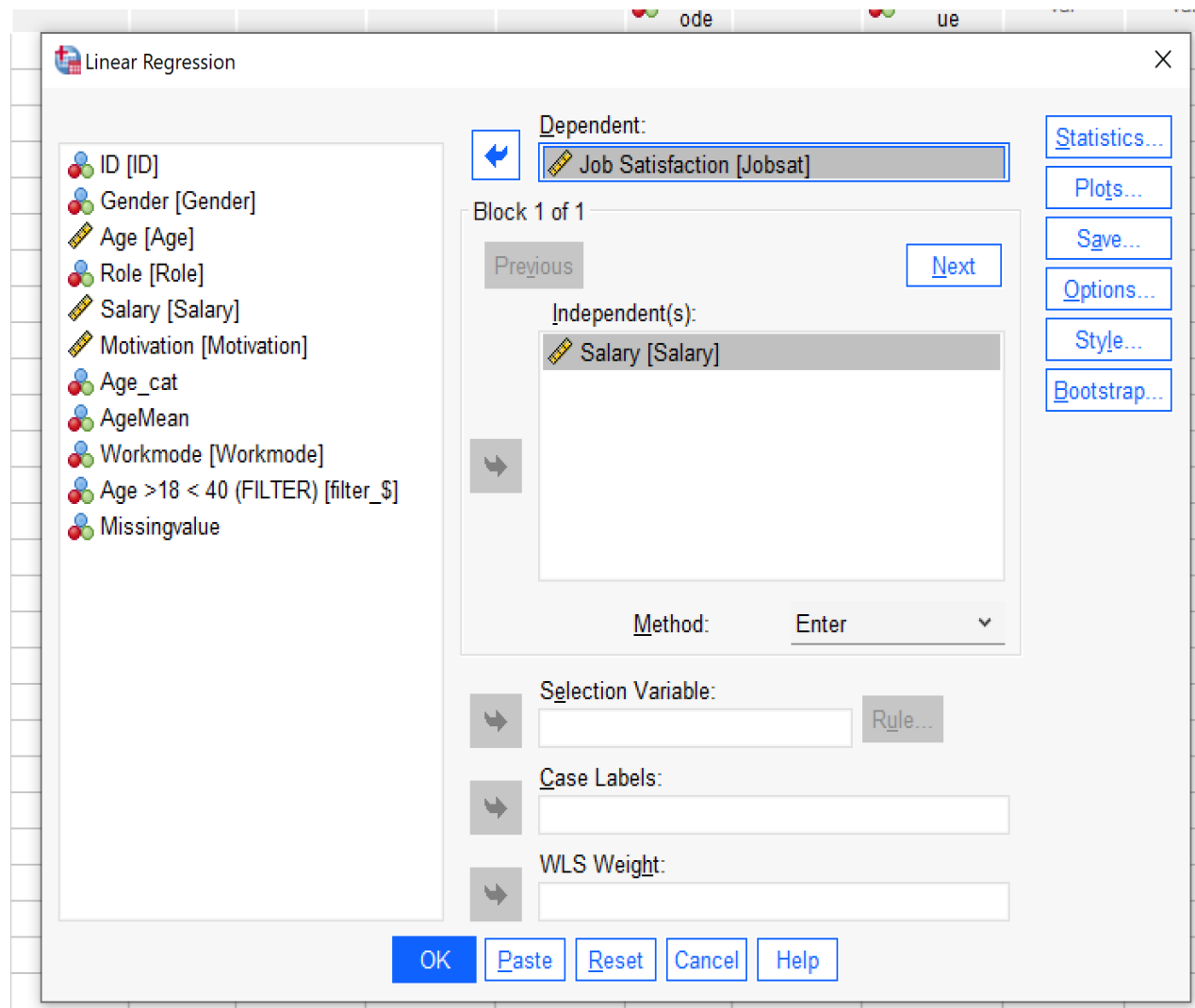
Go to Analyze
=>Regression =>
Linear



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



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

Linear Regression

Dependent: Job Satisfaction [Jobsat]

Block 1 of 1

Previous Next

Independent(s): Salary [Salary]

Method: Enter

Selection Variable: Rule...

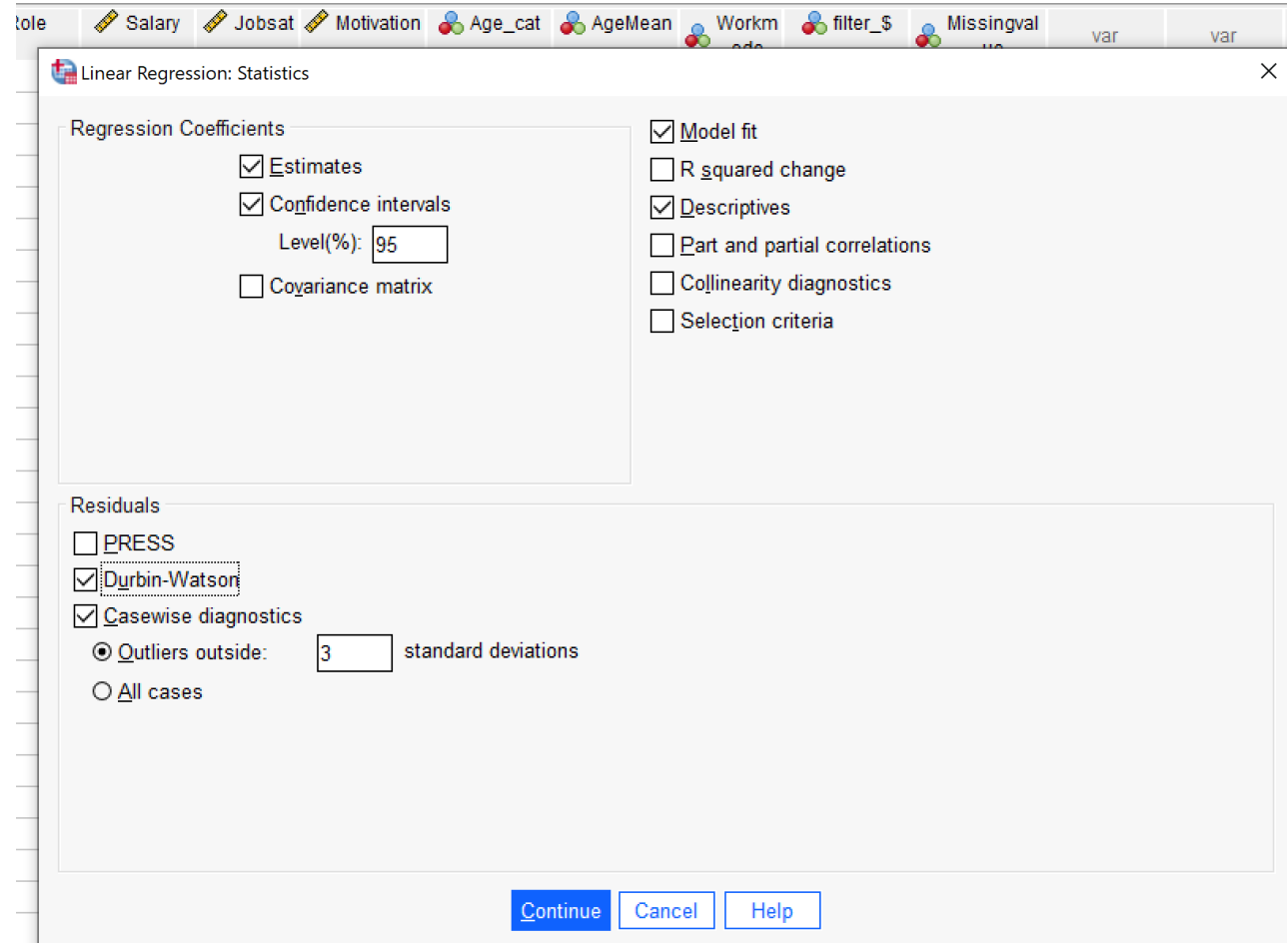
Case Labels:

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

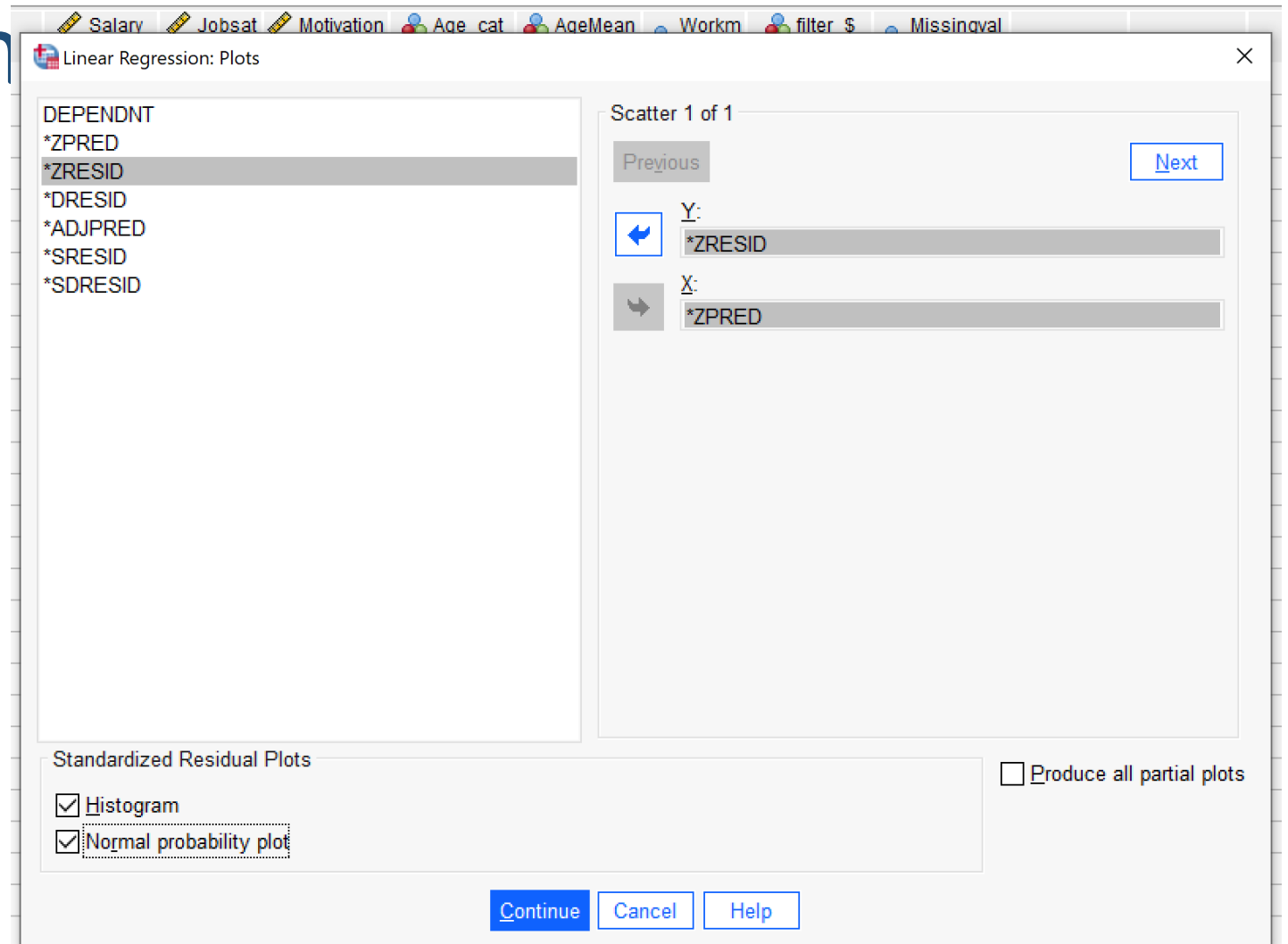
[Continue](#) [Cancel](#) [Help](#)

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

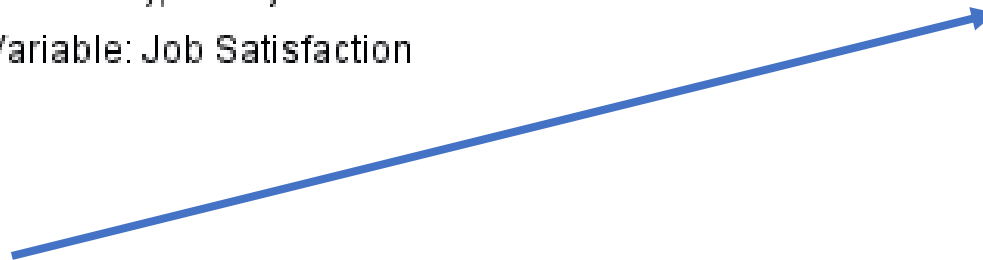
Model Summary^b

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

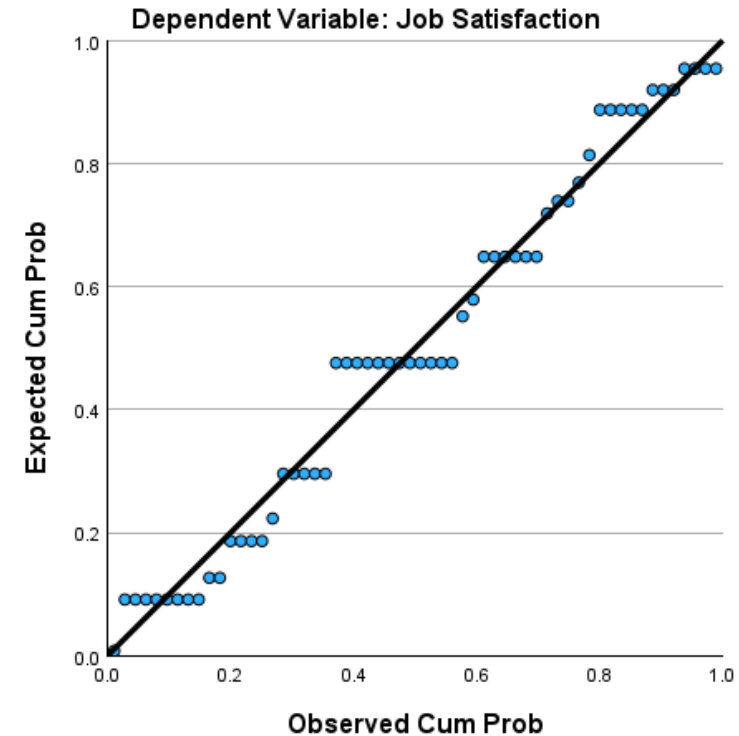
Assumption of independent
observations
 $1 < \text{DW value} < +3$



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	t	Sig.	95.0% Confidence Interval for B	
Model		B	Std. Error	Standardized Coefficients Beta			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 (B):**

- Tells us how much the outcome (dependent variable) changes when one predictor changes by **1 unit**, keeping other variables constant.

- It comes with a **p-value** to show if the result is **statistically significant**.

- **Standardised Beta (β):**

- Helps compare which **predictor has the strongest effect** on the outcome.

- Bigger values (positive or negative) = **stronger impact**.

- Also has a **p-value** for significance.

- **t-statistic:**

- Shows how far the result is from zero (no effect).

- A t-value above **+2** or below **-2** means the result is **usually meaningful**.

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!



Q&A

Thanks