
EUR/USD - 1,35379 - 00:00:00 14 glu (EES) EUR/USD (Bid), Ticks # 300 / 300

PREDICTING THE FUTURE: INTRODUCTION TO REGRESSION ANALYSIS

NURULJANNAH BT NOR AZMI



What is regression analysis?

We often hear of new, complex "machine learning" methods that:

- Allow us to generate human language.
- Very accurately predict changes in the stock market.
- Recognize that an image contains a person or specific object.

What is regression analysis?



- Regression model analysis is utilized in various applications.
- Adrien-Marie Legendre introduced the concept of regression models in 1805.
- Since then, regression-based modeling has remained fundamental in applied statistics!



What is regression analysis?

Regression analysis comprises a set of statistical techniques aimed at estimating the relationship between:

Dependent variable (outcome variable)

One OR More independent variables (predictor variables)

INDEPENDENT VARIABLE

VARIABLE THAT IS CHANGED

Amount of Water

DEPENDENT VARIABLE

VARIABLE AFFECTED BY THE CHANGE

Size of Plant Number of Leaves Living or Dead?

sciencenotes.org

When to apply Regression Analysis?

Regression analysis can address a broad range of questions, such as:

- 1. Is the relationship between two variables linear?
- 2. Which variable contributes the most to the outcome measurement?
- 3. How accurately can we predict future values?
- 4. Is our outcome variable caused by another variable?

AGE

DIET PLAN



Start at monthly

AGE



Treatment methods



Severity of illness

Length of hospital stay

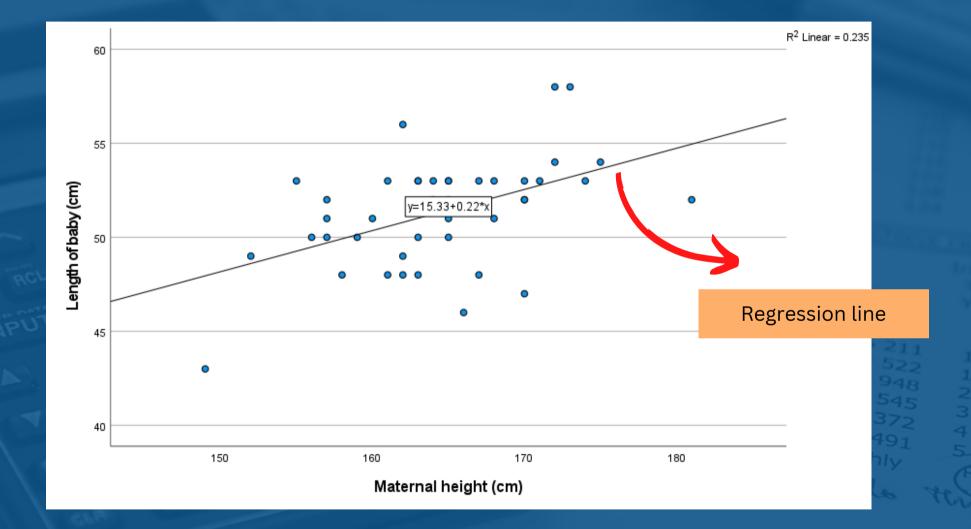
1 859 317 2 499 808 227 076 050 935 2 28 331

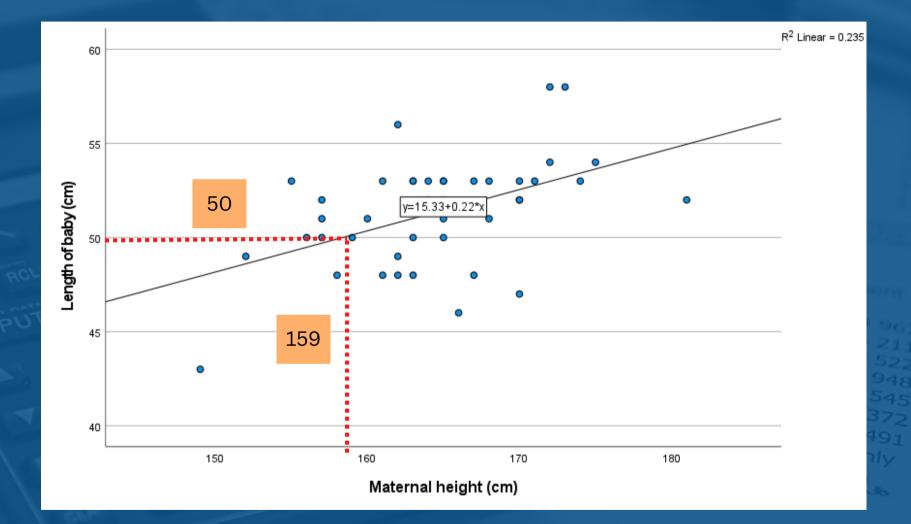
Estimating Coefficient(s)

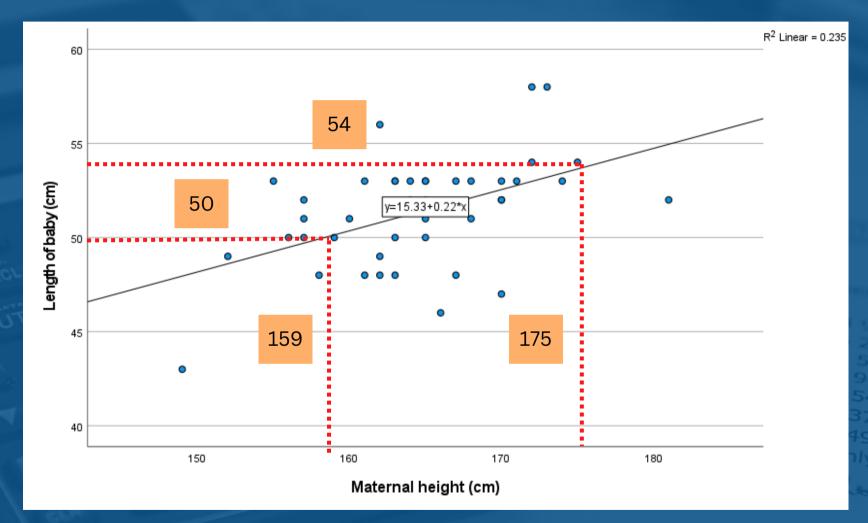
Assuming we are analyzing a basic model consisting of <u>a single</u> <u>predictor</u>, <u>one outcome variable</u>, <u>and one coefficient</u>, we can formally represent this model as follows:

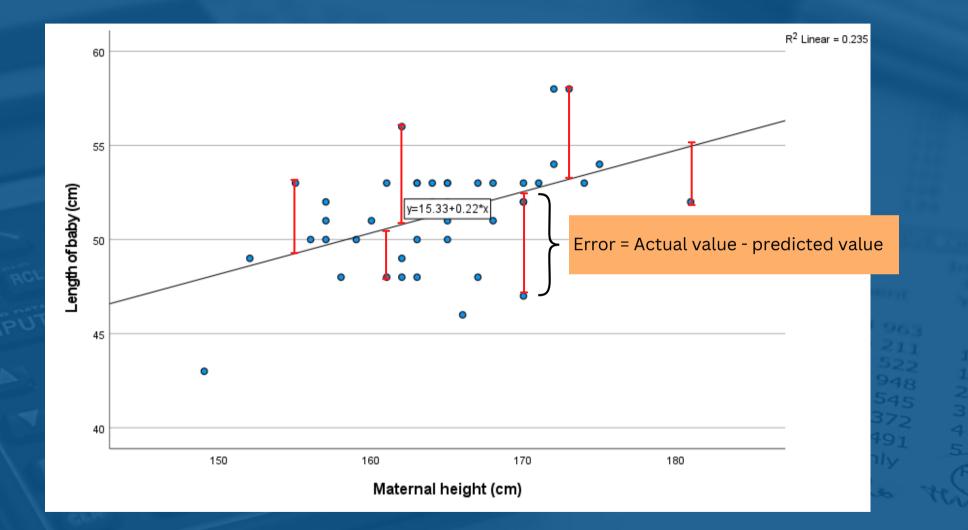
 $Y = \beta_0 + \beta_1 X_1 + error$

 $\begin{array}{l} Y = \text{outcome} \\ \beta_0 = \text{an intercept of the regression line} \\ \beta_1 = \text{a slope of the line / regression coefficient for independent variable} \\ X_1 = \text{independent variable} \\ \text{error} = \text{residual} \end{array}$











SIMPLE LINEAR REGRESSION

Simple Linear Regression is used to estimate the relationship between two quantitative variables.

Dependent variable : numerical

Independent variable : numerical

When to apply Simple Linear Regression?

You can use simple linear regression when you want to identify:

- 1. How strong the relationship between two variables.
- 2.To predict a value of one variable for a given value of the other.

How much the value Y (dependent variable) varies with one unit of change in value X (independent variable)

SIMPLE LINEAR REGRESSION - ONLY ONE INDEPENDENT VARIABLE

Independent variable (x)

Dependent variable (y)

Mother's height



Length of baby

MULTIPLE LINEAR REGRESSION - MORE THAN ONE INDEPENDENT VARIABLES

Independent variables (x)

Dependent variable (y)

Mother's height Mother's weight Age



Length of baby



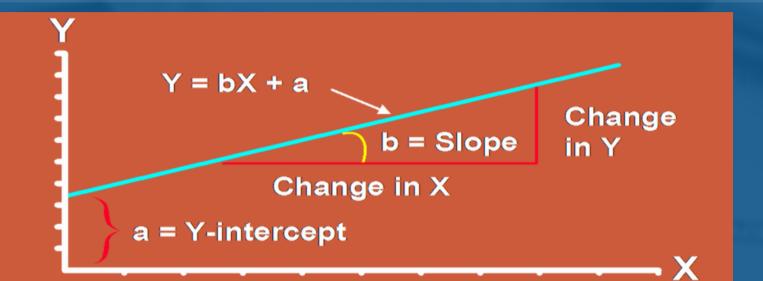
INTRODUCTION

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

Regression Equation



$$y = a + bx$$

x : independent variable y: dependent variable

a: an intercept of the regression line (value of Y when X=0) b: a slope of the line (an amount of change in Y for a unit change in X)

Coefficient of Determination (r^2)

- Ranges from 0 to 1.
- It provides a measure of how well future outcomes are likely to be predicted by the model.
 How much the independent of 'x' is explained by dependent 'y'.

ASSUMPTIONS OF THE MODEL

L

Relationship between the independent and dependent variable is <u>linear</u> <u>(Linearity)</u> Independent observation

Ν

Residuals should be approximately <u>normally</u> distributed

Ε

Homoscedasticity (<u>Equal</u>variances)

CHECKING MODEL ASSUMPTIONS

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

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



	s ID	Headcirc	st Length	Inthweight	🔗 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	

30



Open dataset: birthweight.sav

This dataset contains information on new born babies and their parents. Is there any relationship between maternal height and length of baby?

1.92

40

<

431

20

21 22

EXAMPLE

A study was conducted to determine the relationship between mother's height and the length of baby, with the researcher aiming to forecast the <u>baby's length</u> using the <u>mother's height</u> as a predictor.

Mother's height Length of baby

List down all the variables

Numerical

Numerical

Identify the types of variables

Simple Linear Regression

Identify the right statistical analysis

STEPS IN SIMPLE LINEAR REGRESSION

Step 1: State your research hypothesis

Null hypothesis and Alternative hypothesis Ho : There is no relationship between mother's height and the length of baby H_A : There is a relationship between mother's height and the length of baby

Step 2: Run Simple Linear Regression

Go to: Analyze > Regression > Linear

Eile E	dit ⊻iew	<u>D</u> ata	Transform	<u>A</u> nalyze Po <u>w</u> e	<u>G</u> raphs r Analysis	<u>U</u> tilities	Extensions	<u>W</u> indow	Help		5
				Regor	ts iptive Stati	etice	>		1ન		1
	4	ID	Head	-	sian Statist		>	🔗 Gestati	on	st mage	💰 si
1		136	0	-		lics			44	20	Non
2		101	6	Table			>		40	19	Non
3		46	2	Comp	are Means		>		41	35	Non
4		118	7	Gener	ral Linear N	lodel	>		44	20	Non
5		55	3	Gene	ralized Line	ear Models	>		42	24	Non
6		163	6	Mixed	Models		>		38	29	Non
7		82	0	Correl	ate		>		40	24	Non
8		119		Regre			>				Non
9		108	1 3				$\boldsymbol{\lambda}$	Automat	ic Linear	Modeling	Non
10		82	2	L <u>o</u> glin				Linear			Non
11		168	3	Neura	l Networks		J	Curve Es	timation		Non
12		108	8	Class	ify		>	Reartial Lo	east Sou	ares	Non
13		110	7	Dimer	nsion Redu	ction	>				Non
14		75	5	Scale			>	🔚 Binary L			Non
15		105	В	Nonpa	arametric T	ests	>	Multinon	nial Logis	stic	Non
16		32	1		asting		>	Grdinal			Non
17		69	7				>	Probit			Non
18		80	8	<u>S</u> urviv							Non
19		160	0	Multip	le Respon	se	>	Nonlinea			Non
20		131	3	💋 Missir	ng <u>V</u> alue A	nalysis		Weight E	stimatio	n	Non
21		79	2	Multip	le Imputati	ion	>	> 1 2-Stage Least Squares			
22		138	8	Comp	lex Sample	es	>	Quantile.			
23		57	5	Simula							
24		56	9	-				Optimal	Scaling	(CATREG)	
25		136	3	Qualit	y Control		>		37	20	

ta Linear Regression			×	<				
 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 [fedyrs] Number of cigarettes smoked per day by fath Father's height (cm) [fheight] Low birthweight baby [lowbwt] Mother aged over 35 [mage35] Number of cigarettes smoked per day by mot Vistandardized Predicted Value [PRE_1] Unstandardized Residual [RES_1] Unstandardized Residual [RES_2] Unstandardized Residual [RES_2] 	Dependent:	Next Qptions Style Bootstra		gression Co Estimates Confidence Level(%): 9 Covariance siduals Durbin-Wat Casewise o Outliers o All cases	intervals P intervals P 5 P matrix C tson diagnostics putside: 3	A A A A A A A A A A A A A A		
		Co	efficients ^a					45
Model	Unstandardize B	d Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confider Lower Bound	nce Interval for B Upper Bound	2
1 (Constant)		10.271		1.493	.143	-5.425	36.093	40
Maternal h	eight (cm) .219	.062	.485	3.507	.001	.093	.345	

a. Dependent Variable: Length of baby (cm)

Model Summary ^b									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate					
1	.485 ^a	.235	.216	2.599					
a Predictors: (Constant) Maternal height (cm)									

a. Predictors: (Constant), Maternal height (cm)

b. Dependent Variable: Length of baby (cm)

23.5% of the variation in length of baby is explained by mother's height according to the linear regression model ($r^2 = 0.235$).

ANOVA ^a									
Model	Sum of Squares df Mean Square F								
1	Regression	83.110	1	83.110	12.302	.001 ^b			
	Residual	270.223	40	6.756					
	Total	353.333	41						

a. Dependent Variable: Length of baby (cm)

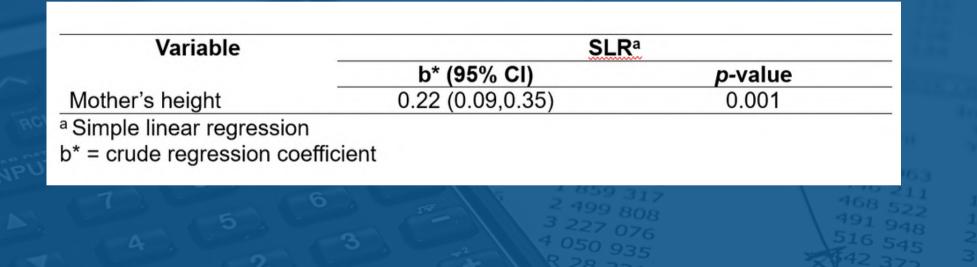
b. Predictors: (Constant), Maternal height (cm)

This table indicates that the regression model predicts the dependent variable significantly well (refer to p-value).

Here, p < 0.05, which is less than 0.05, and indicates that the regression model statistically significantly predicts the outcome variable.

Result presentation for Simple Linear Regression

Table 1: Simple linear regression



Step 3: Checking assumptions

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

Checking assumption : Linearity

Go to: Graph > Legacy Dialogs > Scatter/Dot

X

Simple

Dot

.....

Matrix

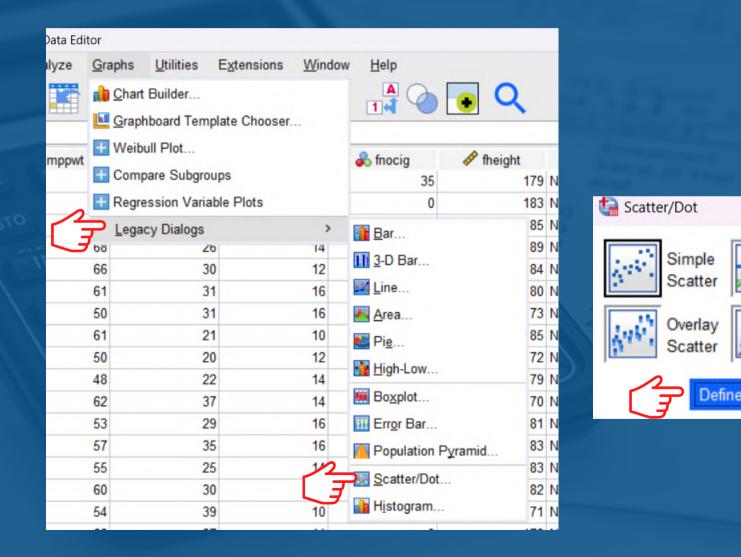
3-D

Cancel

Scatter

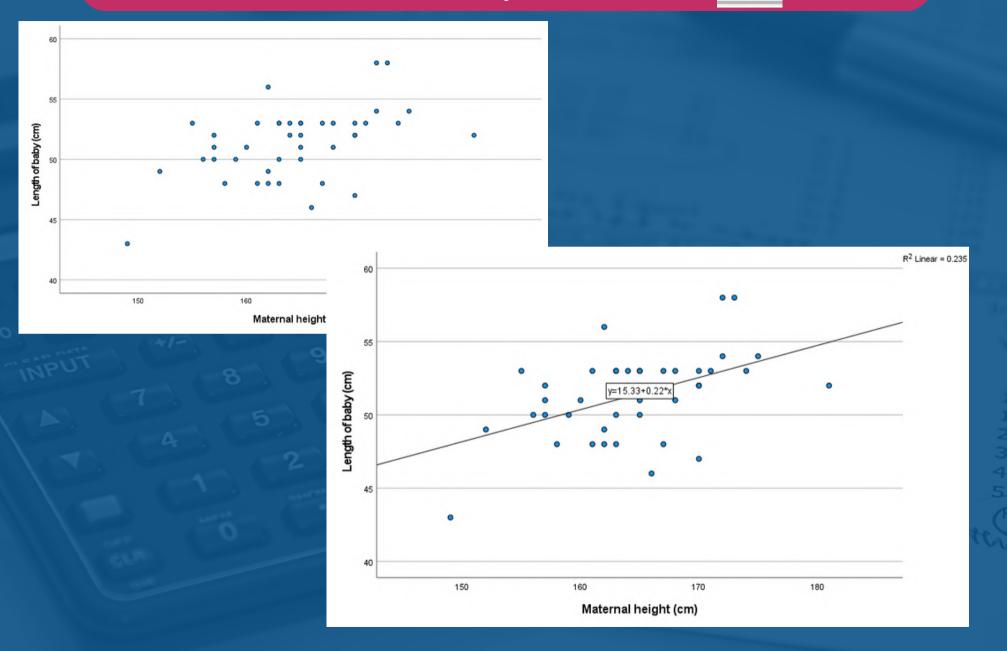
Scatter

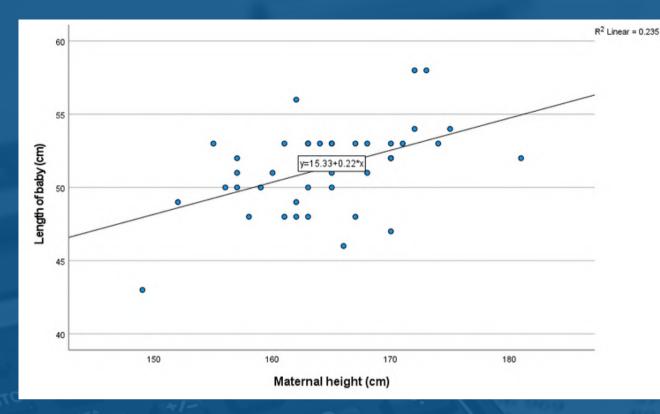
Help



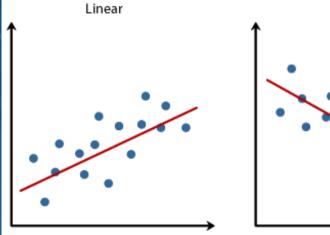
	Y Axis:	<u>T</u> itles	
 ✓ Baby ID [ID] ✓ Head circumference (cm) [Headcirc] 	Length of baby (cm) [Length]	Options	
Birthweight (kg) [Birthweight]	X Axis:		
IGestational age at birth (weeks) [Gest	Maternal height (cm) [mheight]		
Maternal age [mage] Smoker	Set Markers by:	_	
 Mother's pre-pregnancy weight (kg) [Father's age [fage] 	Label <u>C</u> ases by:		
 Years father was in education [fedyrs] Number of cigarettes smoked per day 	Panel by		
Father's height (cm) [fheight]	Ro <u>w</u> s:		
Low birthweight baby [lowbwt]	14		
 Mother aged over 35 [mage35] Number of cigarettes smoked per day 	4		
	Nest variables (no empty rows)	_	
	Columns:		
	100		
	*		
	Nest variables (no empty columns)		
	magest variables (no empty countins)		
Template			
Use chart specifications from:		n	
Eile			

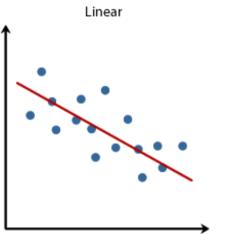
Double click the plot and click

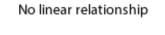




Linear relationship of two continuous variables. Linearity assumption is met.







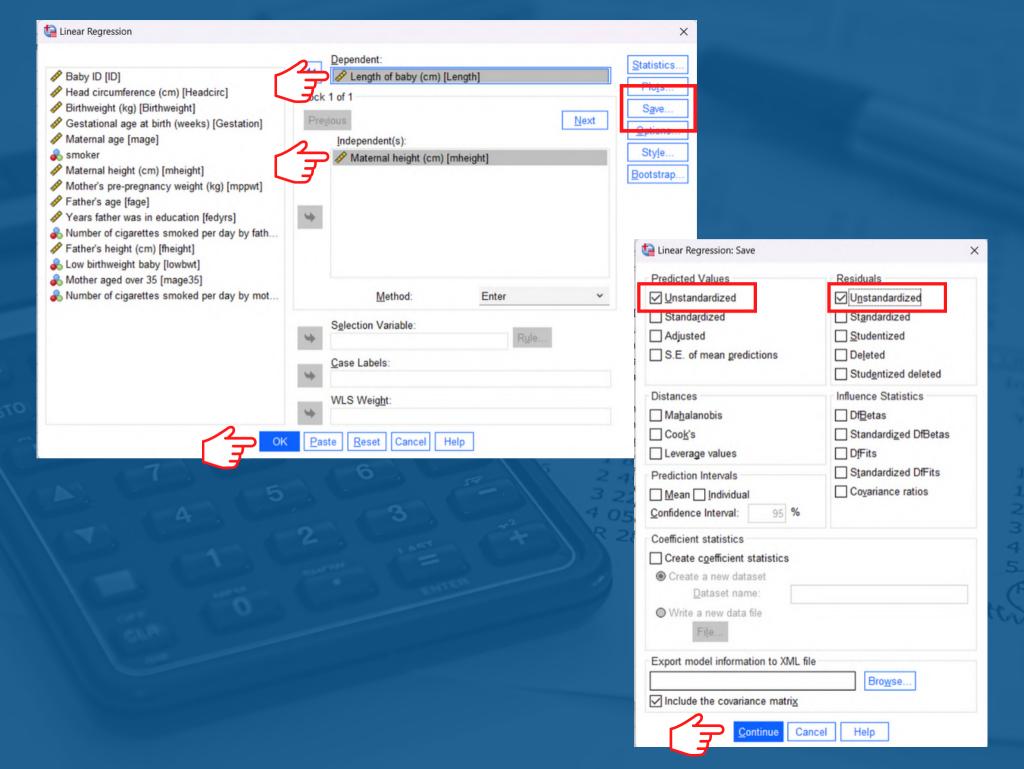
Example of linear and non-linear relationship

Copyright 2014. Laerd Statistics.

Checking assumption: Homoscedasticity (equal variances)

Go to: Analyze > Regression > Linear

<u>Eile</u> <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	Analyze	<u>G</u> raphs	Utilities	Extension	ns <u>W</u> indow	Help		
		10 3	Powe	r Analysis		>		A	0 - (a
		_	Repo	rts		>				
	A 10	A		riptive Stati	stics	>	A		•	
	13 ID	Head	Baye	sian Statist	ics	>	Sestat	10n 44	Mage 20	🗞 sr Non
1	13		Table	s		>		44	20	Non
3		52	Com	are Means		>		40	35	Non
4	11			ral Linear N		>		41	20	Non
5		53	_	ralized Line		>		44	20	Non
6	16			-	ai wouels	,		38	24	Non
7		20	-	d Models				40	23	Non
8	11		<u>C</u> orre			>			24	Non
9	10	_	Regre	ession		\rightarrow	Automat	ic Linear	Modeling	Non
10		22	Loglin	near			Linear			Non
11	16	33	Neura	al Networks			Curve Es	stimation		Non
12	10	1088		Classify		>	Partial L			Non
13	11	07	Dime	nsion Redu	ction	>	_			Non
14	7	55	Scale			>	Binary L	ogistic		Non
15	10	58	Nonp	arametric T	ests	>	Multinon	nial Logis	stic	Non
16	3	21		asting		,	Grdinal			Non
17	6	97	Surviv			>	Probit			Non
18	8	08	_		2.0		Nonlinea			Non
19	16	00		ble Respons		>				Non
20	13	13	💕 Missi	ng <u>V</u> alue Ar	nalysis		Weight I			Non
21		92	Multi	ole Imputati	on	>	2-Stage	Least So	quares	
22	13		Comp	ex Sample	es	>	Quantile			
23	-	75	Bimul	ation			Optimal	Scaling	(CATREG)	
24	-	59	Quali	ty Control		>	The second	-		
25	13		-		poral Model	ing >		37	20	
26	-	00						35	41	_
27	4	31	Direc	t Marketing				33	20	



Predicted value



y = a + bx Length of baby = 15.33 + (0.22*mother's height)

Predicted length of baby whose mother's height is 162cm is:

Length of baby = 15.33 + (0.22*162) = 50.97

Residuals



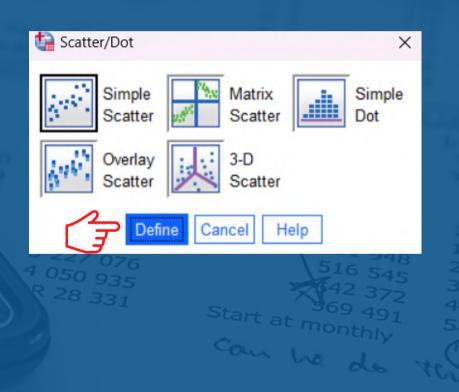
The difference between the observed value of the dependent variable and the value predicted by the regression line.

Residual = observed length - predicted length Residual = 56 - 50.8 = 5.2

Go to: Graph > Legacy Dialogs > Scatter/Dot

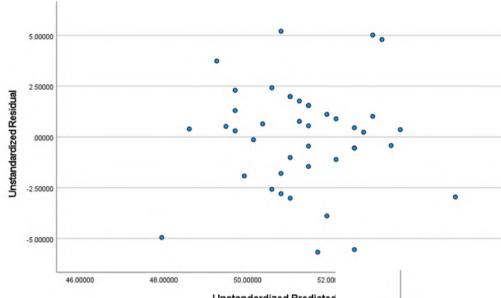
Data	Editor
Jata	Luitor

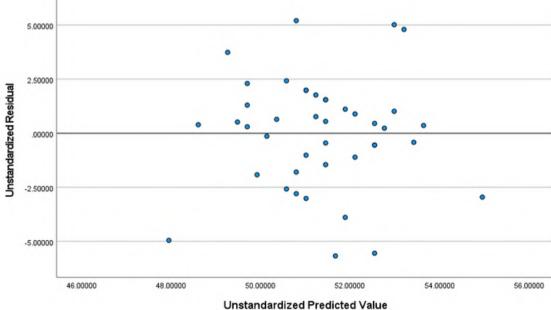
lyze	<u>G</u> raphs	Utilities	Extensions	<u>W</u> indow	<u>H</u> elp		
		t Builder hboard Tem	plate Chooser			● Q	
nppwt	+ Weit	oull Plot			💑 fnocig	I fheight	
	+ Com	pare Subgro	ups		35	179	N
	+ Regr	ession Varia	ble Plots		0	183	N
~2		acy Dialogs		>	1 Bar	85	5 N
J	60	4	b	14		89	N
	66	3	0	12	111 <u>3</u> -D Bar	84	N
	61	3	1	16	📫 <u>L</u> ine	80	N
	50	3	1	16	🐴 <u>A</u> rea	73	B
	61	2	1	10	Ni <u>e</u>	85	N
	50	2	0	12		72	2
	48	2	2	14	High-Low	79	N
	62	3	7	14	I Boxplot	70	N
	53	2	9	16	III Error Bar	81	N
	57	3	5	16	Population P	vramid 83	BN
	55	2	5	12	Scatter/Dot	83	B
	60	3	0	LJ		82	2 N
	54	3	9	10	🚹 Histogram	71	N



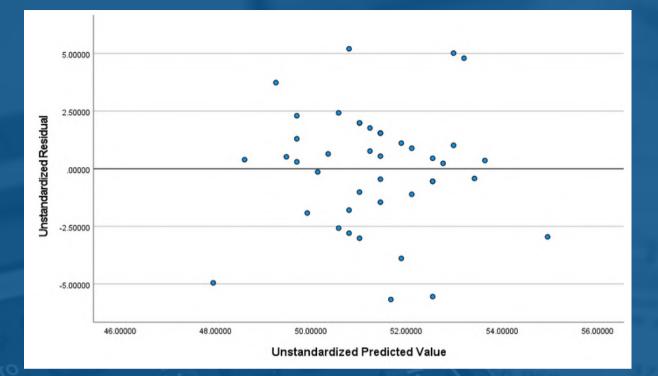
 ✓ Baby ID [ID] ✓ Head circumference (cm) [Headcirc] ✓ Length of baby (cm) [Length] 	Vinstandardized Residual [RES_1]	XP - YR	
 Birthweight (kg) [Birthweight] Gestational age at birth (weeks) [Gest 	Set Markers by:		
Maternal age [mage]	*		
Maternal height (cm) [mheight]	Label <u>C</u> ases by:		
Mother's pre-pregnancy weight (kg) [Father's age [fage]	Panel by		
Years father was in education [fedyrs]	Ro <u>w</u> s:		
Father's height (cm) [fheight]	*		
Low birthweight baby [lowbwt] Mother aged over 35 [mage35]	Nest variables (no empty rows)	Contraction of the Institute of the Inst	
Number of cigarettes smoked per day	Columns:		
	*	7-	
	Nest variables (no empty columns)	49	
Template		2 4.	
Use chart specifications from:		mor	

Double click the plot and click 📘



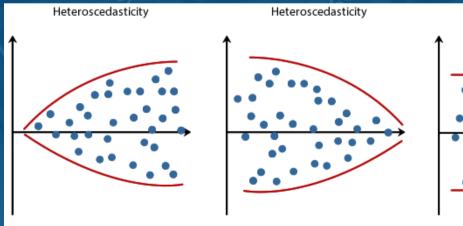






There is no pattern in the scatter. Homoscedasticity assumption is met.

Homoscedasticity



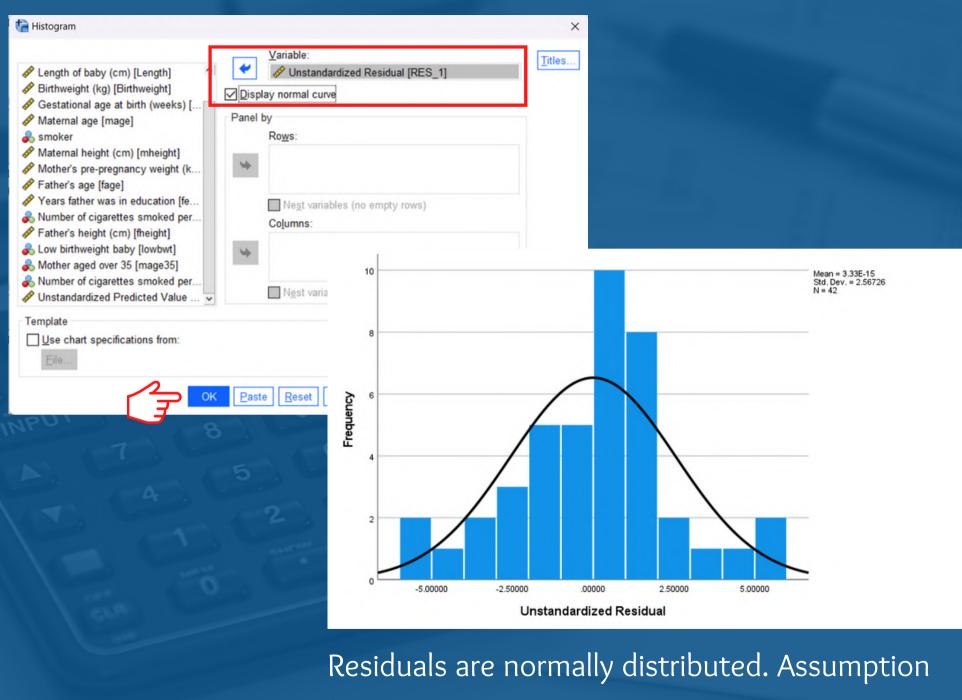
Example of heteroscedasticity and homoscedaticity

Copyright 2014. Laerd Statistics.

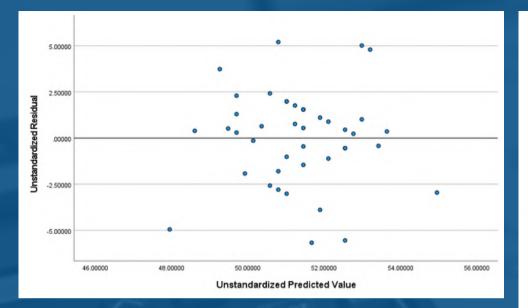
Checking assumption: Normality distribution of residuals

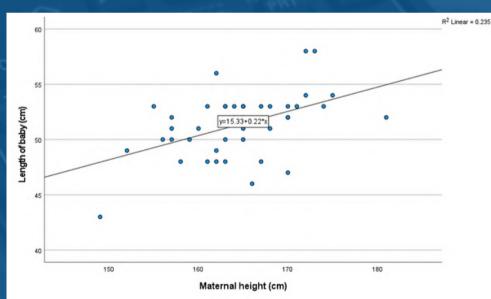
Go to: Graphs > Legacy Dialogs > Histogram

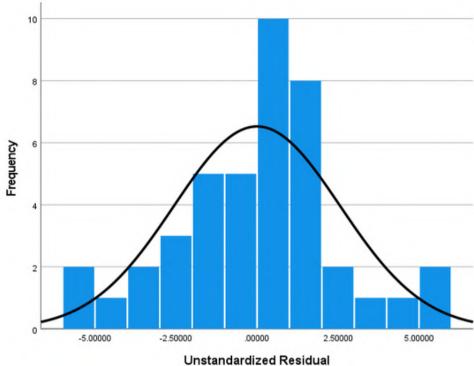
ditor						
<u>G</u> raphs	<u>U</u> tilities	Extensions	Window	w <u>H</u> elp		
hart B		plate Chooser		A 14	 Image: A state Image: A state<td>ર</td>	ર
- H Weibull	Plot			tion	🛷 mage	💰 smo
🛨 Compa	re Subgro	oups		44	20	Non-sr
+ Regres	sion Varia	able Plots		40	19	Non-si
	Dialogs		>	Bar.		Von-si
_	53	4.07				Von-si
	54	3.94		11 <u>3</u> -D		Von-si
	51	3.93		ine.		Von-si
1	52	3.77		🔼 <u>A</u> rea		Von-si
	53	3.65		Nie		Non-si
	54	3.63		High		Von-si
	50	3.42				Von-si
-	53	3.35		I Box	plot	Von-sr
	51	3.27		HI Erro	Bar	Non-si
	52	3.23			ulation Pyramid	Von-si
-	53	3.20			ter/Dot	Von-si
	53	3.15	1			Von-si
	48	3.11	LŦ	Histo	ogram	Von-si
	48	3.03	<u> </u>	39	27	Non-sr



is met.







Assumptions for homoscedasticity, linearity and normally distributed are met.

Step 4: Result Interpretation & Conclusion

			Co	befficients"				
		Unstandardize	ed Coefficients	Standardized Coefficients			95.0% Confider	ice Interval for B
Model		В	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)

Interpretation:

Increasing the mother's height by 1 cm will result in a 0.2 cm increase in the length of the baby (b=0.22, 95% CI 0.09, 0.35, p=0.001).

Regression equation: y = a + bx length of baby = 15.33 + (0.22*mother's height)



0 8 8 8 8 B

100.000.000

REGRESSION ANALYSIS

DEPENDENT VARIABLE	INDEPENDENT VARIABLE	STATISTICAL TEST
Numerical	Numerical	Multiple Linear Regression
Categorical (dichotomous)	Numerical and categorical	Multiple/Binary Logistic Regression
Categorical (polytomous - nominal)	Numerical and categorical	Multinomial Logistic Regression
Categorical (ordinal)	Numerical and categorical	Ordinal Logistic Regression



MDM NURULJANNAH BT NOR AZMI

EMAIL: nuruljannah@mahsa.edu.my

THANK YOU

EUR/USD - 1,35379 - 00:00:00 14 glu (EES) EUR/USD (Bid), Ticks # 300 / 300

PREDICTING THE FUTURE: INTRODUCTION TO REGRESSION ANALYSIS

NURULJANNAH BT NOR AZMI



MULTIPLE LINEAR REGRESSION

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



MULTIPLE LINEAR REGRESSION

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)

Dependent variable (y)

Mother's height



Length of baby

MULTIPLE LINEAR REGRESSION - MORE THAN ONE INDEPENDENT VARIABLES

Independent variables (x)

Dependent variable (y)

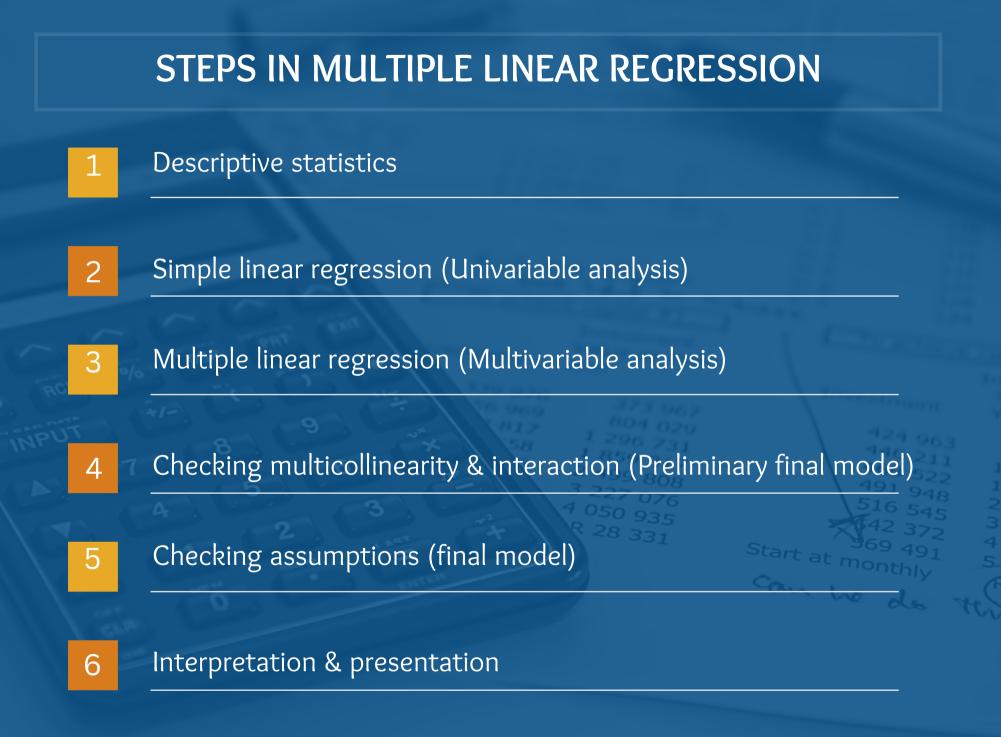
Mother's height Mother's weight Age



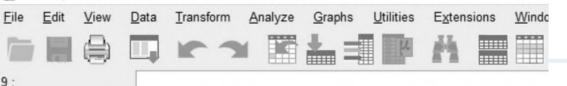
Length of baby

Multiple Linear Regression Model

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$
- $\begin{array}{l} Y = \text{outcome} \\ \beta_0 = \text{intercept} \\ \beta_1 \ldots \beta_n = \text{regression coefficient for independent variable} \\ X_1 \ldots X_n = \text{independent variable} \end{array}$



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



	d ID	Headcirc	Dength	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	



Open dataset: birthweight.sav

This dataset contains information on new born babies and their parents admitted in HKL. A researcher wants to determine the factors that are associated with the length of baby.

1.92

EXAMPLE

RQ: What are the <u>factors</u> that associated with the <u>length of baby</u>?

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.

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-value < 0.25 and/or clinically important.

Go to: Analyze > Regression > Linear

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

Eile Edit	View Da	ta <u>T</u> ransfor	m <u>A</u> nalyze	Graphs	Utilities	Extensions	Window	Help		
			Pow	er Analysis		>				0
			Rep	orts		>		14 4		1
			Des	criptive Stat	istics	>				
	🛷 ID	1	Hear	esian Statis		>	Sestatio	n 🤞	mage	💰 sn
1		1360	Tabl			>		44	20	Non
2		1016						40	19	Non
3		462		npare Means		>		41	35	Non
4		1187	Gen	eral Linear M	Nodel	>		44	20	Non
5		553	Gen	eralized Line	ear Models	>		42	24	Non
6		1636	Mix	ed Models		>		38	29	Non
7		820	Corr	relate		>		40	24	Non
8		1191		ression		>				Non
9		1081	-			~	Automatic	Linear Mo	deling	Non
10		822		linear		[⊐	Linear			Non
11		1683	Neu	ral Networks	3		Zurve Esti	mation		Non
12		1088	Clas	ssify		>	Partial Lea	ast Square	s	Non
13		1107	Dim	ension Redu	uction	>			•	Non
14		755	Sca	le		>	Binary Log			Non
15		1058	Non	parametric 1	ests	>	Multinomia	al Logistic.		Non
16		321	_	ecasting		>	Grdinal			Non
17		697				>	Probit			Non
18		808	Sun							Non
19		1600	Mult	tiple Respon	se	>	Nonlinear.			Non
20		1313	💋 Miss	sing <u>V</u> alue A	nalysis		Weight Es	timation		Non
21		792	Mult	tiple Imputat	ion	>	2-Stage L	east Squar	res	
22		1388	Con	nplex Sampl	es	>	Quantile			
23		575	_	ulation					TDEOL	
24		569					Optimal S	caling (CA	IREG)	
25		1363	_	lity Control	1.11	>		37	20	
26		300	Spa	tial and Tem	poral Mode	ling >		35	41	
27		431	Dire	ct Marketing	1	>		33	20	

Length of baby vs Mother's age

ta Linear Regression					×	ta 1	Linear Regress	ion: Statistics		×
 Baby ID [ID] Head circumference Birthweight (kg) [Birt Gestational age at b Maternal age [mage] smoker Maternal height (cm) Mother's pre-pregnar Father's age [fage] Years father was in d Number of cigarettes Father's height (cm) Low birthweight baby Mother aged over 35 Number of cigarettes 	weight] th (weeks) [Gestation] [mheight] cy weight (kg) [mppwt] ducation [fedyrs] smoked per day by fat [fheight] [lowbwt] [mage35]	Pregiou Anti-	s 3lock 1 of 1 9 Matemal age [mage]	N	Ext Qptions Syve Style Bootstrap		egression Co Estimates Co <u>n</u> fidence Level(%): 9 Co <u>v</u> ariance esiduals	intervals	<u>M</u> odel fit R <u>s</u> quared change Descriptives Part and partial con Co <u>l</u> linearity diagnos	
	smoked per day by m	* Si * <u>C</u> a	Method: glection Variable: use Labels: LS Weight: Reset Cancel Help	Enter Rule			Casewise of Qutliers of All cases	-	standard de	viations
					Coefficients ^a					6 545
	Model		Unstandardize B	d Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confide Lower Bound	nce Interval for B Upper Bound	2 372 9 491 Ithly
	1 (Co	nstant)	60.461	1.400		43.182	<.001	57.631	63.291	de
	Mat	ernal age	318	.048	727	-6.691	<.001	414	222	
	a. Depen	dent Variab	le: Length of bab	y (cm)						

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

Length of baby vs Mother's height

	tinear Regression							×
			2	Dependent:			[automation]	
	Haby ID [ID]			Length of baby (cr	n) [Length]		Statistics	
		ence (cm) [Headcirc]	lock	1 of 1			Plots	_
	Birthweight (kg)		Pres	vious			lext Save	
	Maternal age [n	e at birth (weeks) [Gest	ation	Block 1 of 1			Options.	
	smoker	nagej		Maternal height (c	m) [mheight]		Style	
	Father's age [fa Years father was	egnancy weight (kg) [n ge] is in education [fedyrs] rettes smoked per day (cm) [fheight]	•				Bootstrap	
	Mother aged ov Number of cigat	er 35 [mage35] rettes smoked per day	by mot	Method:	Enter		~	
			\$	Selection Variable:		R <u>u</u> le		1 4
	26		\$	<u>C</u> ase Labels:				
	1		\$	WLS Weight:				1 30
				te Reset Cancel	Help			
	8	-6	X	68	18	59 31	2	
			Co	oefficients ^a				
		Unstandardize	d Coefficients	Standardized Coefficients			95.0% Confide	ence Interval for B
Model		В	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
	Constant)	15.334	10.271		1.493	143	-5.425	36.093

a. Dependent Variable: Length of baby (cm)

.219

Maternal height (cm)

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

.062

.485

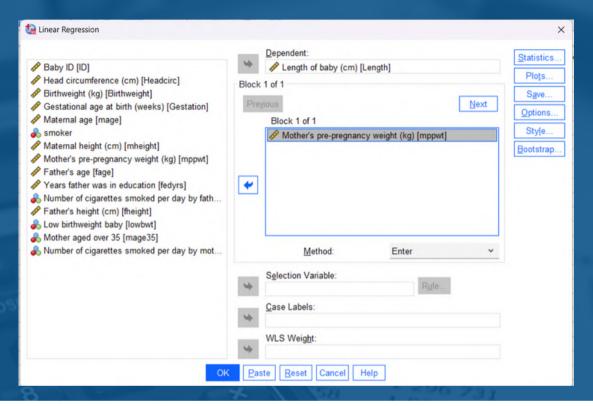
3.507

.001

.093

.345

Length of baby vs Mother's weight



Coefficients^a

		Unstandardize	d Coefficients	Standardized Coefficients			95.0% Confider	ice Interval for B
Model		В	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	41.996	3.427		12.255	<.001	35.070	48.922
	Mother's pre-pregnancy weight (kg)	.162	.059	.398	2.745	.009	.043	.282

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% Cl)	<i>p</i> -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 <u>preliminary main effect model</u>.

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.

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.

STEP 5: CHECKING ASSUMPTIONS

Assumptions	How to check?
1. <u>Independent</u> observation	Done during design stage
2.0verall <u>linearity</u>	Scatter plot between residuals and predicted values (XP - YR)
3.Homoscedasticity (<u>Equal variances</u>)	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

STEP 6: INTERPRETATION AND PRESENTATION

			Co	efficients ^a				
Unstandardized Coefficients			Standardized Coefficients			95.0% Confider	ice Interval for B	
Model		В	Std. Error	Beta	t	Sig.	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 (n=42)

Variable	Simple Linear Regression		Multiple Linear Regression		
	bª (95% Cl)	<i>p</i> -value	b ^ь (95% CI)	<i>p</i> -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	-	-	

Start at monthly

^a Crude regression coefficient

^bAdjusted regression coefficient

All model assumptions are fulfilled.

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

Coefficient of determination (R²) = 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 = 0.621).



0 8 8 8 8 B

100.000.000

REGRESSION ANALYSIS

DEPENDENT VARIABLE	INDEPENDENT VARIABLE	STATISTICAL TEST	
Numerical	Numerical	Multiple Linear Regression	
Categorical (dichotomous)	Numerical and categorical	Multiple/Binary Logistic Regression	
Categorical (polytomous - nominal)	Numerical and categorical	Multinomial Logistic Regression	
Categorical (ordinal)	Numerical and categorical	Ordinal Logistic Regression	



MDM NURULJANNAH BT NOR AZMI

EMAIL: nuruljannah@mahsa.edu.my

THANK YOU