

The background of the slide is a blurred image of financial market data. It features several overlapping elements: a blue semi-transparent rectangle on the left containing the title; a white line chart on the right showing price fluctuations; and a table of market data at the bottom. The table includes columns for market indices like S&P, NASDAQ, and Dow Jones, along with their respective changes and percentages. The overall aesthetic is professional and data-driven.

# MULTIPLE LINEAR REGRESSION

NURULJANNAH BT NOR AZMI

# INTRODUCTION

Multiple linear regression is used to estimate the relationship between **two or more independent variables** and **one dependent variable**.

Dependent (outcome) : numerical

Independent (predictor) : 2 or more numerical variables





# INTRODUCTION

If independent variables are combination of numerical and categorical or categorical only -  
**General Linear Regression**

Dependent (outcome) : numerical

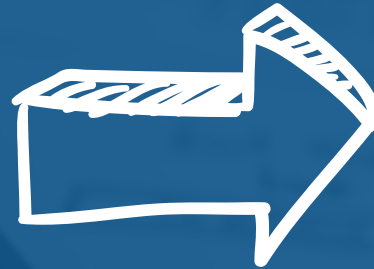
Independent (predictor) : 2 or more combination of numerical and categorical or categorical only



## SIMPLE LINEAR REGRESSION - ONLY ONE INDEPENDENT VARIABLE

Independent variable (x)

Mother's height



Dependent variable (y)

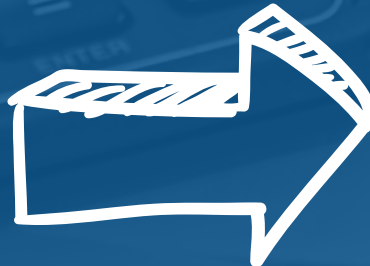
Length of baby

---

## MULTIPLE LINEAR REGRESSION - MORE THAN ONE INDEPENDENT VARIABLES

Independent variables (x)

Mother's height  
Mother's weight  
Age



Dependent variable (y)

Length of baby



# When to apply Multiple Linear Regression?

You can use multiple linear regression when you want to know:

1. How strong the relationship is between two or more independent variables and one dependent variable (e.g. how mother's height, weight and age affect length of baby).
2. The value of the dependent variable at a certain value of the independent variables (e.g. the expected length of baby at certain levels of mother's height, weight and age)

# Multiple Linear Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots \beta_n X_n$$

$Y$  = outcome

$\beta_0$  = intercept

$\beta_1 \dots \beta_n$  = regression coefficient for independent variable

$X_1 \dots X_n$  = independent variable



# STEPS IN MULTIPLE LINEAR REGRESSION

- 1 Descriptive statistics

---
- 2 Simple linear regression (Univariable analysis)

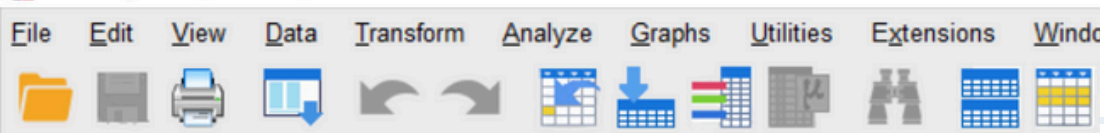
---
- 3 Multiple linear regression (Multivariable analysis)

---
- 4 Checking multicollinearity & interaction (Preliminary final model)

---
- 5 Checking assumptions (final model)

---
- 6 Interpretation & presentation

---



9 :

	ID	Headcirc	Length	Birthweight	Ges
1	1360	34	56	4.55	
2	1016	36	53	4.32	
3	462	39	58	4.10	
4	1187	38	53	4.07	
5	553	37	54	3.94	
6	1636	38	51	3.93	
7	820	34	52	3.77	
8	1191	33	53	3.65	
9	1081	38	54	3.63	
10	822	35	50	3.42	
11	1683	33	53	3.35	
12	1088	36	51	3.27	
13	1107	36	52	3.23	
14	755	33	53	3.20	
15	1058	34	53	3.15	
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27	431	30	48	1.92	

# EXAMPLE

Open dataset:  
birthweight.sav

This dataset contains information on new born babies and their parents admitted in Hospital Kuala Lumpur. A researcher is interested to determine the factors that are associated with the length of baby.



# EXAMPLE

RQ: What are the factors that associated with the length of baby?



Length of baby (DV)



Factors (IV)

- Mother's age
- Mother's height
- Mother's weight

List down all the variables



Numerical



Numerical

Identify the types of variables



Multiple Linear Regression

Identify the right statistical analysis

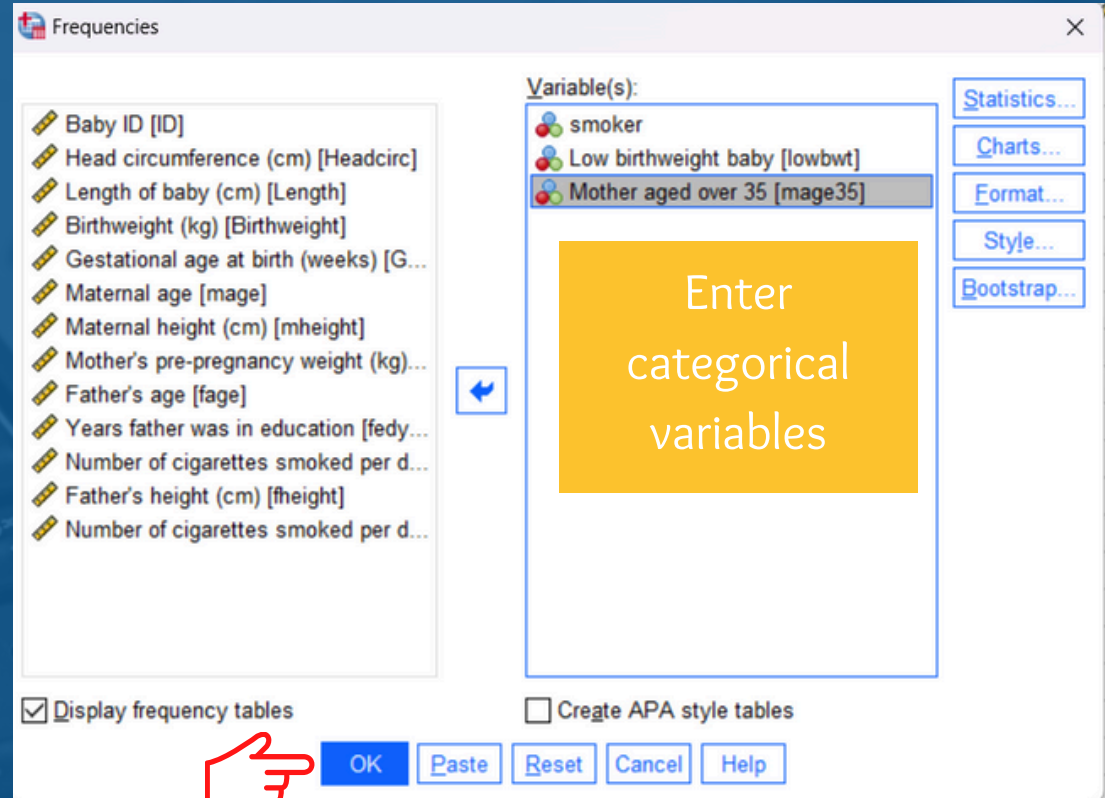
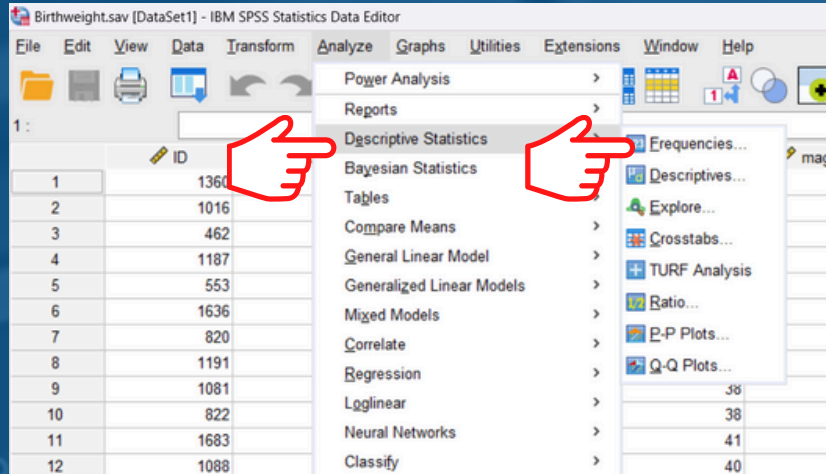
## STEP 1: DESCRIPTIVE STATISTICS

- 1.Data exploration and cleaning.
- 2.For categorical data, run the data by using Frequencies in SPSS.
- 3.For numerical data, run the data by using Descriptives/Explore in SPSS.



# Run frequencies for categorical data

Go to: Analyze > Descriptive statistics > Frequencies



### smoker

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Non-smoker	20	47.6	47.6	47.6
	Smoker	22	52.4	52.4	100.0
	Total	42	100.0	100.0	

### Low birthweight baby

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not low birthweight	36	85.7	85.7	85.7
	Low birthweight	6	14.3	14.3	100.0
	Total	42	100.0	100.0	

### Mother aged over 35

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Aged < 35	38	90.5	90.5	90.5
	Aged 35+	4	9.5	9.5	100.0
	Total	42	100.0	100.0	

# Run descriptive for numerical data

Go to: Analyze > Descriptive statistics > Descriptives/Explore

The screenshot shows the IBM SPSS Statistics Data Editor window with the 'Analyze' menu open. The 'Descriptive Statistics' option is highlighted, and the 'Descriptives...' option is selected. A red hand icon points to the 'Descriptives...' option. The 'Descriptives' dialog box is open, showing the list of variables on the left and the 'Variable(s):' list on the right. A red hand icon points to the 'OK' button. A yellow box with the text 'Enter numerical variables' is overlaid on the 'Variable(s):' list.

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

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

1: ID 1360 1016 462 1187 553 1636 820 1191 1081 822 1683 1088

Analyze > Descriptive Statistics > Descriptives...

Descriptives

Variable(s):

- Head circumference (cm) [Headcirc]
- Length of baby (cm) [Length]
- Birthweight (kg) [Birthweight]
- Gestational age at birth (weeks) [Gestation]
- Maternal age [mage]
- Maternal height (cm) [mheight]
- Mother's pre-pregnancy weight (kg) [mppwt]
- Father's age [fage]
- Years father was in education [fedyrs]
- Number of cigarettes smoked per day by father [fnocig]
- Father's height (cm) [fheight]
- Number of cigarettes smoked per day by mother [mnocig]

Enter numerical variables

Save standardized values as variables ☐

OK Paste Reset Cancel Help

## Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Head circumference (cm)	42	30	39	34.60	2.400
Length of baby (cm)	42	43	58	51.33	2.936
Birthweight (kg)	42	1.92	4.57	3.3129	.60390
Gestational age at birth (weeks)	42	33	45	39.19	2.643
Maternal age	42	20	43	28.71	6.711
Maternal height (cm)	42	149	181	164.45	6.504
Mother's pre-pregnancy weight (kg)	42	45	78	57.50	7.198
Father's age	42	19	46	28.90	6.864
Years father was in education	42	10	16	13.67	2.160
Number of cigarettes smoked per day by father	42	0	50	17.19	17.308
Father's height (cm)	42	169	200	180.50	6.978
Number of cigarettes smoked per day by mother	42	0	50	9.43	12.512
Valid N (listwise)	42				



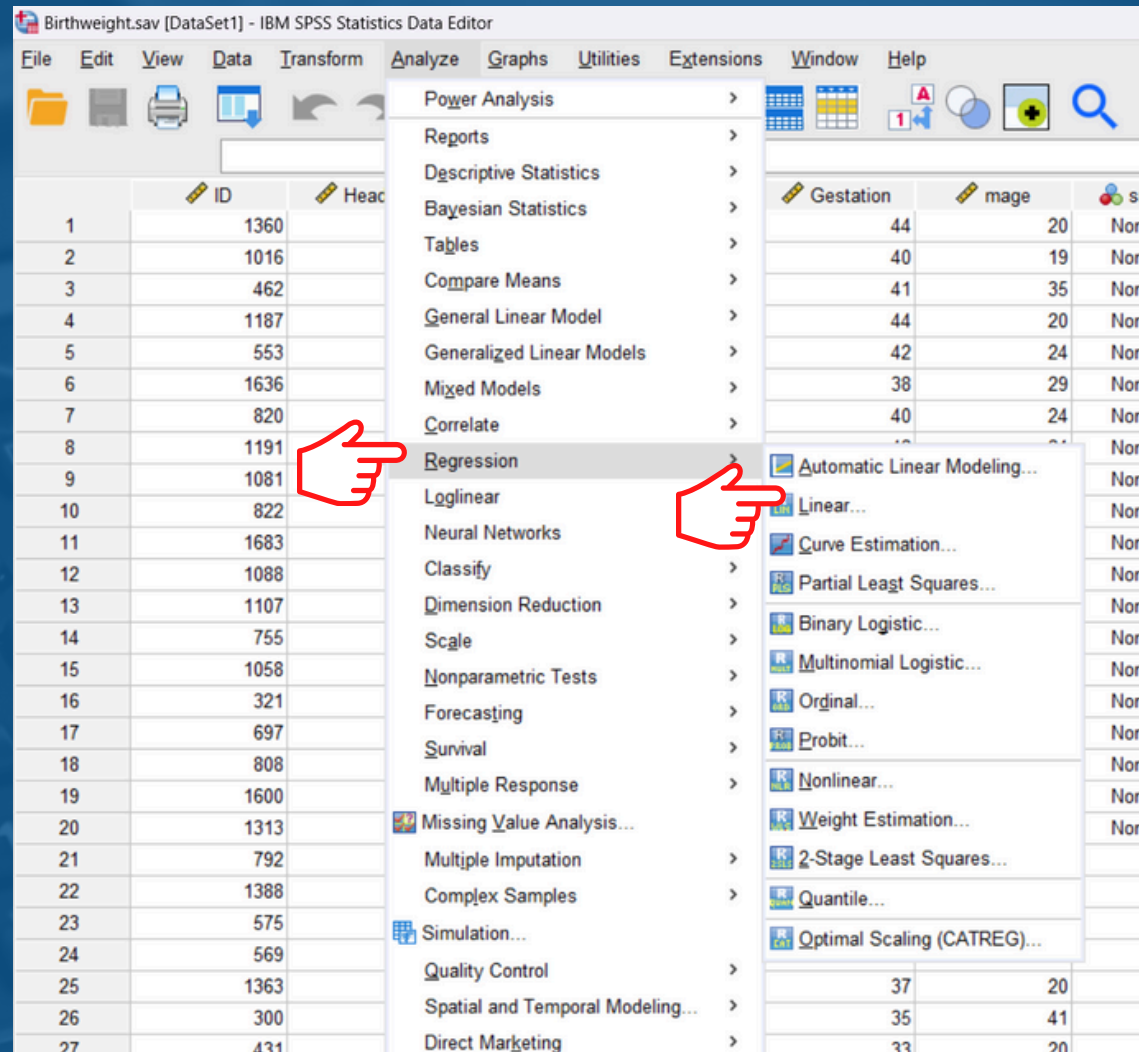
## STEP 2: SIMPLE LINEAR REGRESSION (UNIVARIABLE ANALYSIS)

1. Do Simple Linear Regression analysis for each independent variable:

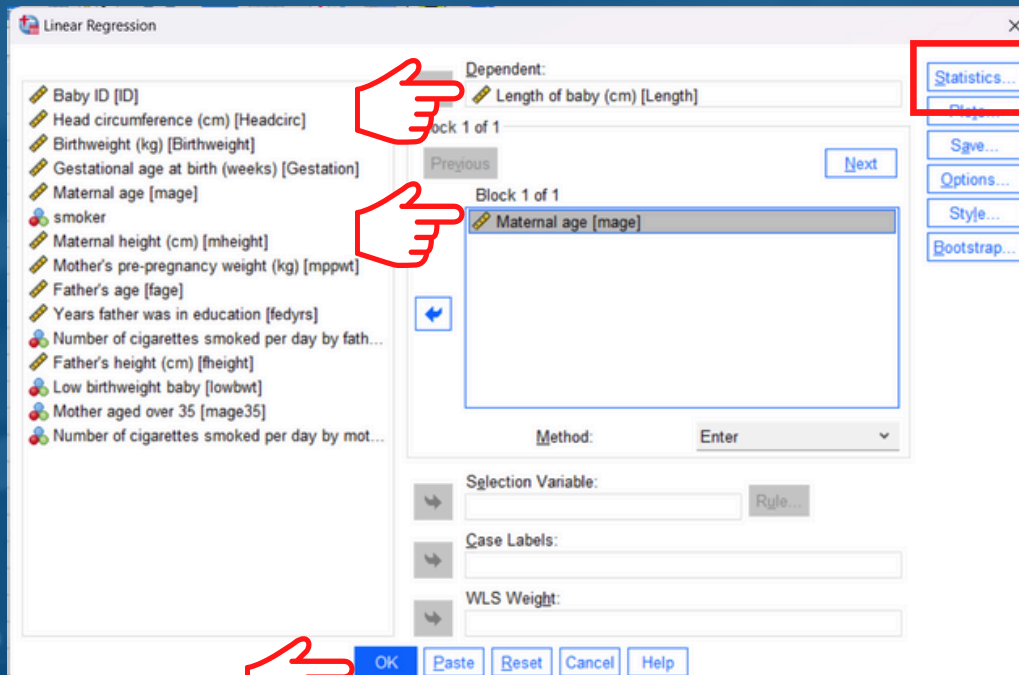
- Mother's age
- Mother's height
- Mother's weight

2. At the end, choose variables with  $p\text{-value} < 0.25$  and/or clinically important.

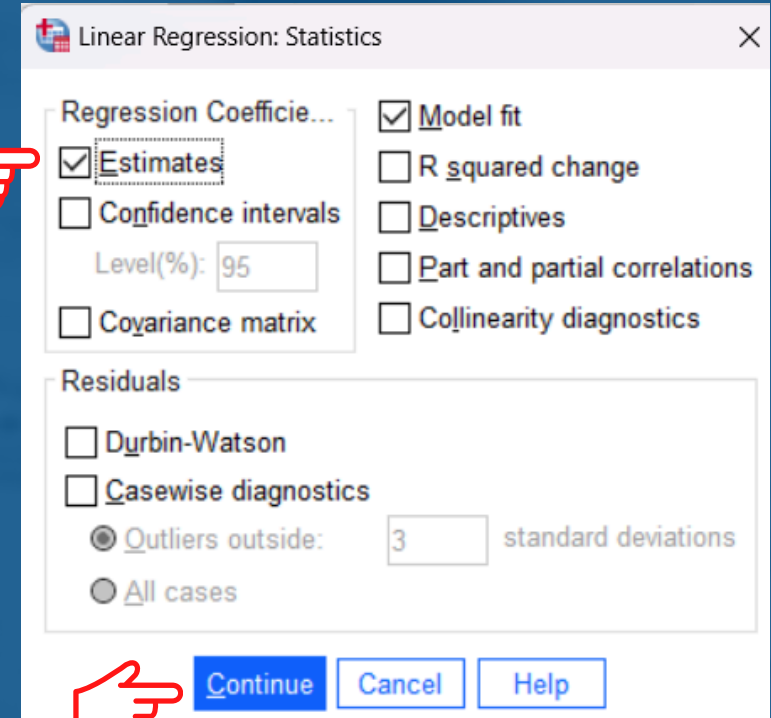
Go to: Analyze > Regression > Linear



# Length of baby vs Mother's age



The 'Linear Regression' dialog box in SPSS. The 'Dependent' variable is 'Length of baby (cm) [Length]'. The 'Block 1 of 1' list contains 'Maternal age [mage]'. The 'Method' is set to 'Enter'. Red arrows point to the 'Statistics...' button, the 'Maternal age [mage]' variable, and the 'OK' button.



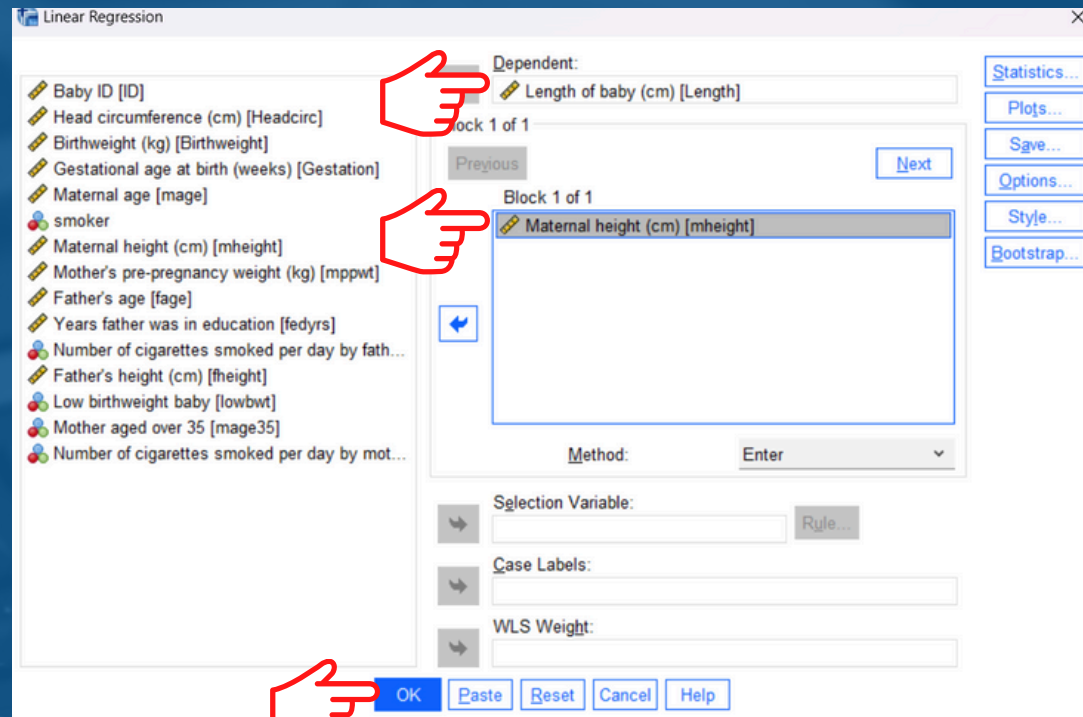
The 'Linear Regression: Statistics' sub-dialog box. Under 'Regression Coefficients', 'Estimates' is checked. Under 'Residuals', 'Durbin-Watson' and 'Casewise diagnostics' are unchecked. 'Outliers outside' is set to 3 standard deviations. Red arrows point to the 'Continue' button and the 'Estimates' checkbox.

Coefficients <sup>a</sup>							
		Unstandardized Coefficients		Standardized Coefficients		95.0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	
1	(Constant)	60.461	1.400		43.182	<.001	57.631 63.291
	Maternal age	-.318	.048	-.727	-6.691	<.001	-.414 -.222

a. Dependent Variable: Length of baby (cm)

There is a significant relationship between mother's age and the length of baby.

# Length of baby vs Mother's height



The image shows the SPSS Linear Regression dialog box. The dependent variable is 'Length of baby (cm) [Length]'. The independent variable is 'Maternal height (cm) [mheight]'. The method is set to 'Enter'. The 'OK' button is highlighted with a red hand icon.

Dependent: Length of baby (cm) [Length]

Block 1 of 1

Maternal height (cm) [mheight]

Method: Enter

OK Paste Reset Cancel Help

Coefficients <sup>a</sup>								
		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	15.334	10.271		1.493	.143	-5.425	36.093
	Maternal height (cm)	.219	.062	.485	3.507	.001	.093	.345

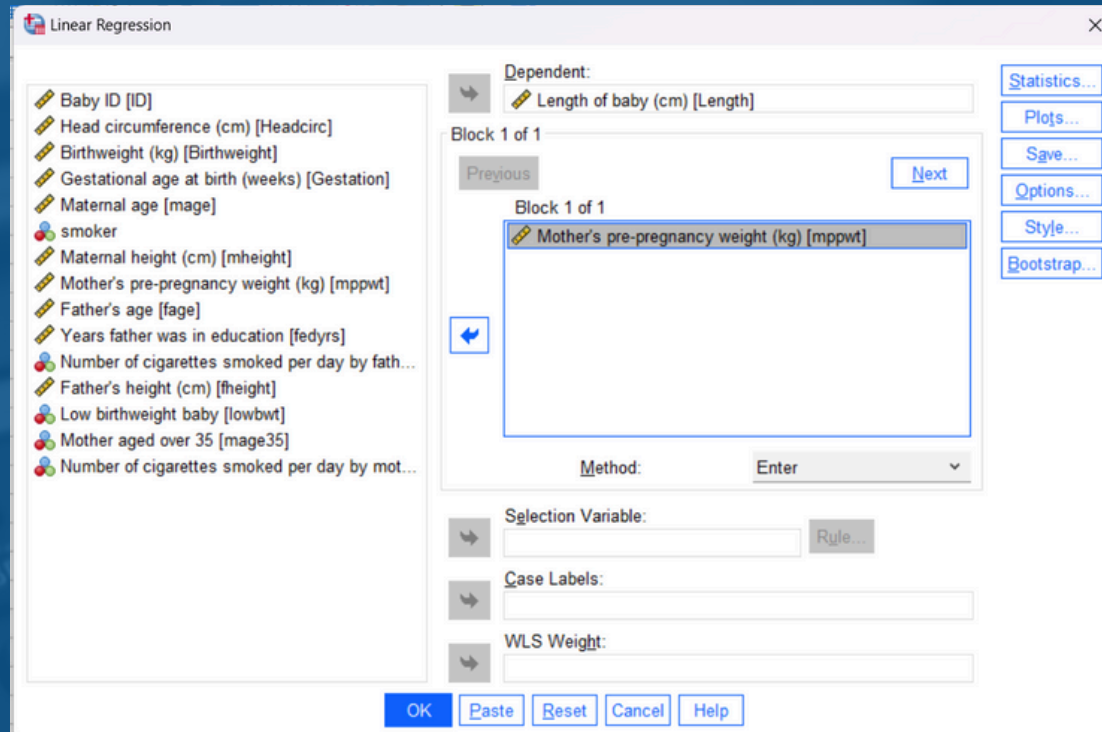
a. Dependent Variable: Length of baby (cm)



There is a significant relationship between mother's height and the length of baby.



# Length of baby vs Mother's weight



The image shows the 'Linear Regression' dialog box in SPSS. The 'Dependent' variable is 'Length of baby (cm) [Length]'. The 'Block 1 of 1' list contains 'Mother's pre-pregnancy weight (kg) [mppwt]'. The 'Method' is set to 'Enter'. The left list of variables includes 'Baby ID [ID]', 'Head circumference (cm) [Headcirc]', 'Birthweight (kg) [Birthweight]', 'Gestational age at birth (weeks) [Gestation]', 'Maternal age [mage]', 'smoker', 'Maternal height (cm) [mheight]', 'Father's age [fage]', 'Years father was in education [fedys]', 'Number of cigarettes smoked per day by fath...', 'Father's height (cm) [fheight]', 'Low birthweight baby [lowbwt]', 'Mother aged over 35 [mage35]', and 'Number of cigarettes smoked per day by mot...'. The bottom buttons are 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. On the right, there are buttons for 'Statistics...', 'Plots...', 'Save...', 'Options...', 'Style...', and 'Bootstrap...'.

Coefficients <sup>a</sup>							
Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B
		B	Std. Error	Beta	t	Sig.	Lower Bound
1	(Constant)	41.996	3.427		12.255	< .001	35.070
	Mother's pre-pregnancy weight (kg)	.162	.059	.398	2.745	.009	.043

a. Dependent Variable: Length of baby (cm)



There is a significant relationship between mother's pre-pregnancy weight and the length of baby.

Table 1: Associated factors of the length of baby by Simple Linear Regression

Variable	Simple Linear Regression	
	b* (95% CI)	p-value
Mother's age	-0.32 (-0.41,-0.22)	<0.001
Mother's height (cm)	0.22 (0.09,0.35)	0.001
Mother's pre-pregnancy weight (kg)	0.16 (0.04,0.28)	0.009

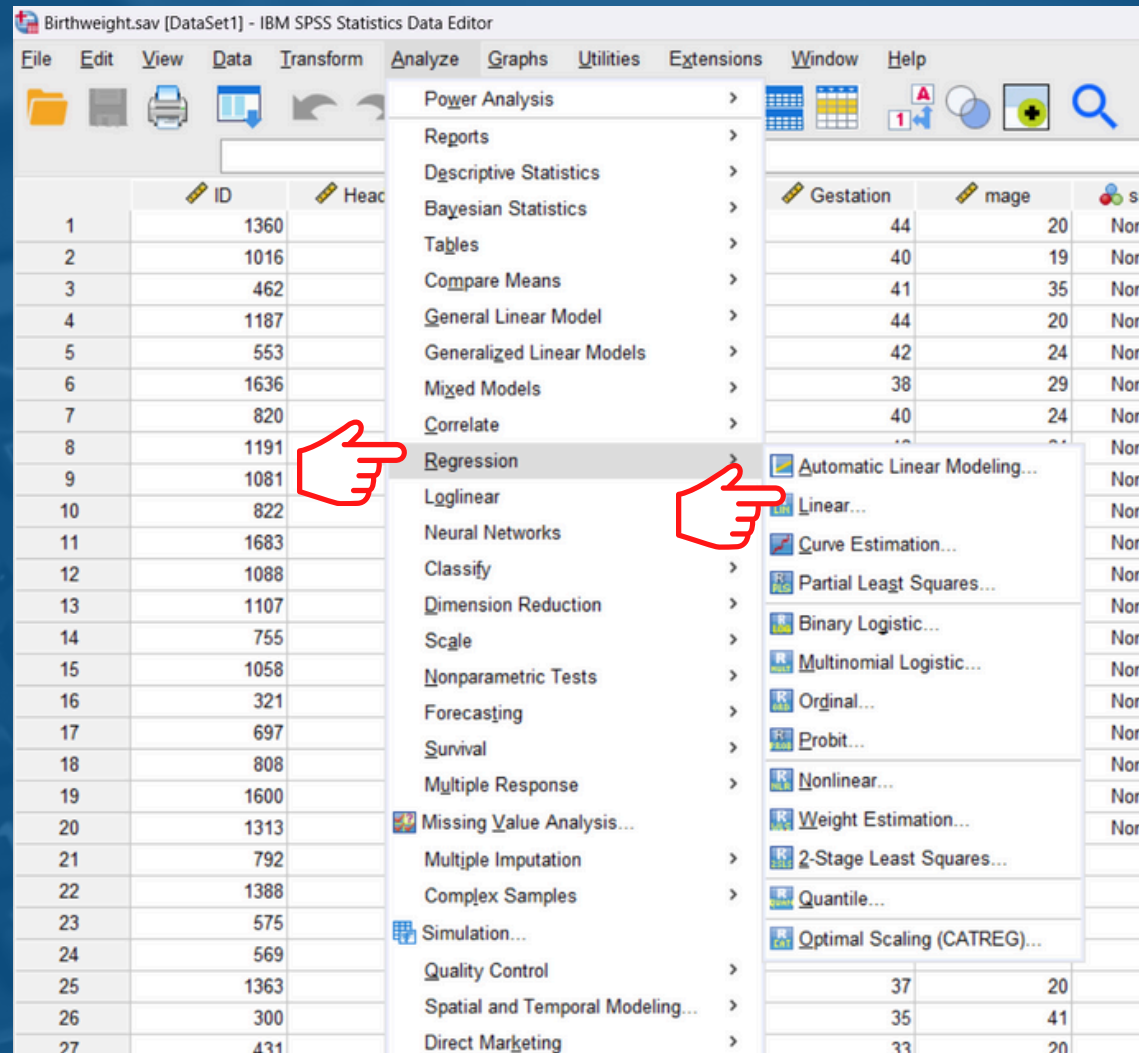
## STEP 3: MULTIPLE LINEAR REGRESSION (MULTIVARIABLE ANALYSIS)

1. Variables selection can be done by using following methods:

- Forward
- Backward
- Stepwise

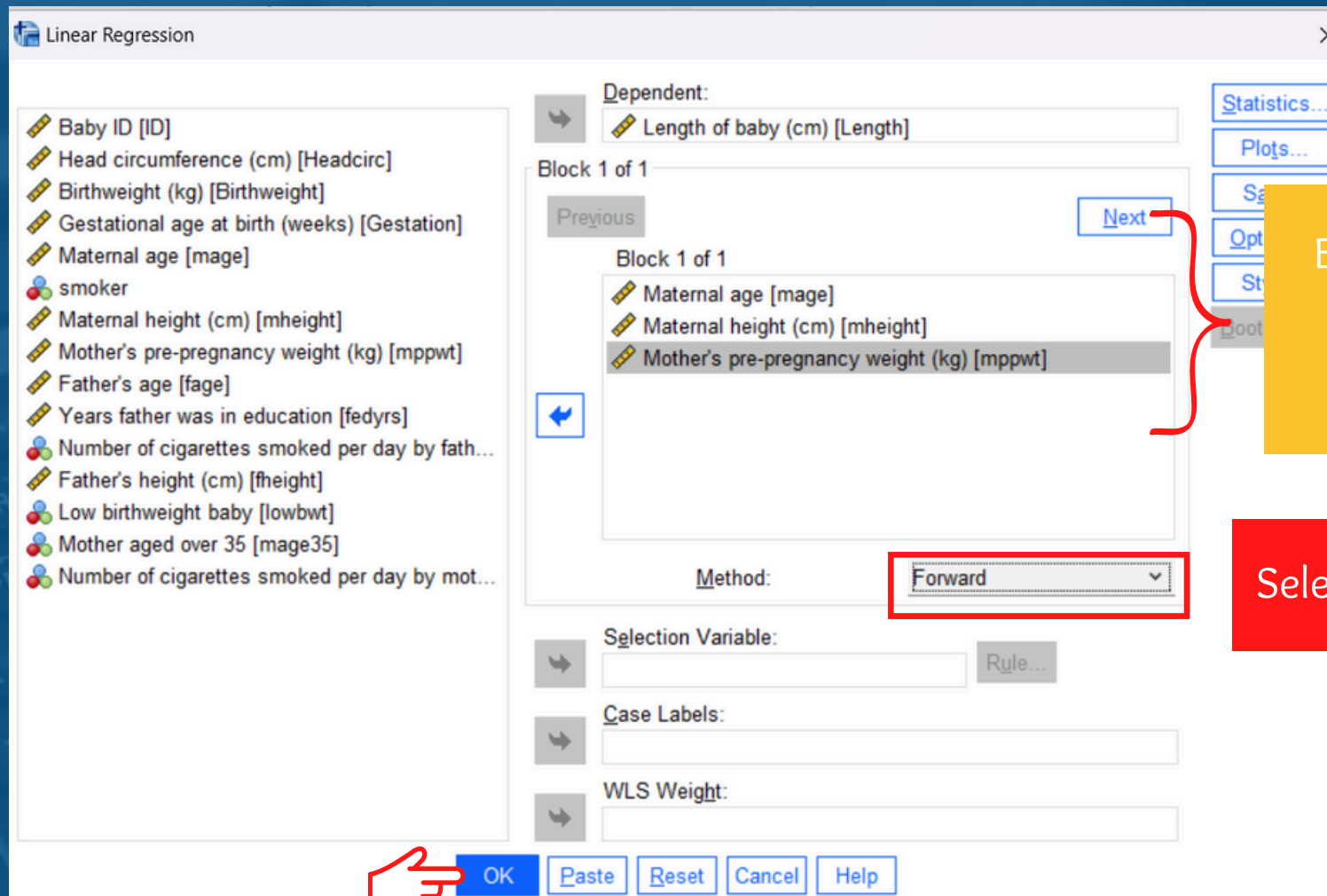
2. Perform all the methods and select the model with all variables significant as the preliminary main effect model.

# Go to: Analyze > Regression > Linear





METHOD: FORWARD (Automatically enters the IMPORTANT independent variable into the model)



The image shows the 'Linear Regression' dialog box in SPSS. On the left is a list of variables: Baby ID [ID], Head circumference (cm) [Headcirc], Birthweight (kg) [Birthweight], Gestational age at birth (weeks) [Gestation], Maternal age [mage], smoker, Maternal height (cm) [mheight], Mother's pre-pregnancy weight (kg) [mppwt], Father's age [fage], Years father was in education [fedysr], Number of cigarettes smoked per day by fath..., Father's height (cm) [fheight], Low birthweight baby [lowbwt], Mother aged over 35 [mage35], and Number of cigarettes smoked per day by mot... The 'Dependent' field contains 'Length of baby (cm) [Length]'. Under 'Block 1 of 1', the independent variables 'Maternal age [mage]', 'Maternal height (cm) [mheight]', and 'Mother's pre-pregnancy weight (kg) [mppwt]' are listed. The 'Method' dropdown is set to 'Forward'. The 'OK' button is highlighted with a red hand icon. On the right side of the dialog, there are buttons for 'Statistics...', 'Plots...', 'Save...', 'Options...', and 'Save As...', with a red bracket pointing to them from a yellow callout box.

Linear Regression

Dependent: Length of baby (cm) [Length]

Block 1 of 1

Previous Next

Block 1 of 1

Maternal age [mage]  
Maternal height (cm) [mheight]  
Mother's pre-pregnancy weight (kg) [mppwt]

Method: Forward

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Enter all the  
selected  
variables

Select Forward

## METHOD: FORWARD

Excluded Variables <sup>a</sup>						
Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	Maternal height (cm)	.316 <sup>b</sup>	3.094	.004	.444	.931
	Mother's pre-pregnancy weight (kg)	.267 <sup>b</sup>	2.576	.014	.381	.962
2	Mother's pre-pregnancy weight (kg)	.108 <sup>c</sup>	.799	.429	.129	.536

a. Dependent Variable: Length of baby (cm)

b. Predictors in the Model: (Constant), Maternal age

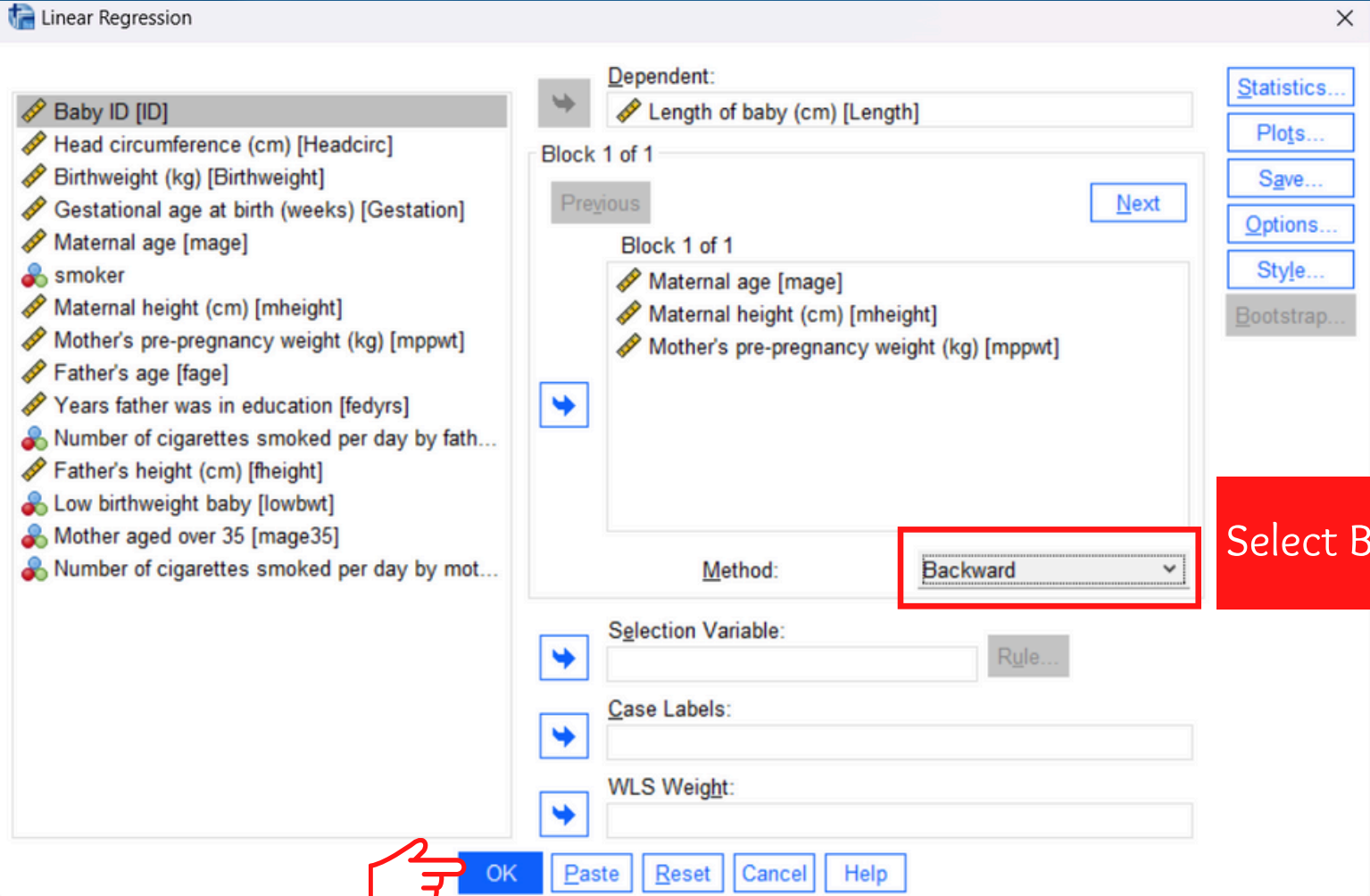
c. Predictors in the Model: (Constant), Maternal age, Maternal height (cm)

Coefficients <sup>a</sup>								
Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	60.461	1.400		43.182	<.001	57.631	63.291
	Maternal age	-.318	.048	-.727	-6.691	<.001	-.414	-.222
2	(Constant)	35.959	8.019		4.484	<.001	19.738	52.180
	Maternal age	-.282	.045	-.644	-6.303	<.001	-.372	-.191
	Maternal height (cm)	.143	.046	.316	3.094	.004	.049	.236

a. Dependent Variable: Length of baby (cm)

Mother's age and height are significant.

METHOD: BACKWARD (Automatically removes the UNIMPORTANT independent variable out of the model)



The image shows the 'Linear Regression' dialog box in SPSS. The 'Dependent' variable is 'Length of baby (cm) [Length]'. The 'Block 1 of 1' list contains 'Maternal age [mage]', 'Maternal height (cm) [mheight]', and 'Mother's pre-pregnancy weight (kg) [mppwt]'. The 'Method' dropdown is set to 'Backward'. A red box highlights the 'Backward' option, and a red arrow points to the 'OK' button. The 'Statistics...', 'Plots...', 'Save...', 'Options...', 'Style...', and 'Bootstrap...' buttons are visible on the right.

Linear Regression

Dependent: Length of baby (cm) [Length]

Block 1 of 1

Previous Next

Block 1 of 1

Maternal age [mage]  
Maternal height (cm) [mheight]  
Mother's pre-pregnancy weight (kg) [mppwt]

Method: Backward

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics...  
Plots...  
Save...  
Options...  
Style...  
Bootstrap...

68 522  
91 948  
16 545  
12 372  
9 491  
nthly  
do the

Select Backward

## METHOD: BACKWARD

Excluded Variables <sup>a</sup>						
Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
2	Mother's pre-pregnancy weight (kg)	.108 <sup>b</sup>	.799	.429	.129	.536

a. Dependent Variable: Length of baby (cm)

b. Predictors in the Model: (Constant), Maternal age, Maternal height (cm)

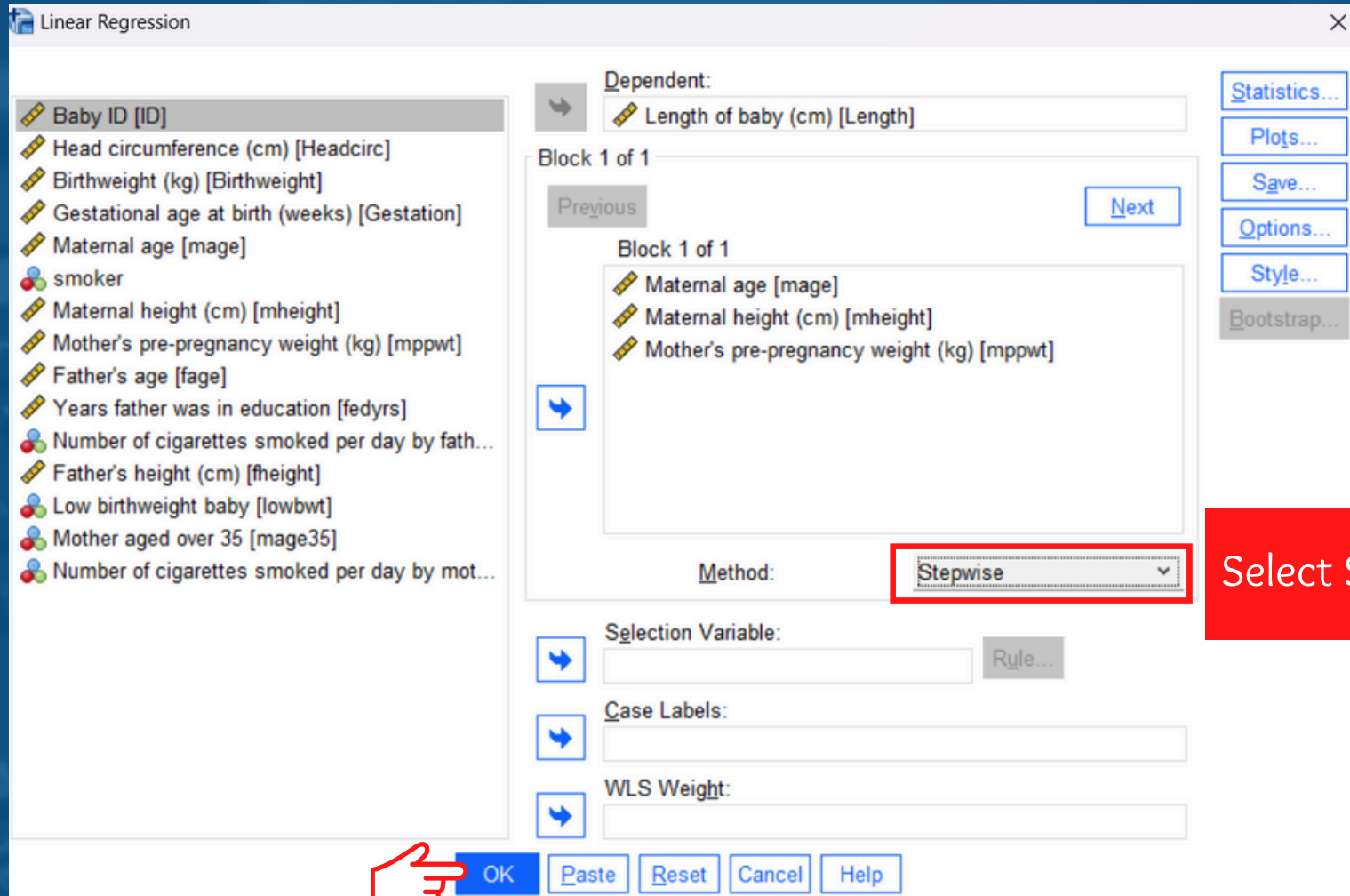
Coefficients <sup>a</sup>								
Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error				Lower Bound	Upper Bound
1	(Constant)	38.823	8.819		4.402	<.001	20.970	56.677
	Maternal age	-.281	.045	-.642	-6.254	<.001	-.372	-.190
	Maternal height (cm)	.110	.062	.243	1.769	.085	-.016	.235
	Mother's pre-pregnancy weight (kg)	.044	.055	.108	.799	.429	-.068	.156
2	(Constant)	35.959	8.019		4.484	<.001	19.738	52.180
	Maternal age	-.282	.045	-.644	-6.303	<.001	-.372	-.191
	Maternal height (cm)	.143	.046	.316	3.094	.004	.049	.236

a. Dependent Variable: Length of baby (cm)

Mother's age and height are significant.



METHOD: STEPWISE (The procedure adds or removes independent variables one at a time using the variable's statistical significance)



The image shows the 'Linear Regression' dialog box in SPSS. The 'Dependent' variable is 'Length of baby (cm) [Length]'. The 'Block 1 of 1' list contains 'Maternal age [mage]', 'Maternal height (cm) [mheight]', and 'Mother's pre-pregnancy weight (kg) [mppwt]'. The 'Method' dropdown is set to 'Stepwise' and is highlighted with a red box. A red arrow points to the 'OK' button at the bottom left. A red box on the right contains the text 'Select Stepwise'.

Linear Regression

Dependent:  
Length of baby (cm) [Length]

Block 1 of 1  
Previous Next

Block 1 of 1  
Maternal age [mage]  
Maternal height (cm) [mheight]  
Mother's pre-pregnancy weight (kg) [mppwt]

Method: Stepwise

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Select Stepwise

## METHOD: STEPWISE

Excluded Variables <sup>a</sup>						
Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	Maternal height (cm)	.316 <sup>b</sup>	3.094	.004	.444	.931
	Mother's pre-pregnancy weight (kg)	.267 <sup>b</sup>	2.576	.014	.381	.962
2	Mother's pre-pregnancy weight (kg)	.108 <sup>c</sup>	.799	.429	.129	.536

a. Dependent Variable: Length of baby (cm)

b. Predictors in the Model: (Constant), Maternal age

c. Predictors in the Model: (Constant), Maternal age, Maternal height (cm)

Coefficients <sup>a</sup>								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	60.461	1.400		43.182	<.001	57.631	63.291
	Maternal age	-.318	.048	-.727	-6.691	<.001	-.414	-.222
2	(Constant)	35.959	8.019		4.484	<.001	19.738	52.180
	Maternal age	-.282	.045	-.644	-6.303	<.001	-.372	-.191
	Maternal height (cm)	.143	.046	.316	3.094	.004	.049	.236

a. Dependent Variable: Length of baby (cm)

Mother's age and height are significant.

- During this step, mother's age and height were found to be significant in all methods.
- Run the model once again using 'Enter' method by using the chosen variables.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	35.959	8.019		4.484	<.001	19.738	52.180
	Maternal age	-.282	.045	-.644	-6.303	<.001	-.372	-.191
	Maternal height (cm)	.143	.046	.316	3.094	.004	.049	.236

a. Dependent Variable: Length of baby (cm)

- This will be the preliminary main effect model.

## STEP 4: CHECKING MULTICOLLINEARITY

1. Multicollinearity occurs when independent variables in a regression model are correlated.
2. This correlation is a problem because independent variables should be independent.
3. If the degree of correlation between variables is high enough, it can cause problems when you fit the model and interpret the results.
4. There is a high chance of getting inaccurate p-values and wide confidence interval of regression coefficient.



## STEP 4: CHECKING MULTICOLLINEARITY

5. Multicollinearity can be checked by using Variance Inflation Factor (VIF).

6. If VIF is more than 10, then there is a multicollinearity amongst independent variables.

# Go to: Analyze > Regression > Linear

Linear Regression

Dependent: Length of baby (cm) [Length]

Block 1 of 1

Maternal age [mage]  
Maternal height (cm) [mheight]

Method: Enter

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options... Style... Bootstrap...

Linear Regression: Statistics

Regression Coefficients

☒ Estimates ☒ Model fit

☒ Confidence intervals ☐ R squared change

Level(%): 95 ☐ Descriptives

☐ Covariance matrix ☐ Part and partial correlations

☒ Collinearity diagnostics

Residuals

☐ Durbin-Watson

☐ Casewise diagnostics

☒ Outliers outside: 3 standard deviations

☐ All cases

Continue Cancel Help

Coefficients <sup>a</sup>										
		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	35.959	8.019		4.484	<.001	19.738	52.180		
	Maternal age	-.282	.045	-.644	-6.303	<.001	-.372	-.191	.931	1.074
	Maternal height (cm)	.143	.046	.316	3.094	.004	.049	.236	.931	1.074

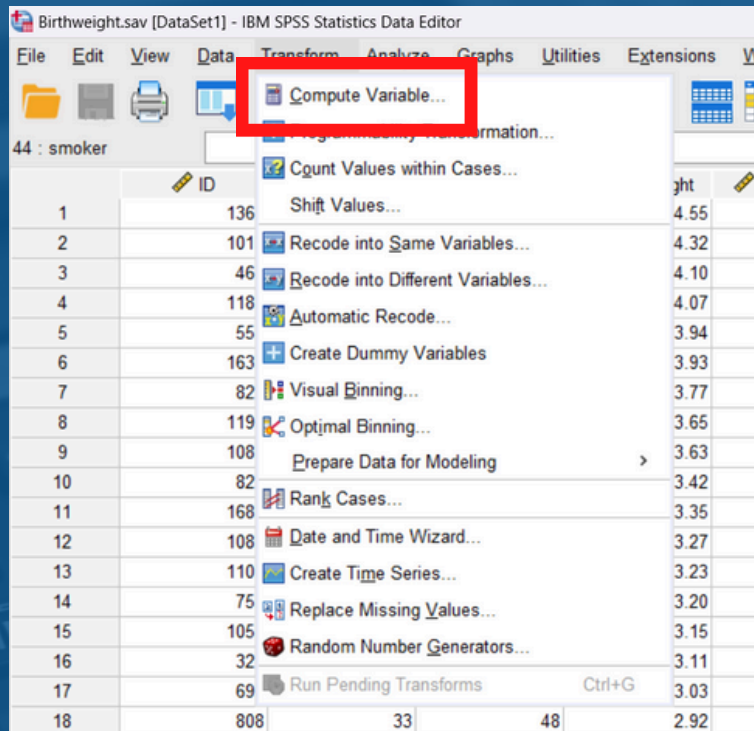
a. Dependent Variable: Length of baby (cm)

The values of VIF for both variables are less than 10. There is no multicollinearity problem in this model.

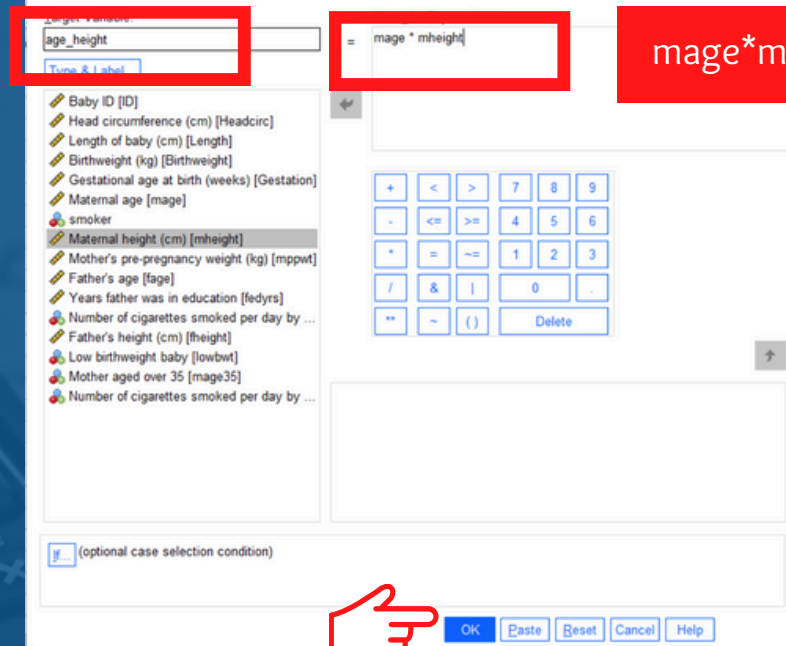
## STEP 4: CHECKING INTERACTION

1. An interaction effect occurs when the effect of one variable depends on the value of another variable.
2. The interaction terms need to be biologically meaningful.
3. The interaction term needs to be computed in SPSS and then added to the model as an independent variable. If you have more than one interaction term, add to the model one by one.
4. If the interaction term is statistically significant, include the term in the model.

# Go to: Transform > Compute variable



age\_height



mage\*mheight

fnocig	fheight	lowbwt	mage35	mnocig	age_height
35	179	Not low birthwei...	Aged < 35		4050.00
0	183	Not low birthwei...	Aged < 35		4446.00
25	185	Not low birthwei...	Aged 35+		3440.00
25	189	Not low birthwei...	Aged < 35		3828.00
0	184	Not low birthwei...	Aged < 35		4550.00
0	180	Not low birthwei...	Aged < 35		4455.00
0	173	Not low birthwei...	Aged < 35		4396.00
25	185	Not low birthwei...	Aged < 35		3960.00
7	172	Not low birthwei...	Aged < 35		4988.00
0	179	Not low birthwei...	Aged < 35		3611.00
0	170	Not low birthwei...	Aged < 35		4428.00
0	181	Not low birthwei...	Aged < 35		3528.00
0	183	Not low birthwei...	Aged < 35		4592.00
25	183	Not low birthwei...	Aged < 35		4485.00



# Go to: Analyze > Regression > Linear

Linear Regression

Dependent: Length of baby (cm) [Length]

Block 1 of 1

Previous Next

Block 1 of 1

Maternal age [mage]  
Maternal height (cm) [mheight]  
age\_height

Method: Enter

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics...  
Plots...  
Save...  
Options...  
Style...

add interaction term





Linear Regression dialog box showing the list of variables on the left and the dependent variable, Block 1 of 1, and the interaction term age\_height selected. A red arrow points to the 'add interaction term' button. Another red arrow points to the 'OK' button.

Coefficients <sup>a</sup>								
		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	32.127	34.224		.939	.354	-37.156	101.411
	Maternal age	-.153	1.121	-.349	-.136	.892	-2.422	2.116
	Maternal height (cm)	.166	.209	.368	.795	.432	-.257	.589
	age_height	-.001	.007	-.286	-.115	.909	-.015	.013

a. Dependent Variable: Length of baby (cm)

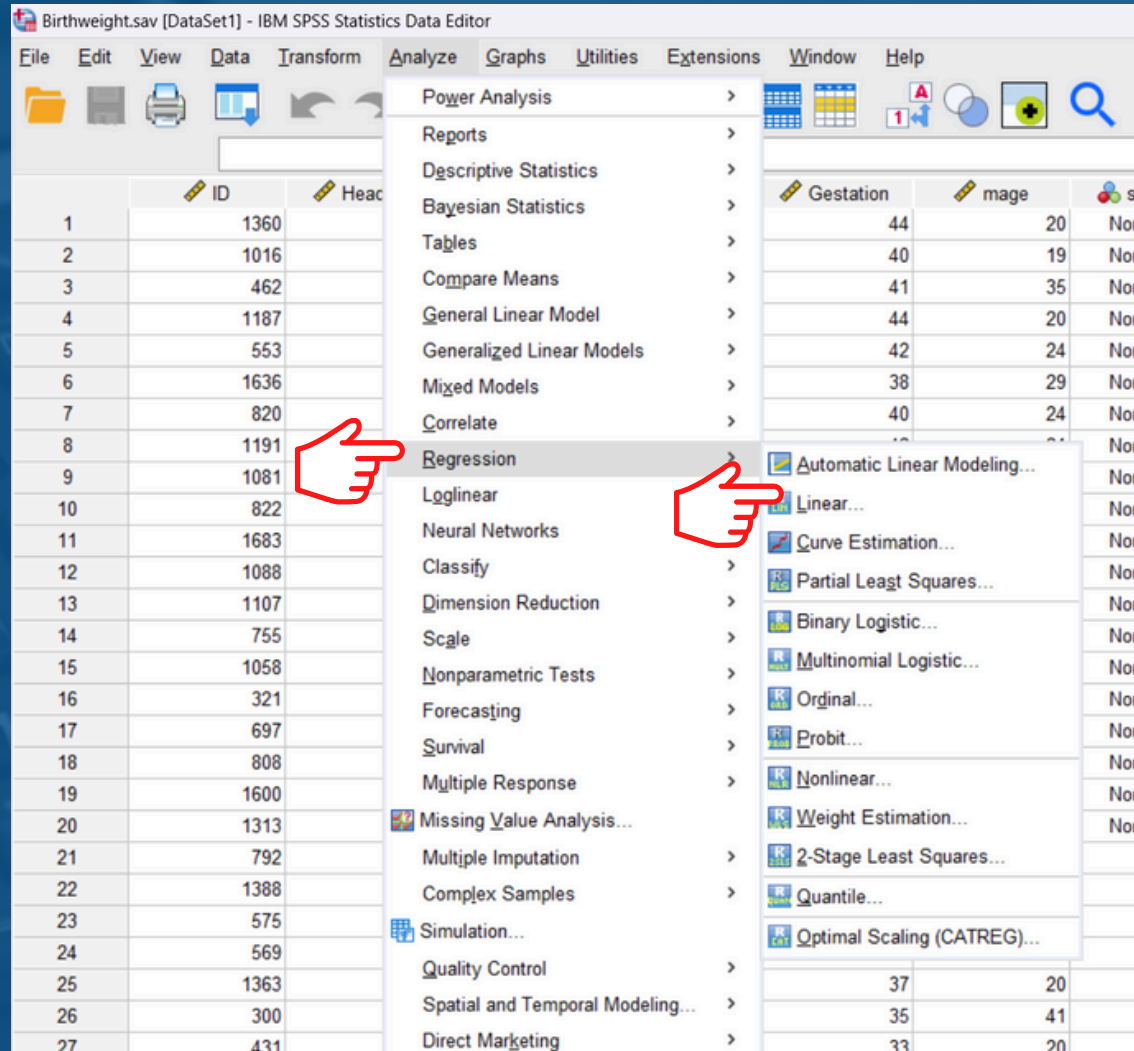
The interaction age\_height is not statistically significant ( $p=0.909$ ). Thus, remove the interaction term from the analysis.

## STEP 5: CHECKING ASSUMPTIONS

Assumptions	How to check?
1. <u>Independent</u> observation	Done during design stage
2.Overall <u>linearity</u>	Scatter plot between residuals and predicted values (XP - YR) 
3.Homoscedasticity ( <u>Equal</u> variances)	Scatter plot between residuals and predicted values (XP - YR) 
4. <u>Linearity</u> of each independent variable	Scatter plot residual vs each independent variable (XI - YR) 
5.Residuals should be approximately <u>normally</u> distributed	Histogram with overlaid normal curve of residuals 

# Checking assumption: Overall linearity & Homoscedasticity

Go to: Analyze > Regression > Linear





Linear Regression

Dependent: Length of baby (cm) [Length]

Block 1 of 1

Maternal age [mage]  
Maternal height (cm) [mheight]

Method: Enter

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics...  
Plot...  
Save...  
Options...  
Style...  
Bootstrap...

Linear Regression: Save

Predicted Values  
☒ Unstandardized  
☐ Standardized  
☐ Adjusted  
☐ S.E. of mean predictions

Residuals  
☒ Unstandardized  
☐ Standardized  
☐ Studentized  
☐ Deleted  
☐ Studentized deleted

Distances  
☐ Mahalanobis  
☐ Cook's  
☐ Leverage values

Influence Statistics  
☐ DfBetas  
☐ Standardized DfBetas  
☐ DfFits  
☐ Standardized DfFits  
☐ Covariance ratios

Prediction Intervals  
☐ Mean ☐ Individual  
Confidence Interval: 95 %

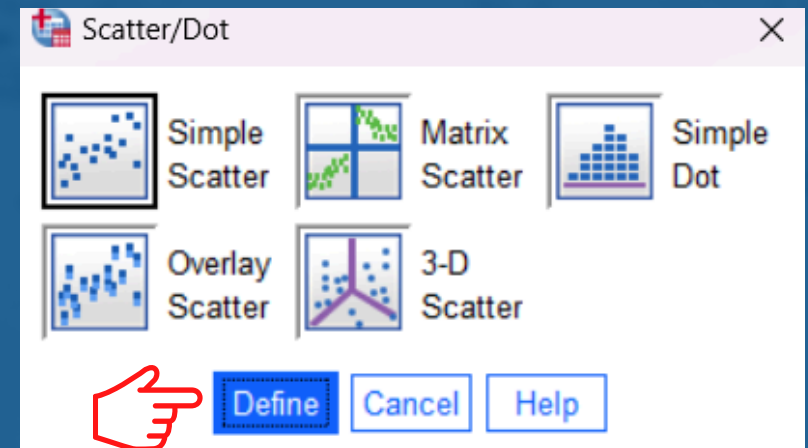
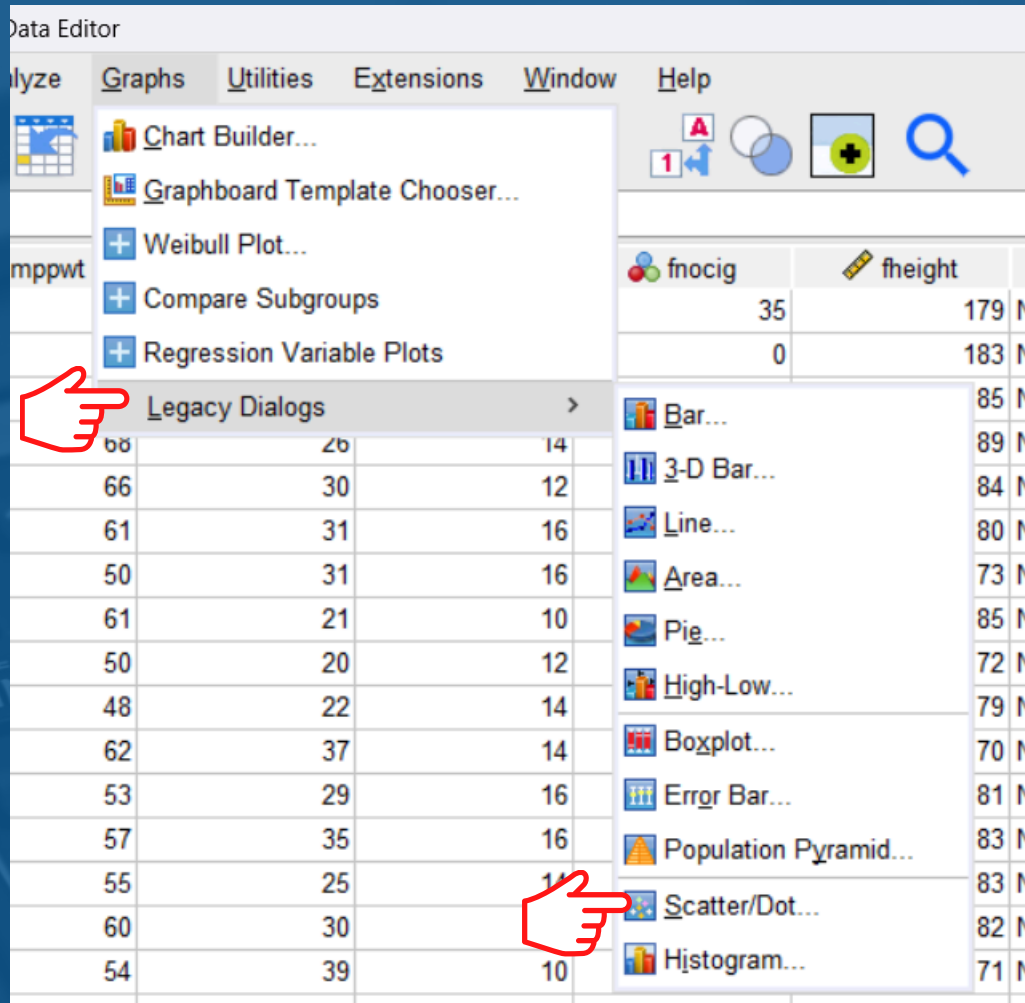
Coefficient statistics  
☐ Create coefficient statistics  
☒ Create a new dataset  
Dataset name:   
☐ Write a new data file  
File:

Export model information to XML file  
 Browse...  
☒ Include the covariance matrix

Continue Cancel Help

wt	mage35	mnocig	PRE_1	RES_1
hwei...	Aged < 35	0	52.02945	3.97055
hwei...	Aged < 35	0	53.03179	-.03179
hwei...	Aged 35+	0	54.86410	3.13590
hwei...	Aged < 35	0	54.58620	-1.58620
hwei...	Aged < 35	0	53.60243	.39757
hwei...	Aged < 35	0	51.89421	-.89421
hwei...	Aged < 35	0	50.47132	1.52868
hwei...	Aged < 35	0	52.73904	.26096
hwei...	Aged < 35	0	52.32962	1.67038
hwei...	Aged < 35	0	51.87936	-1.87936
hwei...	Aged < 35	0	51.75155	1.24845
hwei...	Aged < 35	0	54.01185	-3.01185
hwei...	Aged < 35	0	51.46995	.53005
hwei...	Aged < 35	0	50.46761	2.53239
hwei...	Aged < 35	0	53.58758	-.58758
hwei...	Aged < 35	0	46.95307	1.04693

# Go to: Graph > Legacy Dialogs > Scatter/Dot



Simple Scatterplot

Y Axis:  
Unstandardized Residual [RES\_1]

X Axis:  
Unstandardized Predicted Value [PRE\_1]

Set Markers by:

Label Cases by:

Panel by

Rows:

☐ Nest variables (no empty rows)

Columns:

☐ Nest variables (no empty columns)

Template

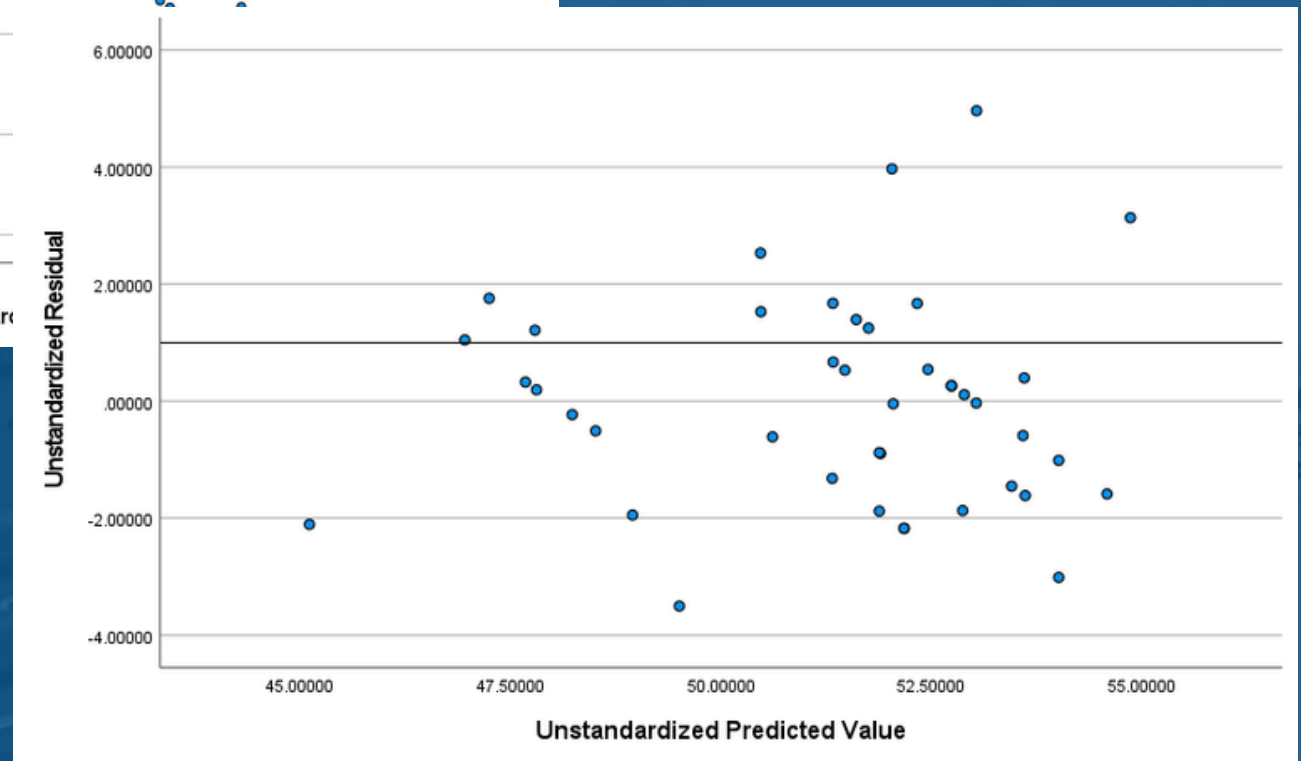
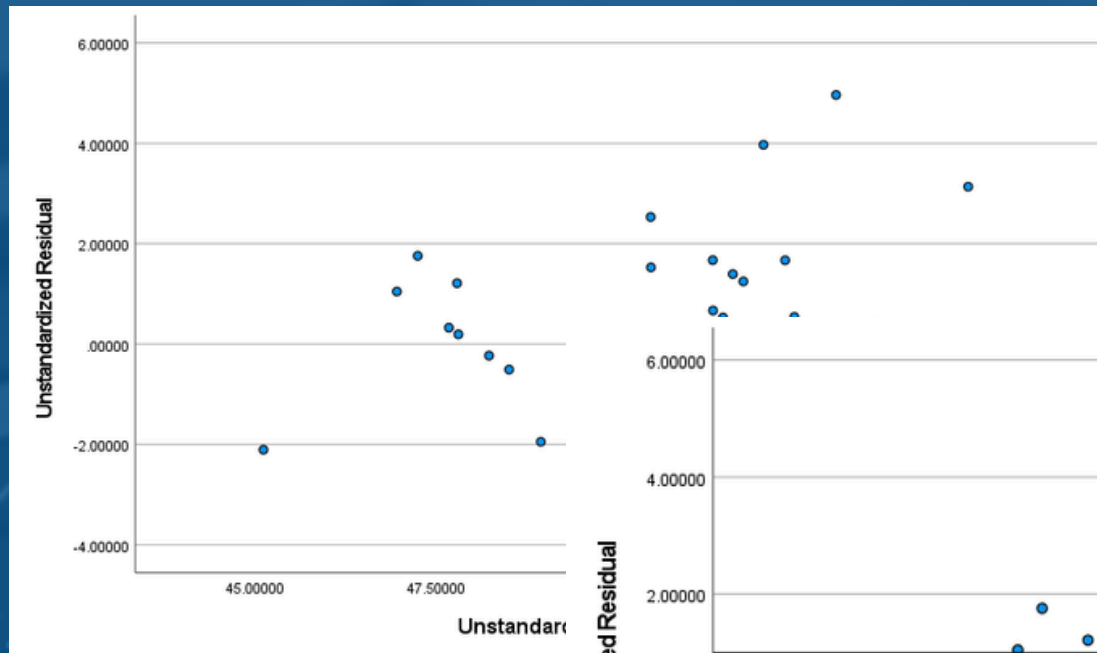
☐ Use chart specifications from:

File...

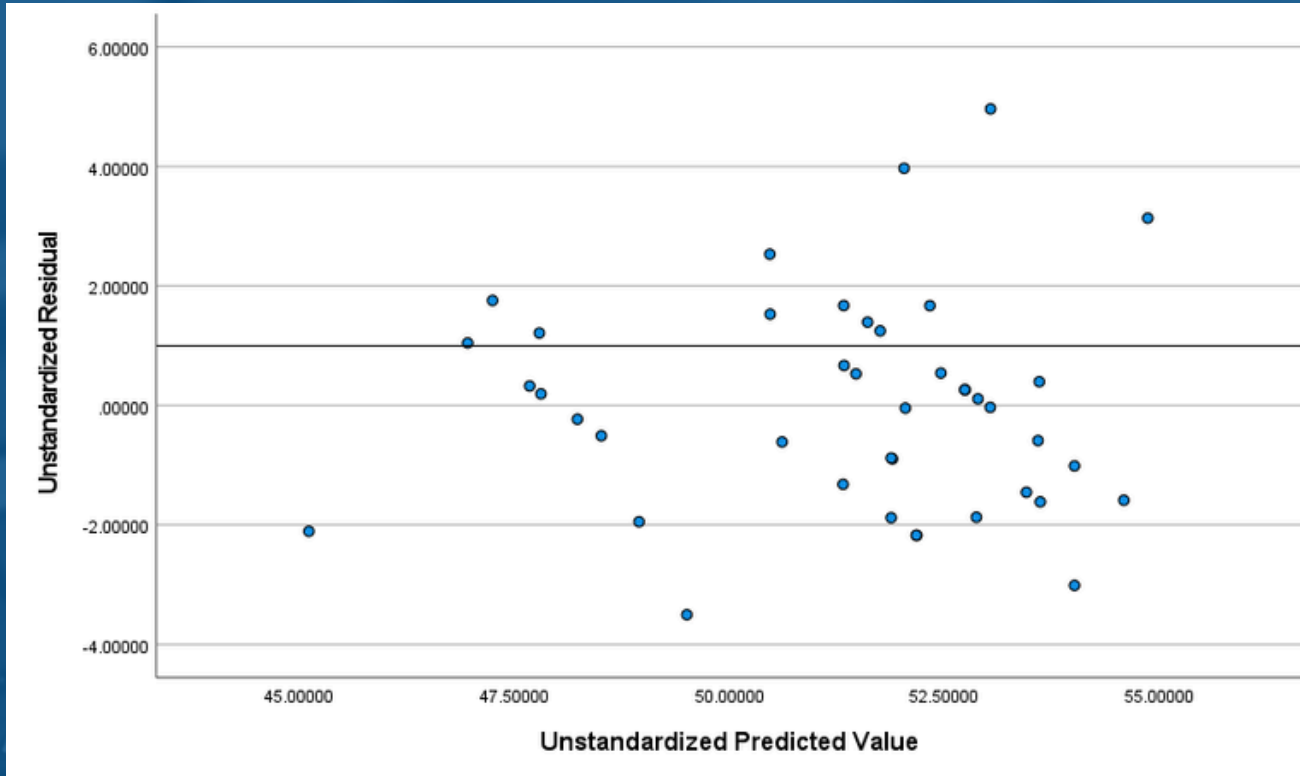
OK Paste Reset Cancel Help

XP - YR

Double click the plot and click



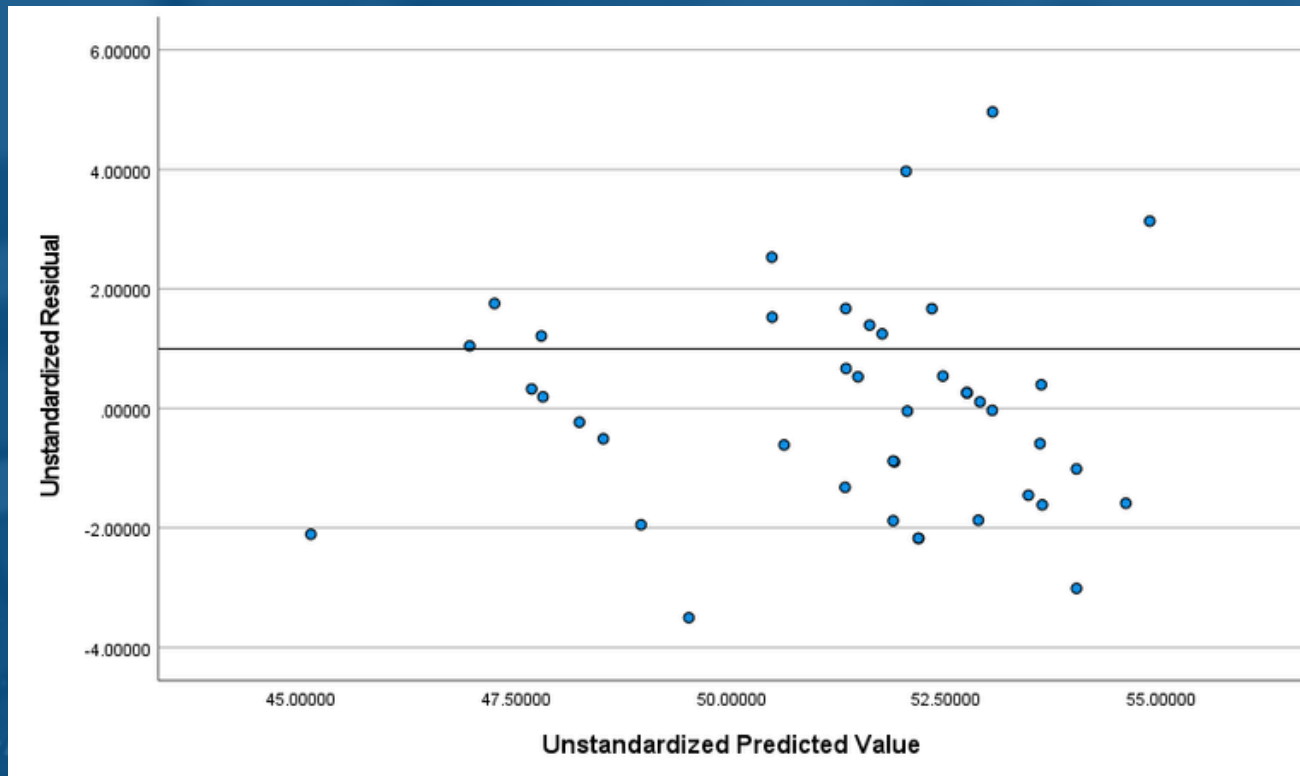




## Linearity:

If there is a peculiar shape of concavity or convexity, then assumption is NOT MET.

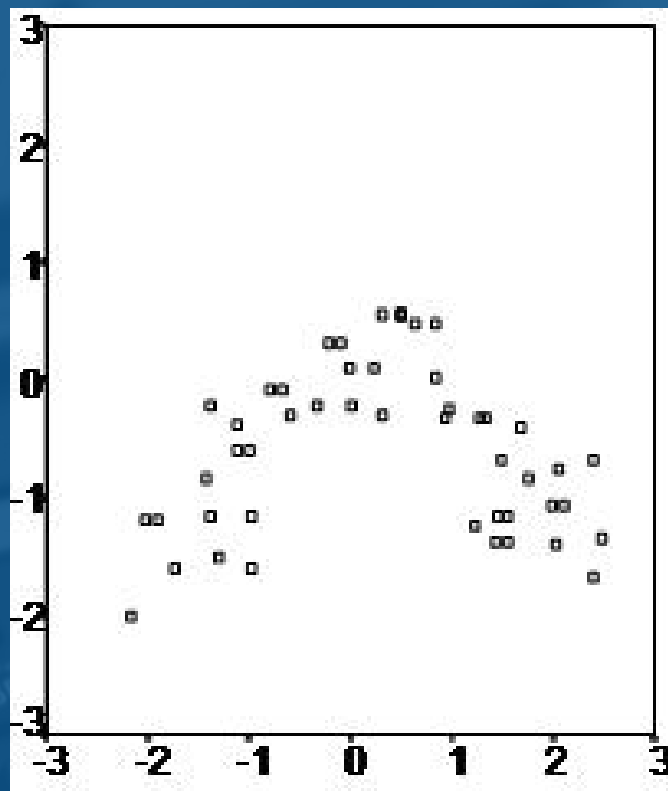
Since there is no peculiar shape, linearity assumption is MET.



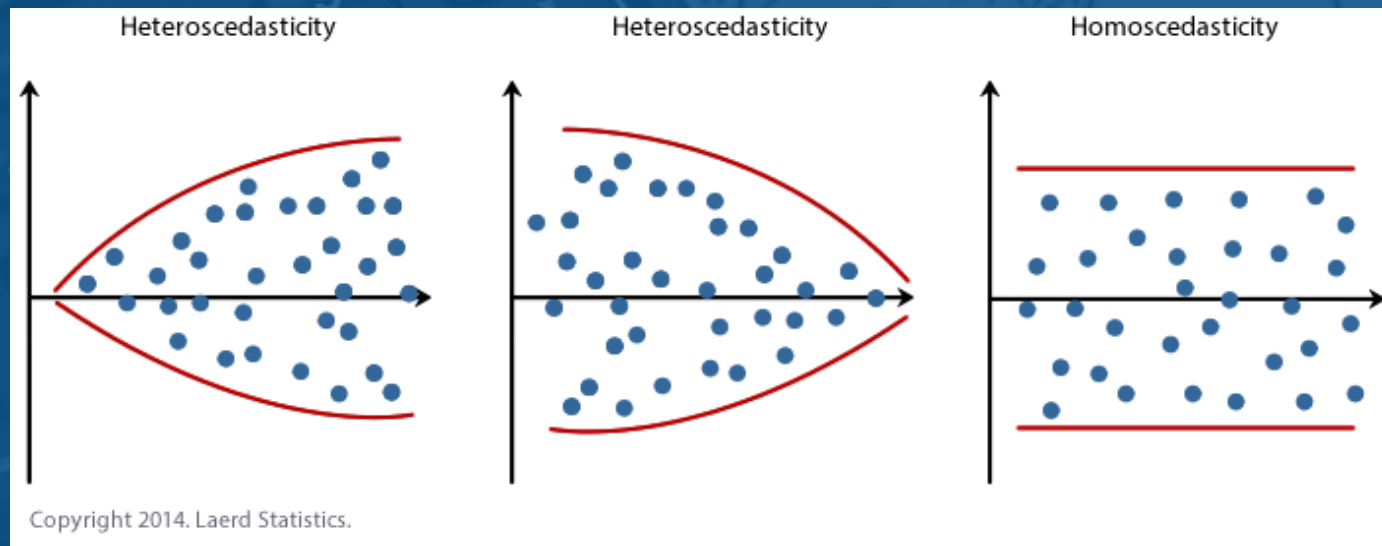
Homoscedasticity (Equal variance):

If there is a peculiar shape of divergence or convergence or fan-shape, then assumption is NOT MET.

Since there is no peculiar shape, homoscedasticity assumption is MET.

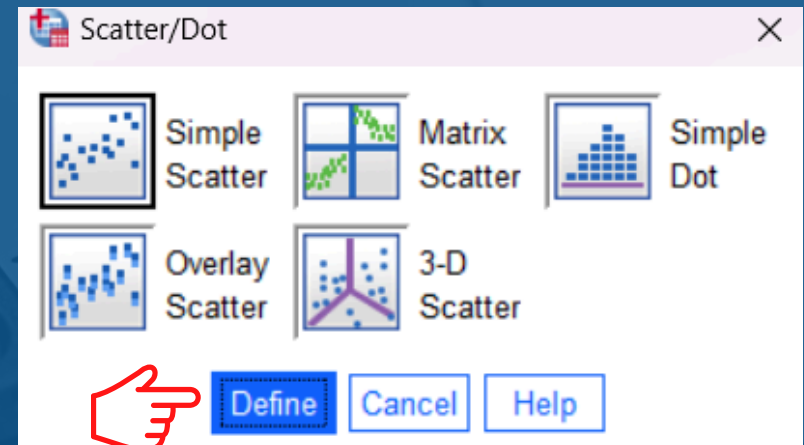
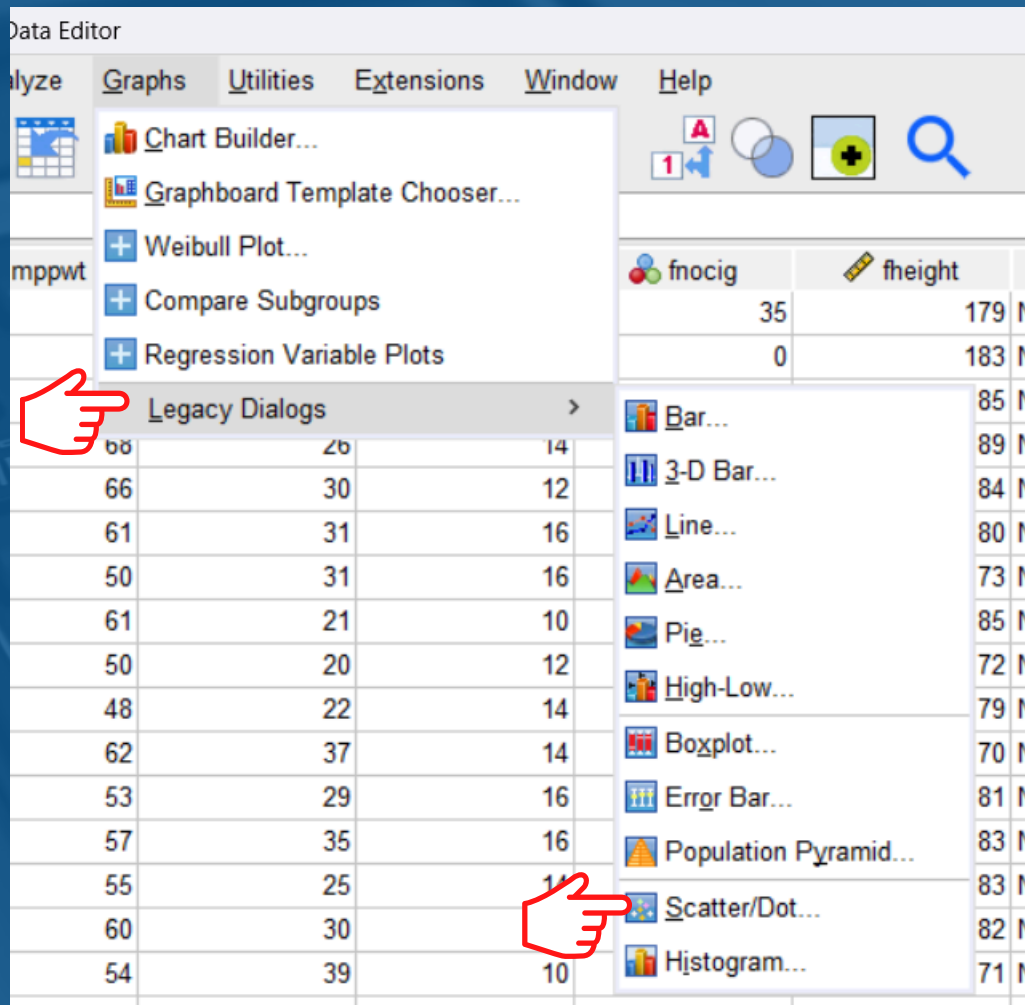


Example of non-linear relationship



# Checking assumption: Linearity of each independent variable

Go to: Graph > Legacy Dialogs > Scatter/Dot





# Mother's age vs Residual

Simple Scatterplot

Y Axis:  
Unstandardized Residual [RES\_1]

X Axis:  
Maternal age [mage]

Set Markers by:

Label Cases by:

Panel by

Rows:

Columns:

Template

☐ Use chart specifications from:

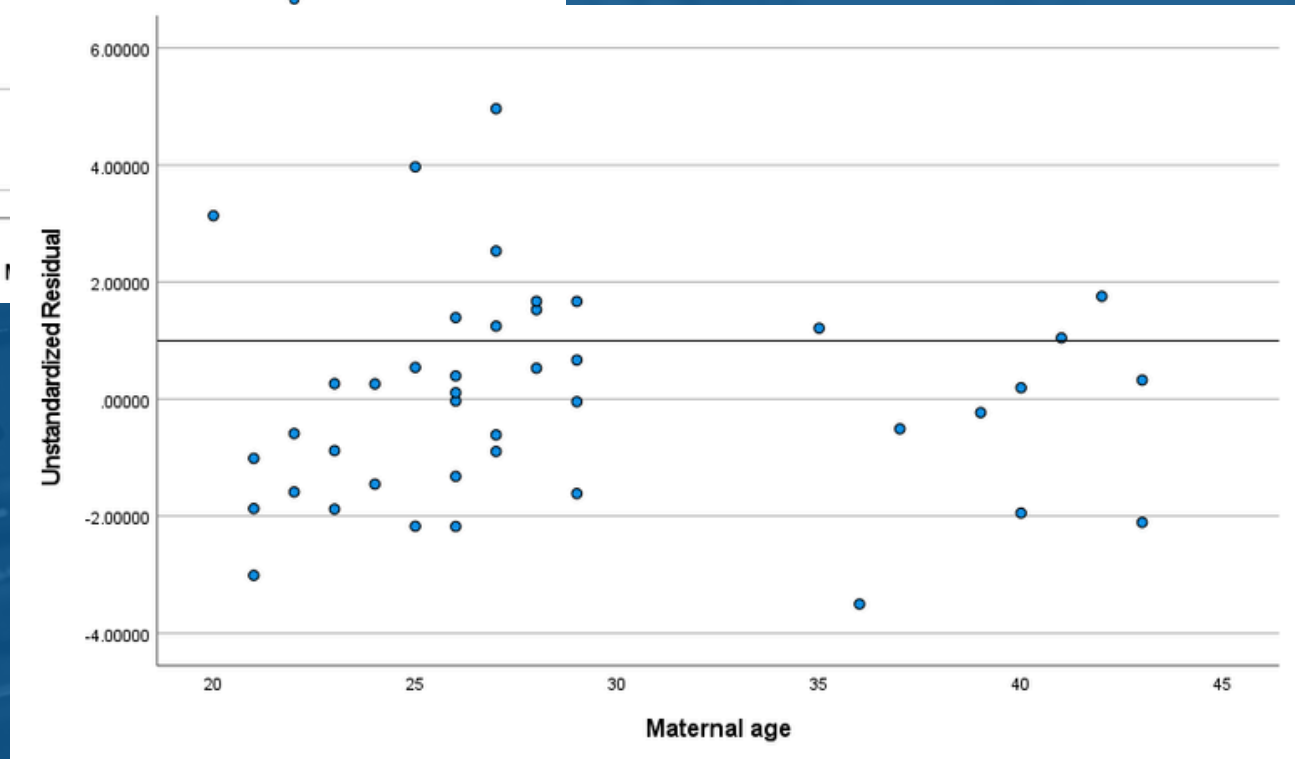
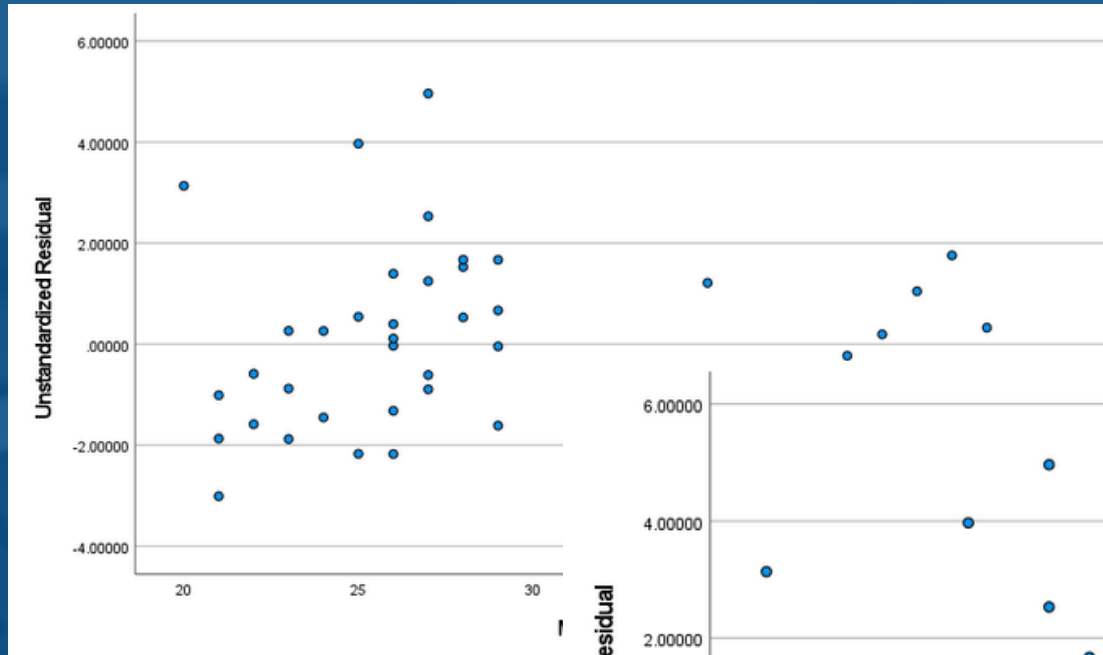
File...

OK Paste Reset Cancel Help

Handwritten red arrows point to the Y Axis variable (Unstandardized Residual [RES\_1]), the X Axis variable (Maternal age [mage]), and the OK button.

XI - YR

Double click the plot and click



There is no peculiar shape, linearity assumption is MET.

# Mother's height vs Residual

Simple Scatterplot

**Y Axis:**  
Unstandardized Residual [RES\_1]

**X Axis:**  
Maternal height (cm) [mheight]

**Set Markers by:**

**Label Cases by:**

**Panel by**

Rows:

☐ Nest variables (no empty rows)

Columns:

☐ Nest variables (no empty columns)

**Template**

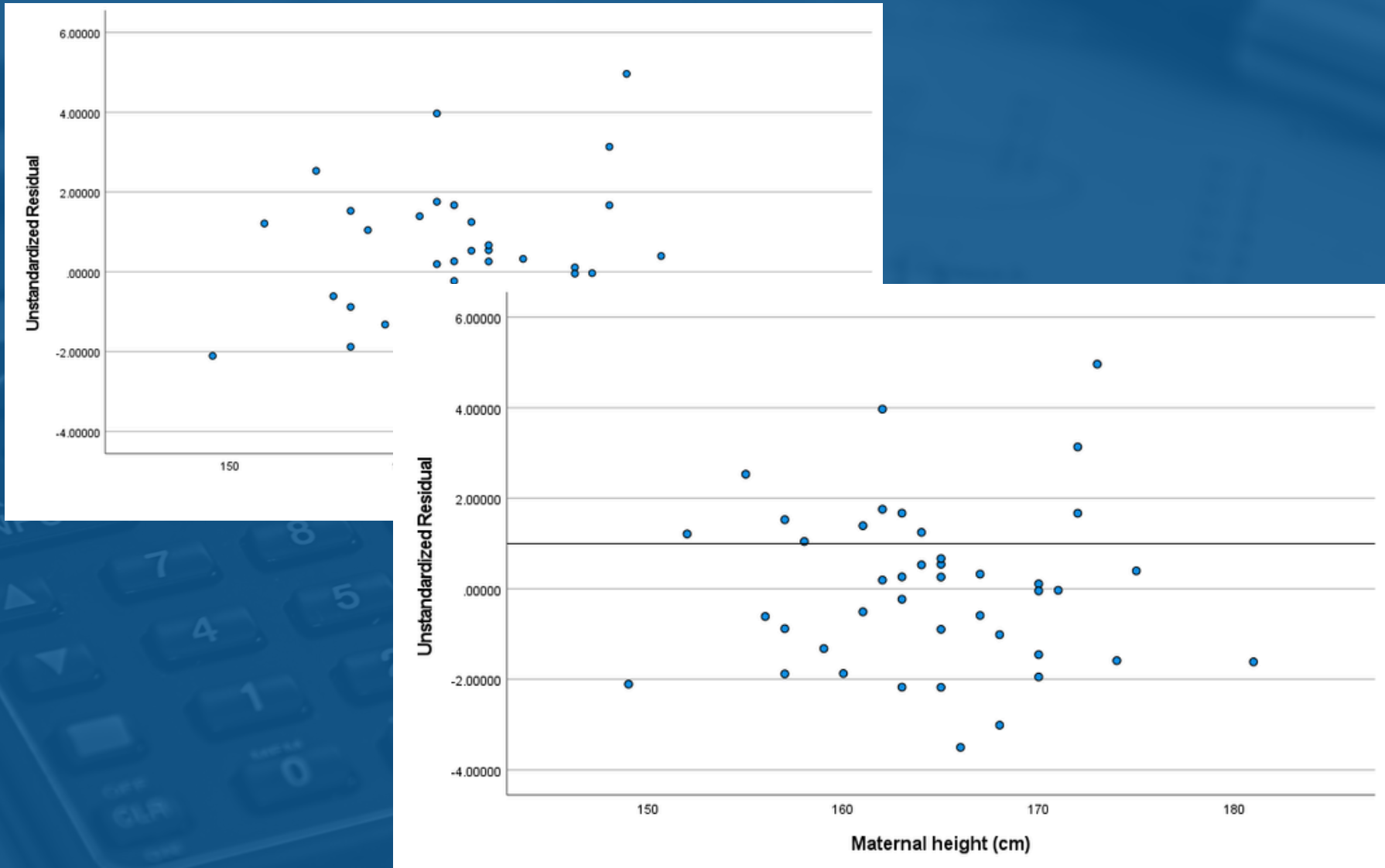
☐ Use chart specifications from:

File...

OK Paste Reset Cancel Help

XI - YR

Double click the plot and click

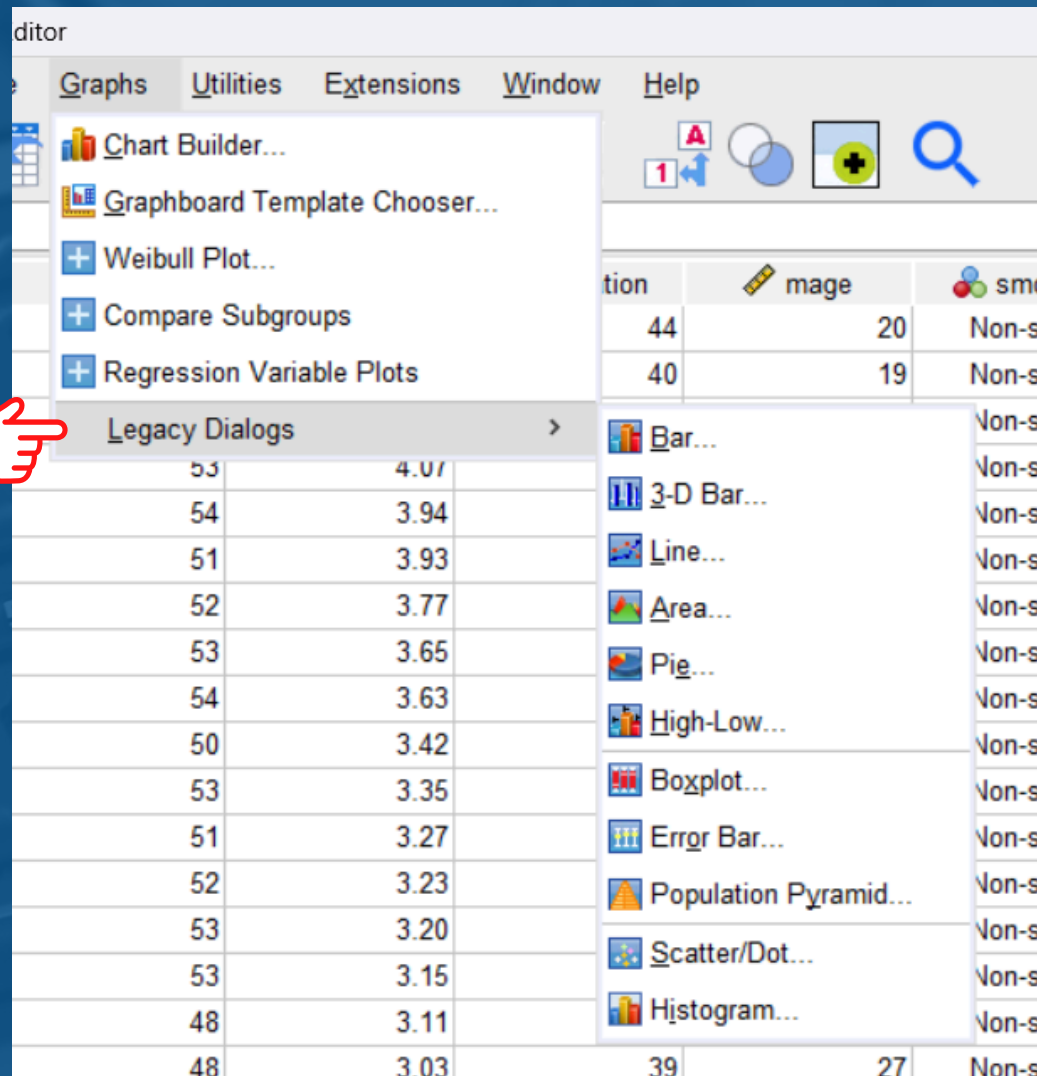


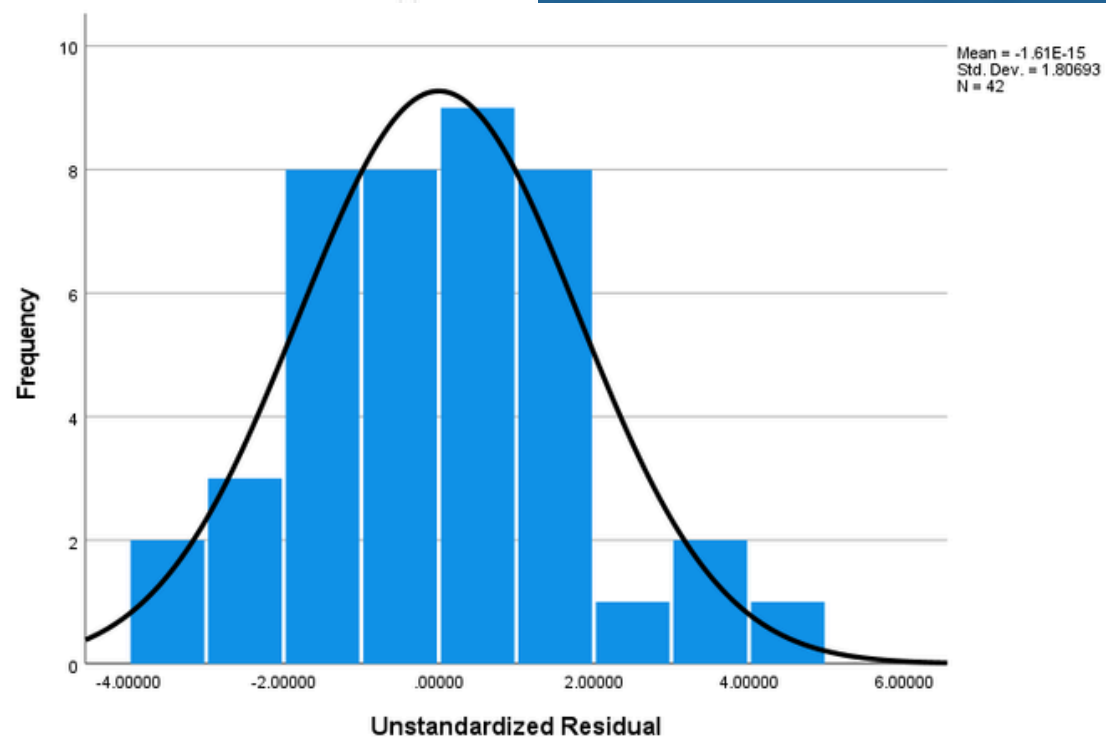
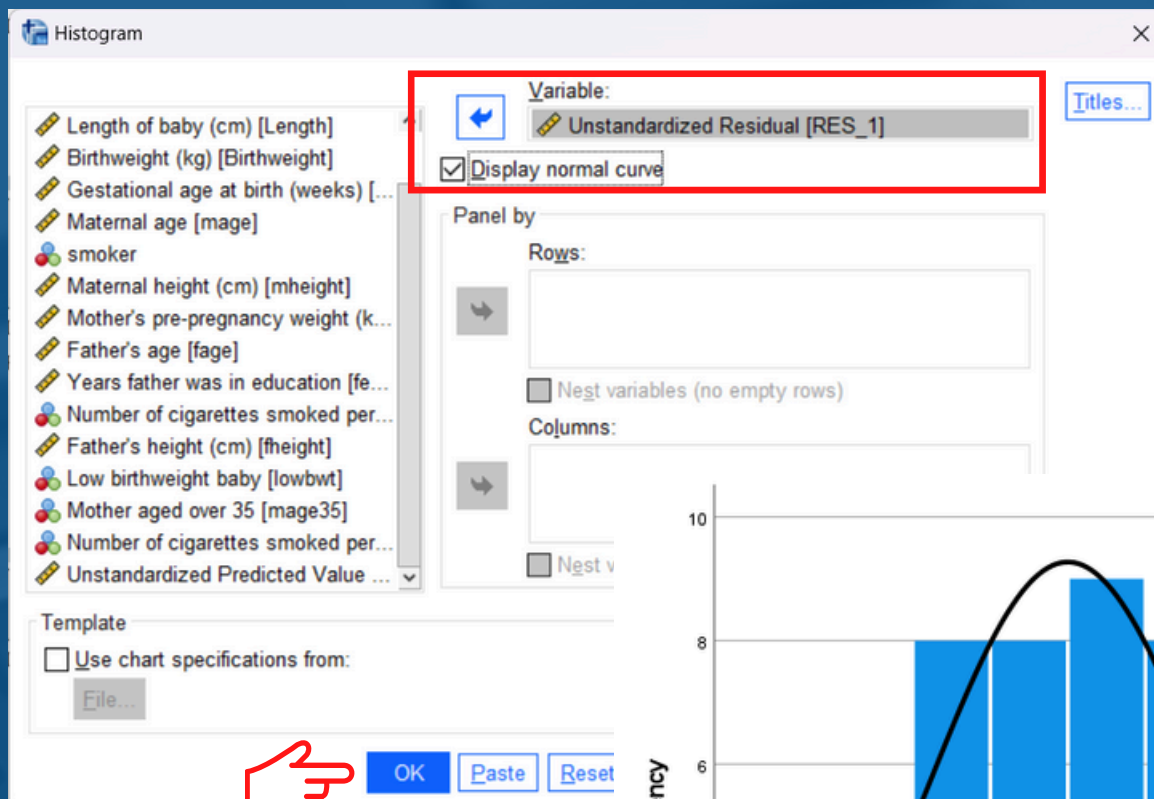
There is no peculiar shape, linearity assumption is MET.



# Checking assumption: Normality distribution of residuals

Go to: Graphs > Legacy Dialogs > Histogram





Residuals are normally distributed. Assumption is met.

## STEP 6: INTERPRETATION AND PRESENTATION

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	35.959	8.019		4.484	<.001	19.738	52.180
	Maternal age	-.282	.045	-.644	-6.303	<.001	-.372	-.191
	Maternal height (cm)	.143	.046	.316	3.094	.004	.049	.236

a. Dependent Variable: Length of baby (cm)

Run the final model. All the assumptions were checked and MET.

## STEP 6: INTERPRETATION AND PRESENTATION

Table 2: Factors associated with the length of baby in HKL (n=42)

Variable	Simple Linear Regression		Multiple Linear Regression	
	b <sup>a</sup> (95% CI)	p-value	b <sup>b</sup> (95% CI)	p-value
Mother's age	-0.32 (-0.41,-0.22)	<0.001	-0.28 (-0.37,-0.19)	<0.001
Mother's height (cm)	0.22 (0.09,0.35)	0.001	0.14 (0.05,0.24)	0.004
Mother's pre-pregnancy weight (kg)	0.16 (0.04,0.28)	0.009	-	-

<sup>a</sup> Crude regression coefficient

<sup>b</sup> Adjusted regression coefficient

All model assumptions are fulfilled.

No multicollinearity problem detected and there were no interaction among the independent variables.

Coefficient of determination ( $R^2$ ) = 0.621

Final model equation:

Length of baby = 35.96 – (0.28\*mother's age) + (0.14\*mother's height)



## STEP 6: INTERPRETATION AND PRESENTATION

- There is a significant linear negative relationship between mother's age and the length of baby. For every one-year increase in the mother's age, the baby's length is 0.28 cm lower. (adjusted  $b = -0.28$ ; 95% CI -0.37,-0.19;  $p < 0.001$ )
- There is a significant linear positive relationship between mother's height and the length of baby. For every 1 cm increase in the mother's height, the baby's length increases by 0.14 cm. (adjusted  $b = 0.14$ ; 95% CI 0.05,0.24;  $p = 0.004$ )
- 62.1% of the variation in the length of baby is explained by mother's age and height according to the multiple linear regression model ( $R^2 = 0.621$ ).



**MDM NURULJANNAH  
BT NOR AZMI**

EMAIL: [nuruljannah@mahsa.edu.my](mailto:nuruljannah@mahsa.edu.my)





**THANK  
YOU**