Multivariate Analysis: The Introduction

Jamalludin Ab Rahman MD MPH
Associate Professor & Head
Department of Community Medicine
Kulliyyah (Faculty) of Medicine

www.iiumedic.net/biostatistics

My assumptions

You already have the following basic knowledge:

- Basic computing skill
- Basic biostatistics knowledge
- Basic SPSS knowledge

What is "Basic" SPSS skill

Able to:

- Create dataset & enter data into SPSS
- Filter, compute & recode data
- Check for Normality of data
- Do univariate or bivariate analysis
- Do correlation

In other words, you should these tests

- Check for Normality
- t-test
- χ^2 test
- One-way ANOVA
- Kruskall-Wallis
- Mann-Whitney U Test
- Pearson/Spearman Correlation

…and therefore, I won't cover them at all but I expect you to be using them!

Content

Day 1

- Introduction to causal inference & multivariate analysis
- Preparing data for multivariate analysis
- Linear Regression

Day 2

- GLM – Univariate
- GLM - Multivariate
- GLM – Repeated measure

Day 3

- Logistic Regression

Biostatistics for Medical Researchers (Advanced)

THE CONCEPTS

Basic I - Variance

The variance of a random variable is one measure of statistical dispersion, averaging the squared distance of its possible values from the expected value.

Basic II - Covariance

- Measures of how much **two variables vary together**
- The covariance becomes **more positive** for each pair of values which differ from their mean in the **same direction**
- The covariance becomes **more negative** with each pair of values which differ from their mean in **opposite directions**
- If the values are independent, then their covariance is zero *but the converse, however, is not true i.e. if the pair of values have covariance zero, they need not be independent.*

Basic III - Correlation

- The strength and direction of a linear relationship between two variables
- Standardized covariance

$$\rho X,Y = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y}$$

Basic IV – Conditional Independence

- Two events, A and B, are said to be independent if the two events do not affect one another
  i.e. $P(A/B) = P(A)$ and $P(B/A) = P(B)$
- Conditional independence occurred when there is the 3rd event

Understanding **correlation** is very important in understanding the concept underlying multivariate analysis!!

Biostatistics for Medical Researchers (Advanced)

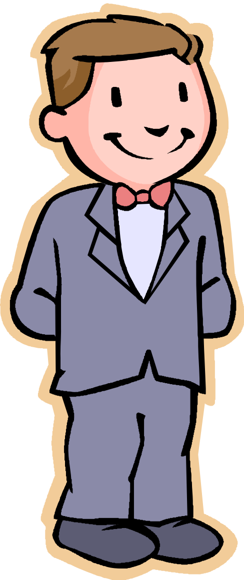
CAUSAL INFERENCE

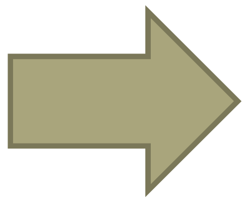
The past

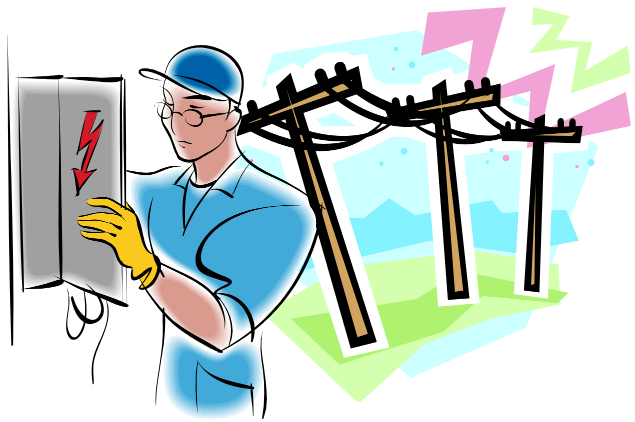
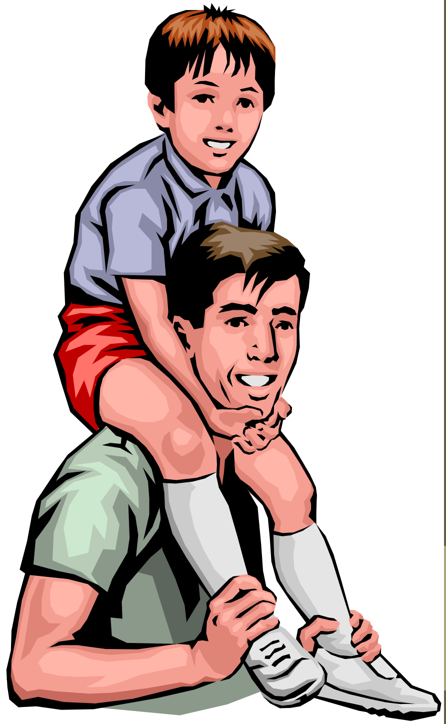
- **Miasma theory** – 18th century - disease due to 'bad air' or contamination from poor sanitation
- **Germ theory** – 19th century - one-disease-one-cause
- **One-to-many** – 20th century – e.g. smoking causing many diseases (discovered by Ernst Wynder)

Causal – the classic example

- A boy learning that moving a light switch causes the light to turn on



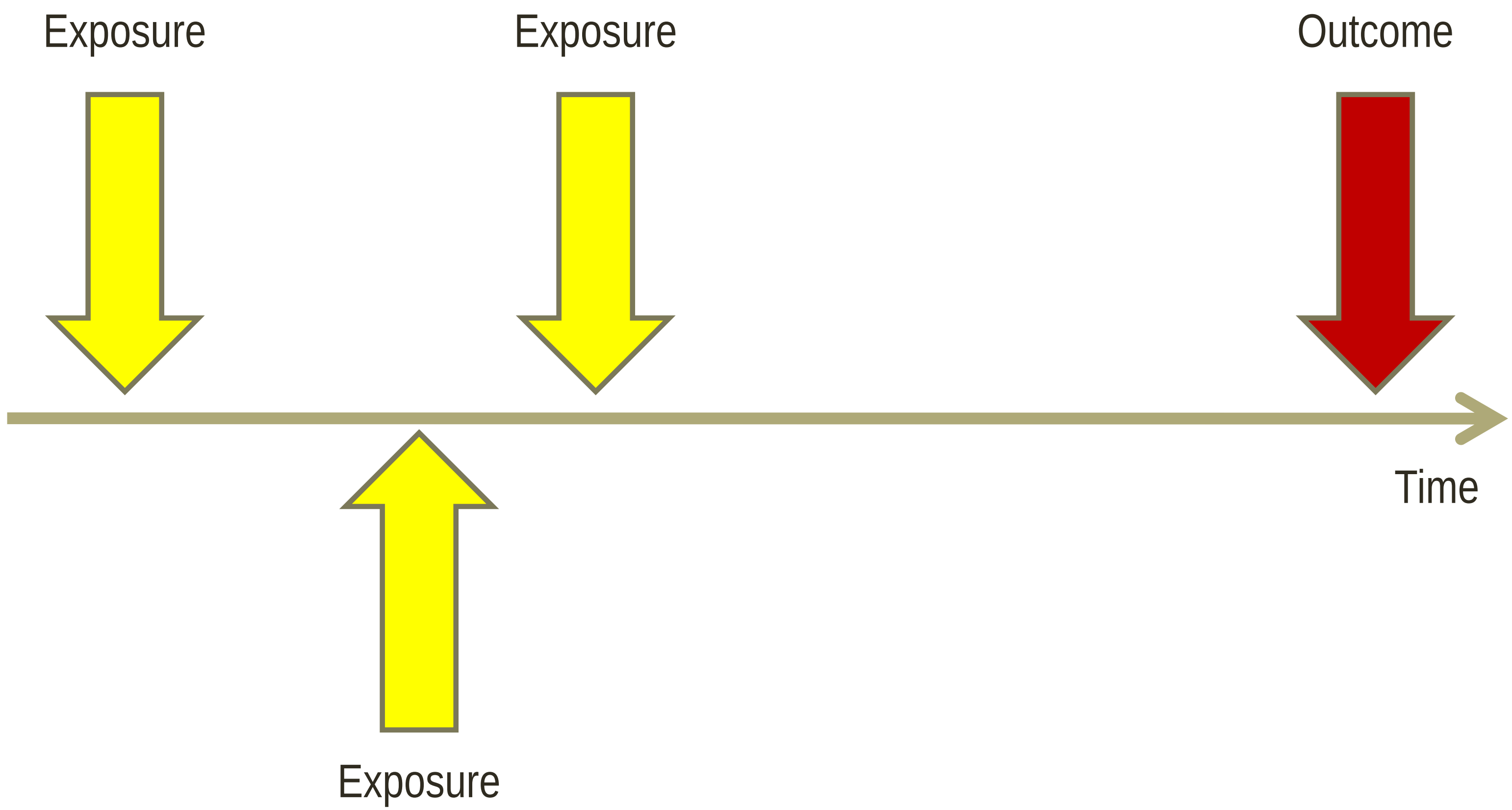


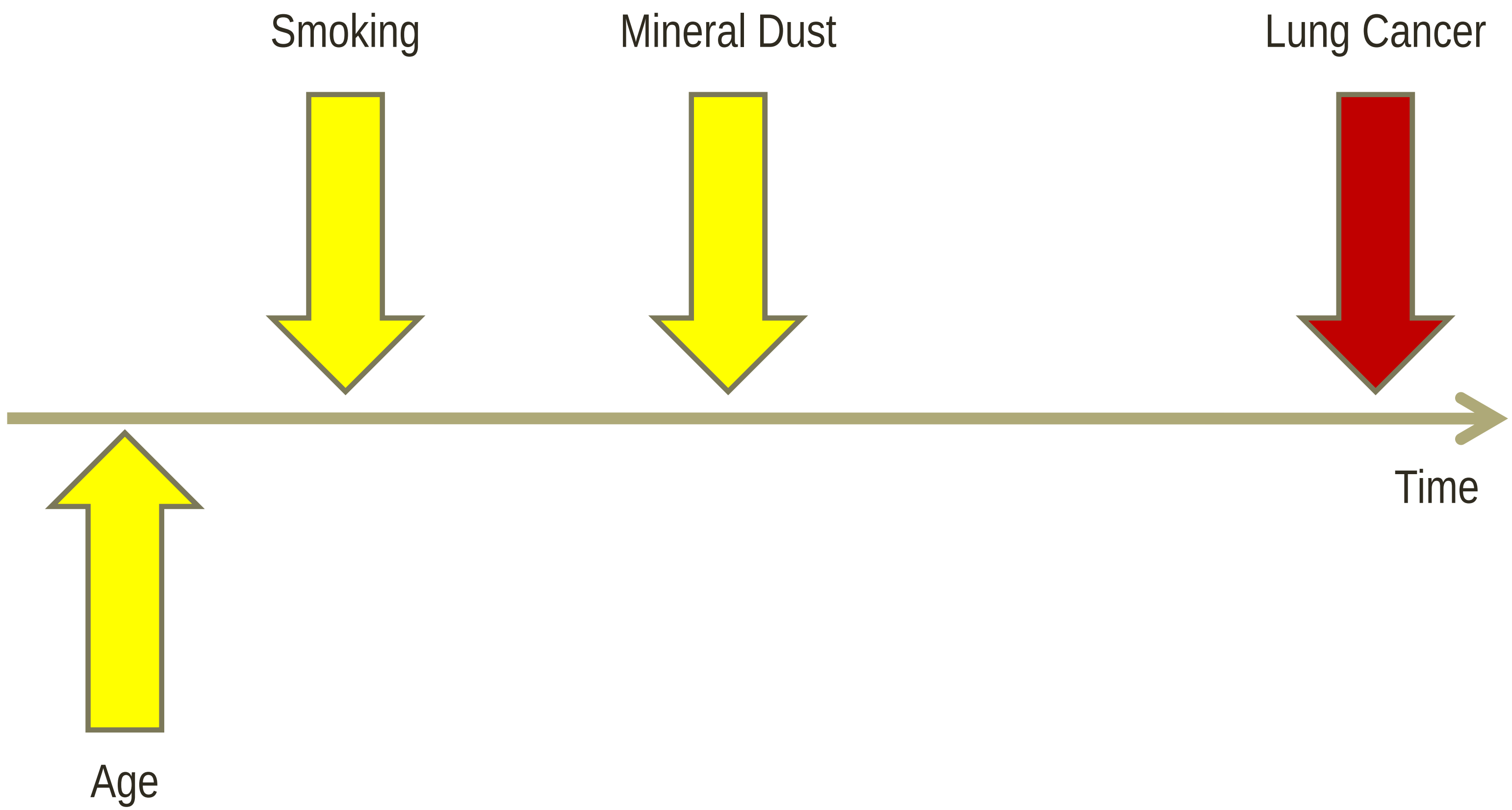
Direct observation vs. Inference

- Tuberculosis – CXR vs. AFB vs. DNA
- HIV – Antibody vs. Antigen
- Hypertension – Peripheral BP vs. Central BP
- Depression – Psychiatric assessment vs. biomarker
- Obesity – BMI vs. Waist vs. Fat%
- Smoking behavior – social influence vs. genetic?

Disease model



Disease model (example)



Bivariate Disease Model

Smoking
High cholesterol diet
Promiscuity

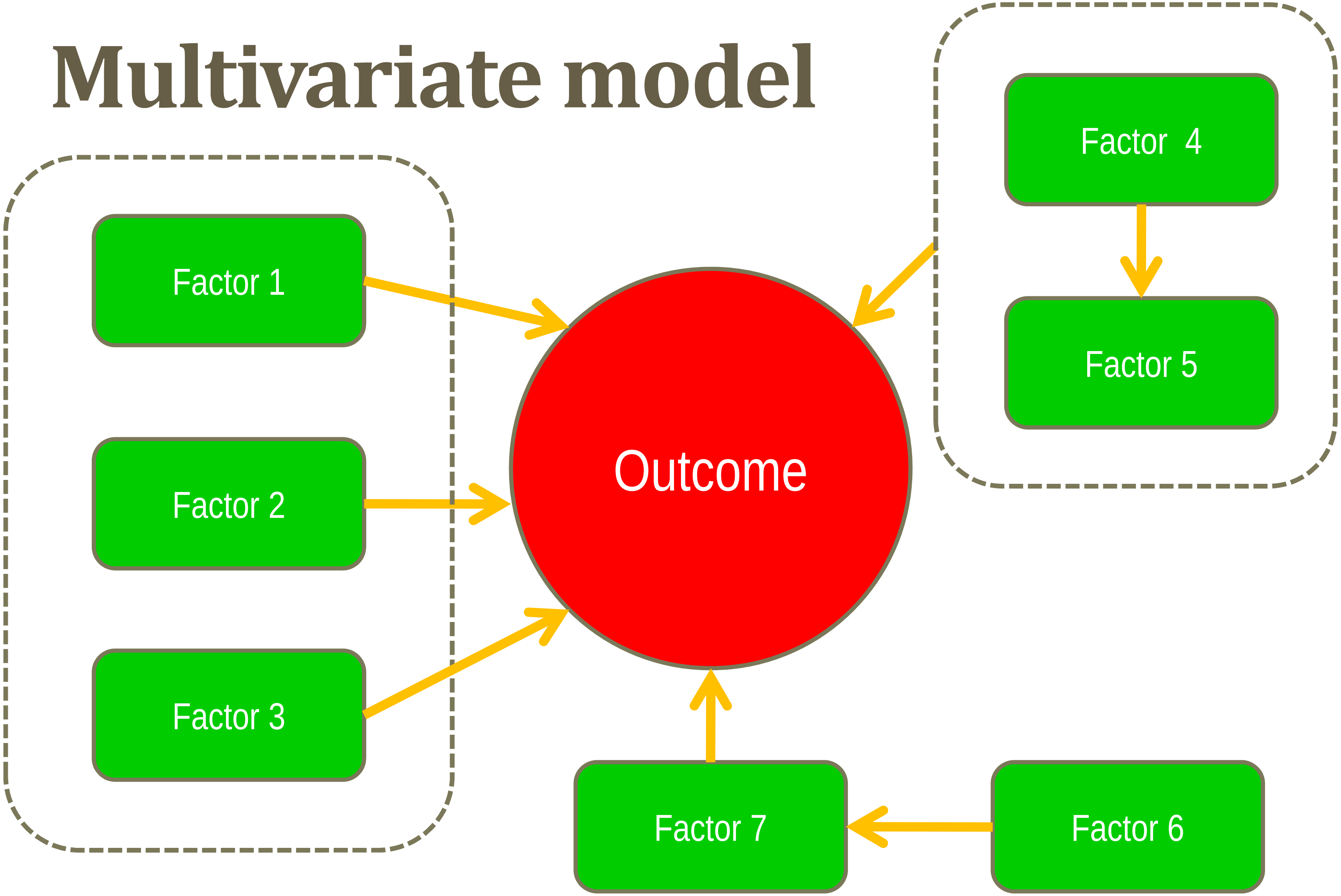
Lung Cancer
CVD
HIV

“the use of alcohol-containing mouthwashes & oro-pharyngeal cancer” (Shapiro et al 1996)

Shapiro, S., Castellana, J. V., & Sprafka, J. M. (1996). Alcohol-containing Mouthwashes and Oropharyngeal Cancer: A Spurious Association due to Underascertainment of Confounders? (Vol. 144, pp. 1091-1095): Oxford Univ Press.

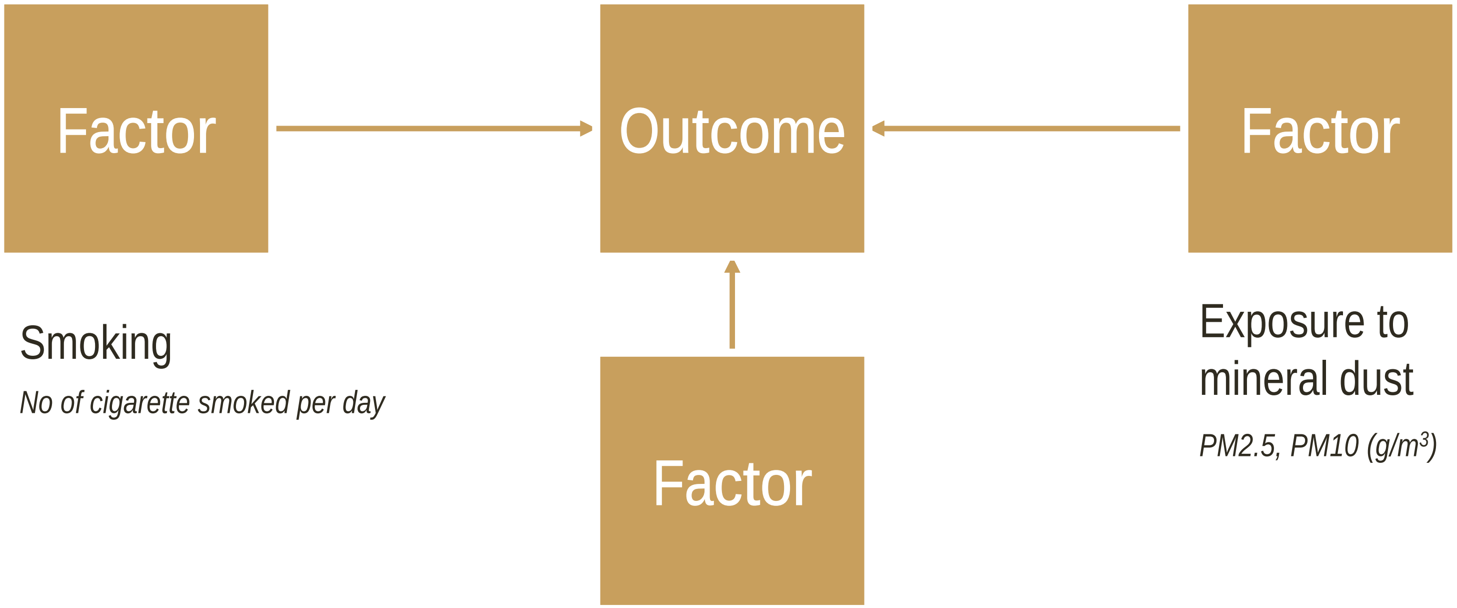
The 3rd ~~person~~ factor???
Jamalludin Ab Rahman 8 Oct. 2012

Multivariate model



Multivariate Model

Lung Cancer



Smoking

No of cigarette smoked per day

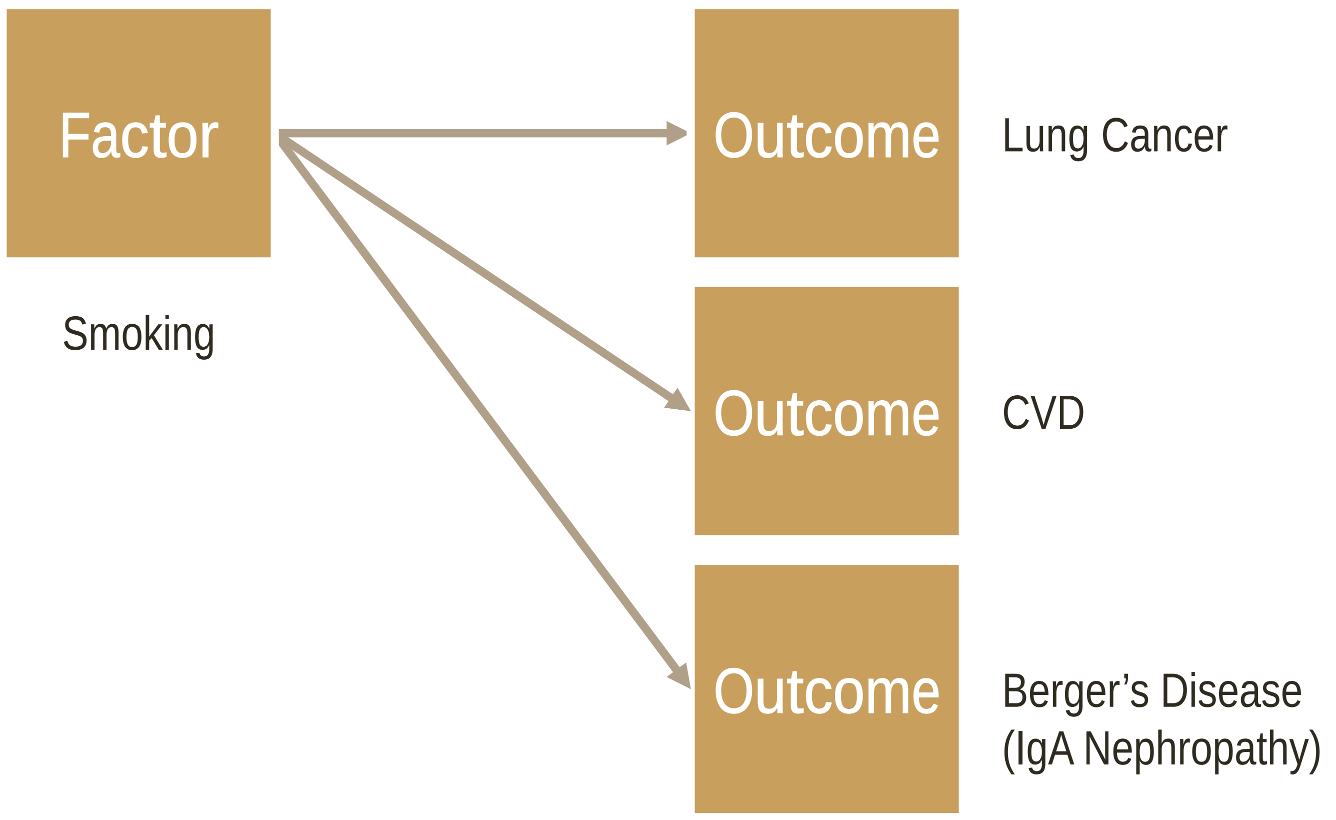
Exposure to mineral dust

PM2.5, PM10 (g/m^3)

Radiation

Radiation Absorbed dose (mGy) per day

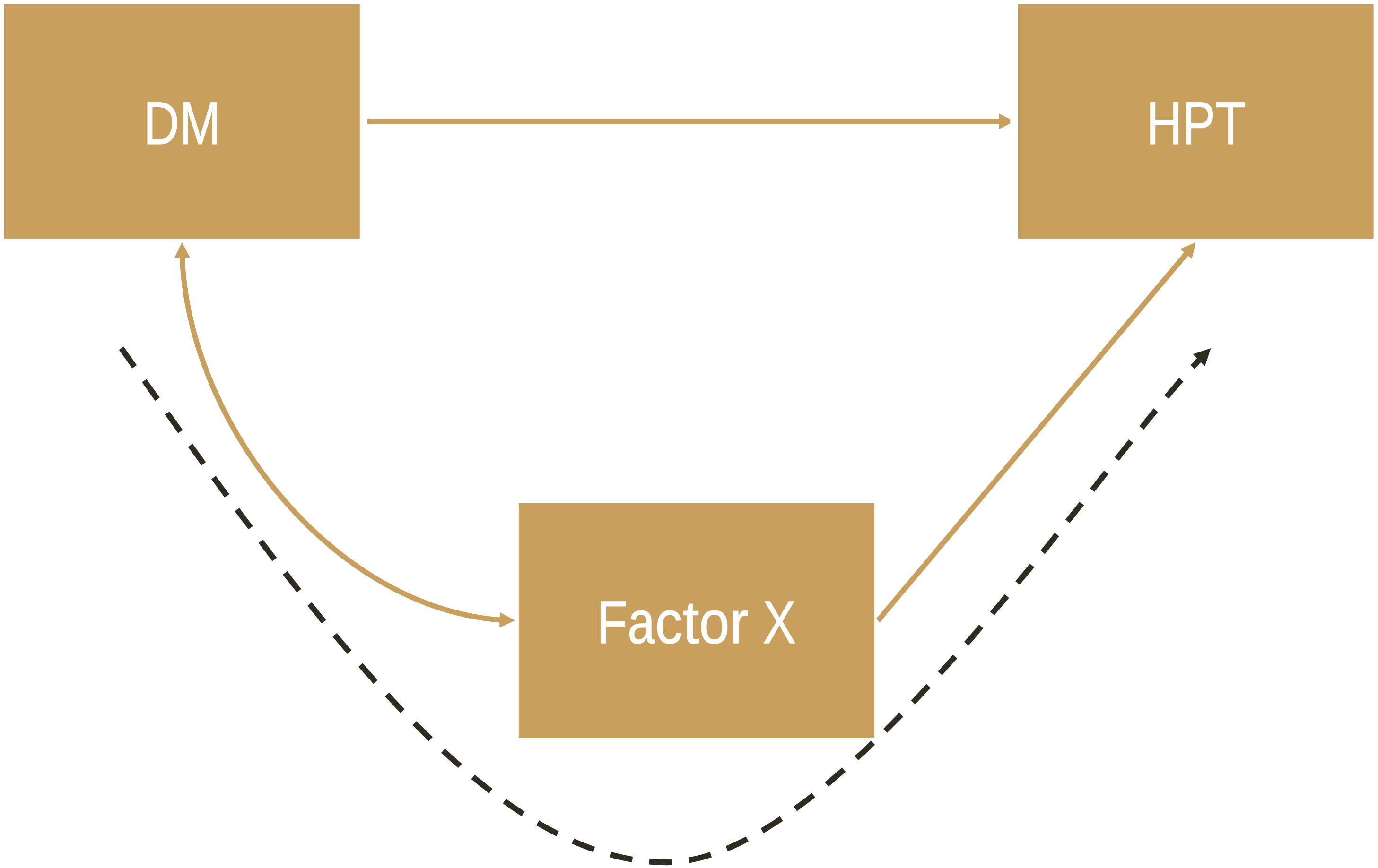
Multiple Outcomes



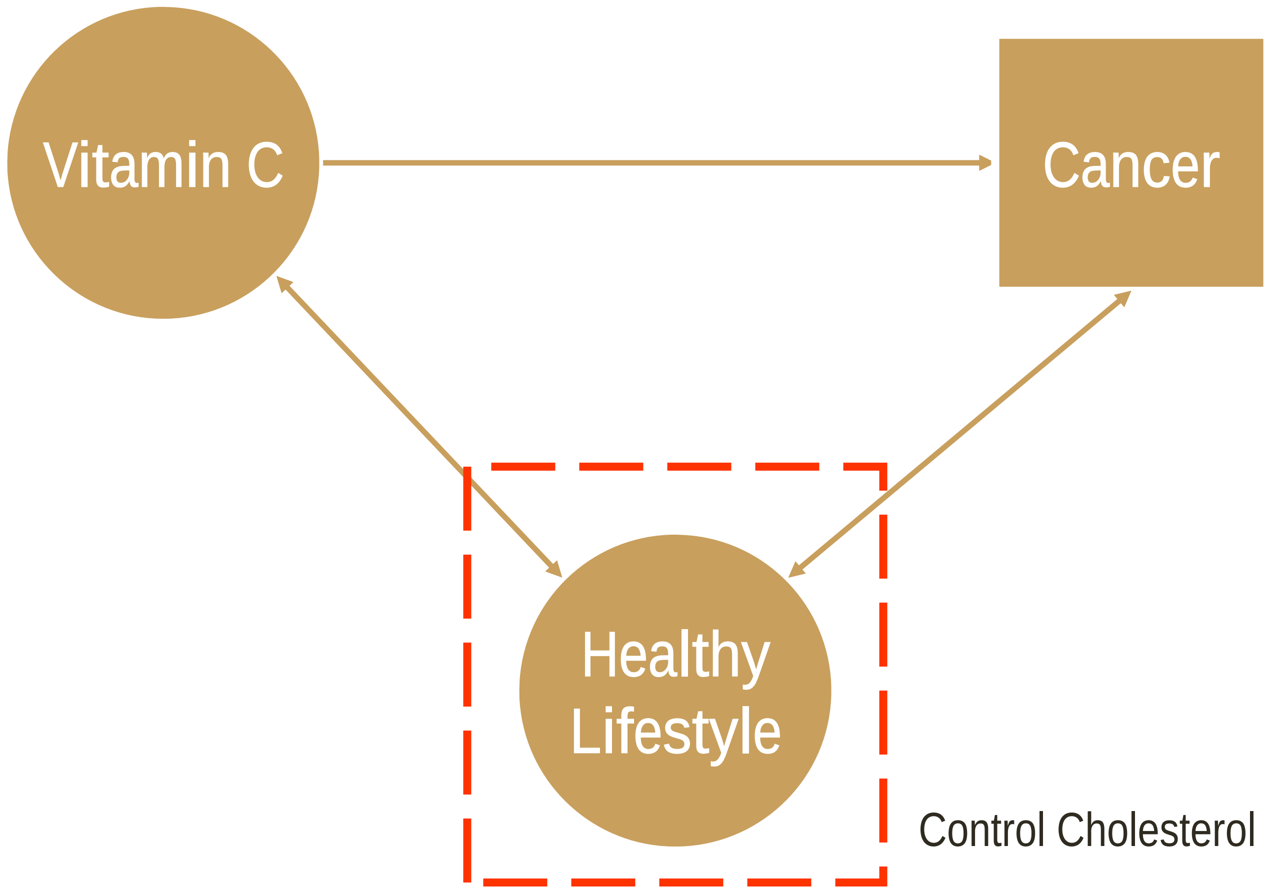
About relationship....

- Diabetes is associated with hypertension.
- Diabetes causing hypertension? or.....
- Hypertension causing diabetes?

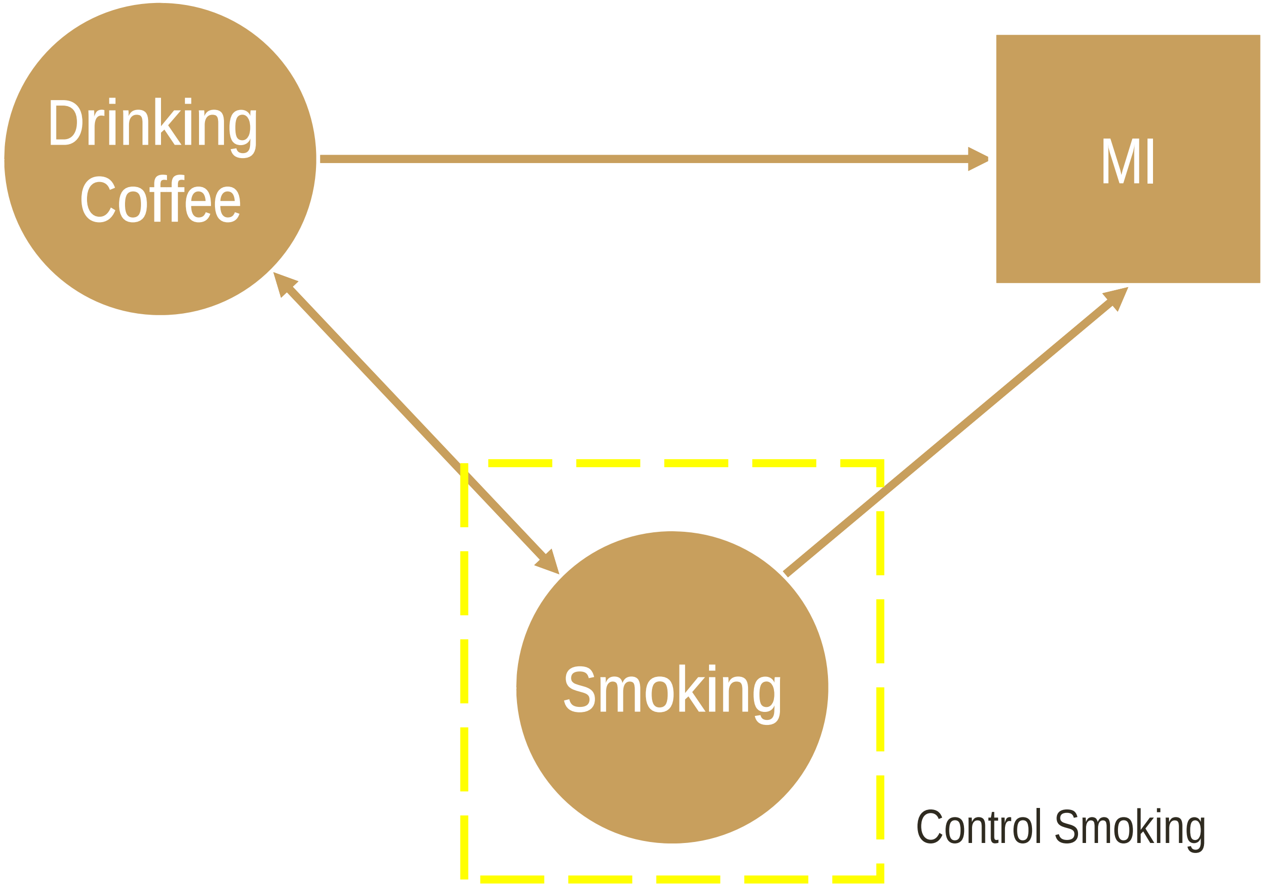
Direct Effect + Indirect Effect = Total Effect



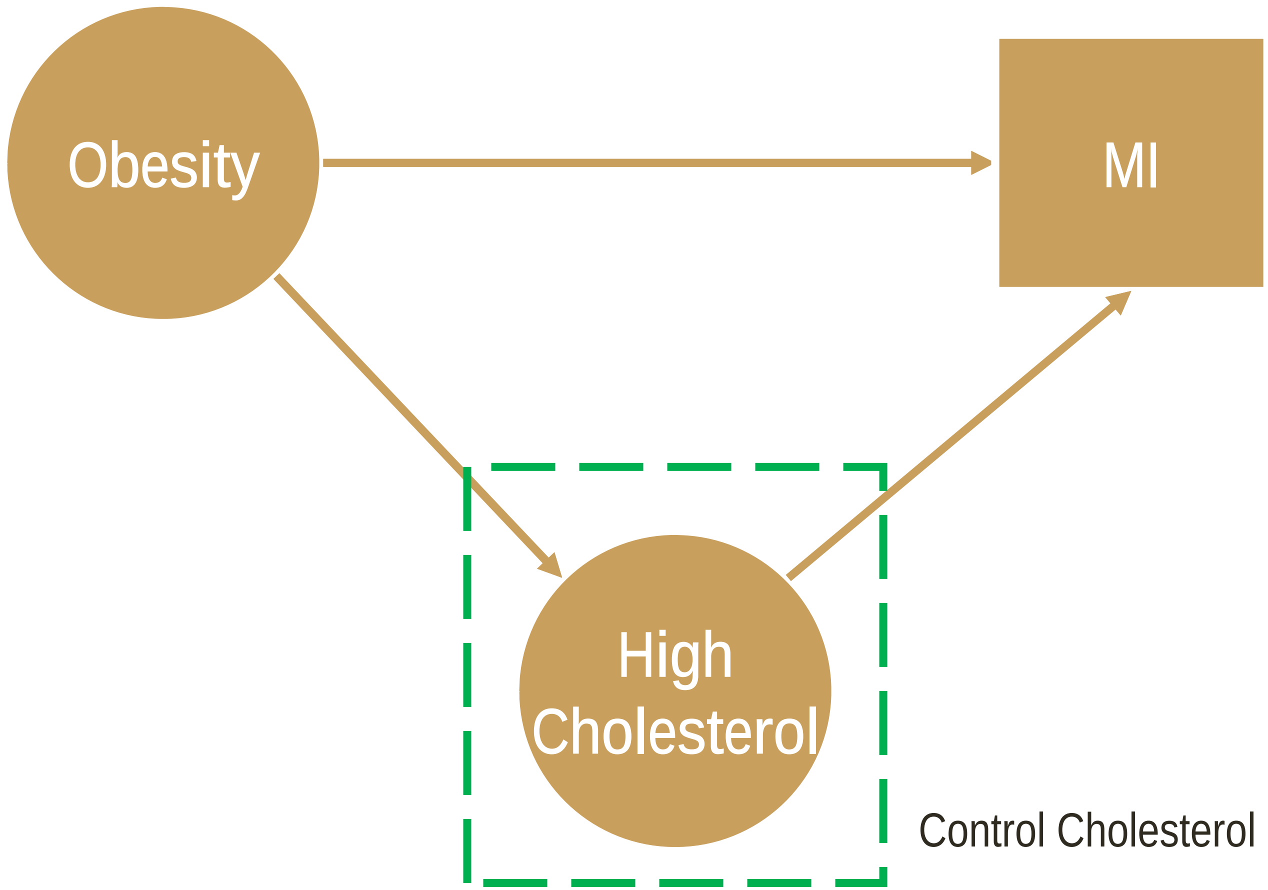
Is vitamin C really prevent cancer?



Can coffee drinking lead to MI?



Is obesity independently leads to MI?



The 3rd factors can be…

- A confounder
- A moderator or effect modifier (interaction)
- A mediating or intervening factor (mediator)

Confounder & Mediator

- Confounder influence a relationship (between two variables) but it is **not a part of the pathway**
- Mediator influence a relationship (between two variables) and it is **also a part of the pathway**

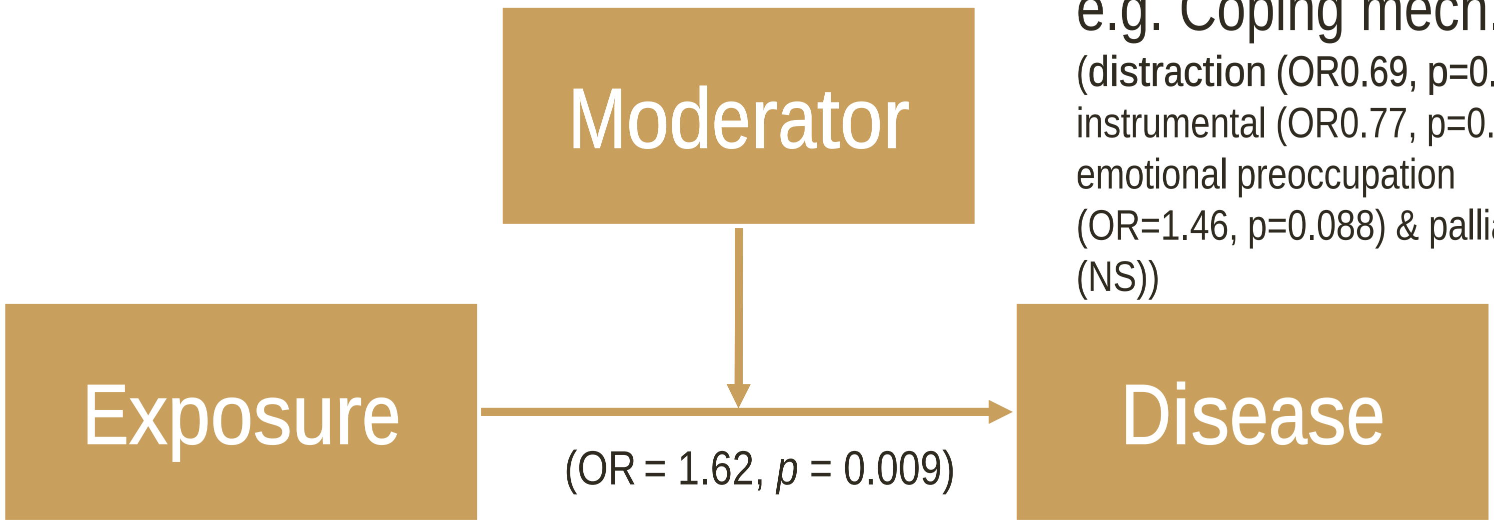
We can't differentiate a confounder from a mediator statistically. It must be made from the ***scientific basis****! Remember science always precedes statistics. Clinical significance precedes statistical significance!*

Moderator

- When an exposure has **different effects on disease at different values** of a variable, that variable is called a effect modifier (or there is 'interaction' happening)

Moderating Effect

e.g. Coping mech.
(distraction (OR0.69, p=0.009), instrumental (OR0.77, p=0.081), emotional preoccupation (OR=1.46, p=0.088) & palliative (NS))



e.g. Stress

e.g. Exacerbation of MS

Mohr, D. C., Goodkin, D. E., Nelson, S., Cox, D., & Weiner, M. (2002). Moderating Effects of Coping on the Relationship Between Stress and the Development of New Brain Lesions in Multiple Sclerosis. *Psychosom Med, 64(5), 803-809.*

Why multivariate?

1. Multi-factorial
   – *which are the significant factors?*
2. Multiple outcomes
3. Multiple unit of measurements
4. Exploration of associations

What is multivariate analysis?

- An analysis of the relationships between more than two variables
- If y = Outcome, x1 = Factor #1, x2 = Factor #2

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

- ε = Error term, Residual

Multivariate notation

Coefficient for Var x_1

Dependent Var

$$Y = \beta o + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \varepsilon$$

Error/Residual

Intercept

Explanatory Var x_1

Example #1

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

- Arterial BP = Constant + Age + Body weight + Pulse rate + Stress + Residual

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.994[a]	.987	.984	.793	1.744

a. Predictors: (Constant), Stress Score, Weight, Pulse Rate, Age

b. Dependent Variable: Arterial Blood Pressure

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	735.108	4	183.777	291.948	.000[a]
	Residual	9.442	15	.629		
	Total	744.550	19			

a. Predictors: (Constant), Stress Score, Weight, Pulse Rate, Age

b. Dependent Variable: Arterial Blood Pressure

$R^2 = 0.986$, meaning 99% of variation in ABP is explained by Age, Body Weight, Pulse Rate & Stress ($F(4,15)=291.948$), $P<0.001$)

Main Result

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	17.308	4.623		3.744	.002	7.454	27.163
	Age	.573	.091	.374	6.285	.000	.379	.767
	Weight	.892	.070	.626	12.756	.000	.743	1.040
	Pulse Rate	.087	.078	.053	1.126	.278	-.078	.253
	Stress Score	.005	.007	.025	.662	.518	-.011	.021

a. Dependent Variable: Arterial Blood Pressure

Arterial BP = *17.3* + 0.6(Age) + 0.9(Body weight) + 0.09(Pulse rate) + 0.01(Stress)

Example #2

- Snoring and risk of cardiovascular disease in women . Hu 2000. From The Nurses' Health Study. Cohort. Baseline, N=71,779 women 40 to 65 years old and without diagnosed CVD or cancer in 1986. Till 31st May 1994.
- CVD = Snoring + Age +Smoking + BMI + Alcohol + Physical Activity + Menopausal status + Family history of MI + DM + Cholesterol + Hours sleeping + Sleeping position

	Frequency of Snoring		
	Never	Occasionally	Regularly
Total cardiovascular events (coronary heart disease + stroke)* *			
# cases	162	729	151
Person-years	143,719	356,530	51,292
Age-adjusted	1.0	1.46 (1.23–1.74)	2.02 (1.62–2.53)
Multivariate	1.0	1.20 (1.01–1.43)	1.33 (1.06–1.67)

Conclusion

- Nowadays, with the advancement of technology, multivariate analysis is the most appropriate approach in most statistical analysis & more importantly it is do-able.
- Therefore it should not be considered as something too advanced.
- The days of bi-variable analysis are numbered!

Approach to Multivariate Analysis

- Based on research question
  - Significant of differences between groups
  - Degree of relationship – correlation
  - Prediction
  - Explore structure
- Based on model (modeling)
  - Theory driven – Hypothesis deduction
  - Data driven – Data reduction

Preparation of Data

1. 'Clean' the data – from missing values, outliers (for both univariate & multivariate)
2. Check for Normal distribution – for numerical data (both univariate & multivariate)
3. Check for linear relationship
4. Check for homoscedasticity (homogeneity) – the data (covariances) are homogenous or heterogeneous
5. Check for multi-collinearity

Type of Statistical Tests

Dependent Variables	Independent Variables	Test
1 – Cont	≥ 2 – All Cont	Linear Regression
1 – Cont	≥ 2 – All Cat	ANOVA
1 – Cont	≥ 2 – Cont + Cat	ANCOVA
> 1 – Cont	All Cat	MANOVA
> 1 – Cont	Cat + Cont	MANCOVA
1 – Dichotomous	≥ 2 – Cont + Cat	Binary Logistic Regression

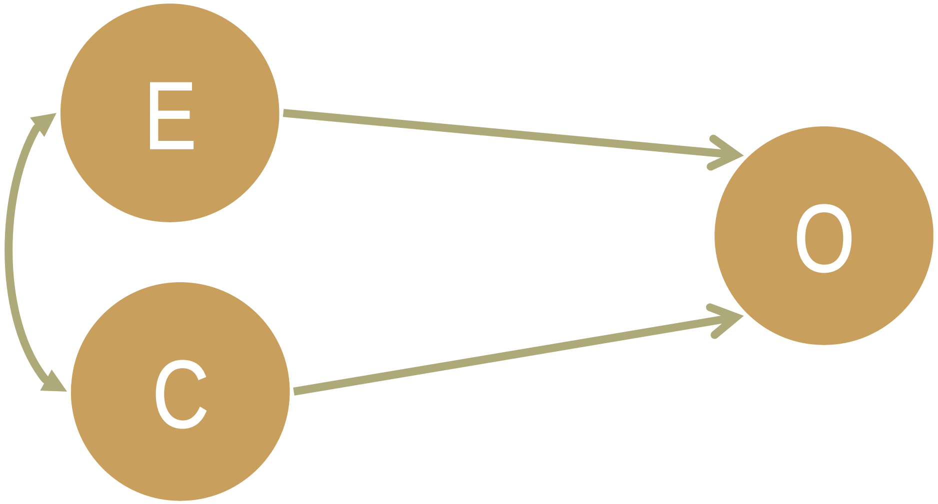
Identifying 3rd Factor

Jamalludin Ab Rahman MD MPH
Department of Community Medicine
International Islamic University Malaysia

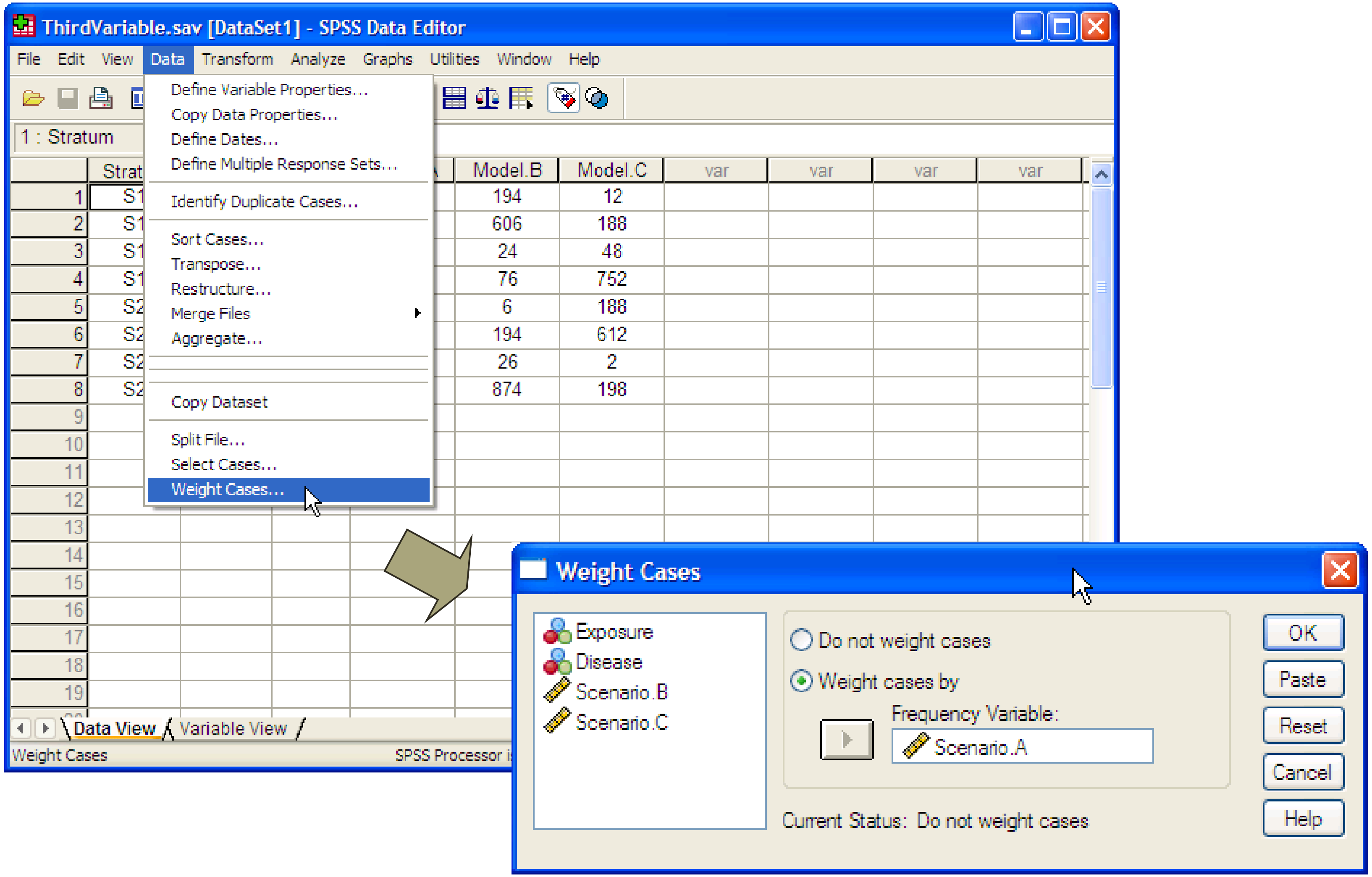
The exercise data

Please download datasets from www.iiumedic.net/biostatistics

Characteristics of a confounder



1. It is a **risk factor** for the disease
2. It is **associated** with risk factor, **independent** of the risk factor.
3. It is not in the causal pathway between exposure and disease.


ThirdVariable.sav [DataSet1] - SPSS Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Window
Help
1 : Stratum
Define Variable Properties...
Copy Data Properties...
Define Dates...
Define Multiple Response Sets...
Identify Duplicate Cases...
Sort Cases...
Transpose...
Restructure...
Merge Files
Aggregate...
Copy Dataset
Split File...
Select Cases...
Weight Cases...
Model.B
Model.C
var
194 12
606 188
24 48
76 752
6 188
194 612
26 2
874 198
Data View
Variable View
Weight Cases
Weight Cases
Exposure
Disease
Scenario.B
Scenario.C
Do not weight cases
Weight cases by
Frequency Variable:
Scenario.A
OK
Paste
Reset
Cancel
Help
Current Status: Do not weight cases

Exposure * Disease Crosstabulation

			Disease		Total
			D+	D-	
Exposure	E+	Count	200	800	1000
		% within Exposure	20.0%	80.0%	100.0%
	E-	Count	50	950	1000
		% within Exposure	5.0%	95.0%	100.0%
Total		Count	250	1750	2000
		% within Exposure	12.5%	87.5%	100.0%

Chi-Square Tests

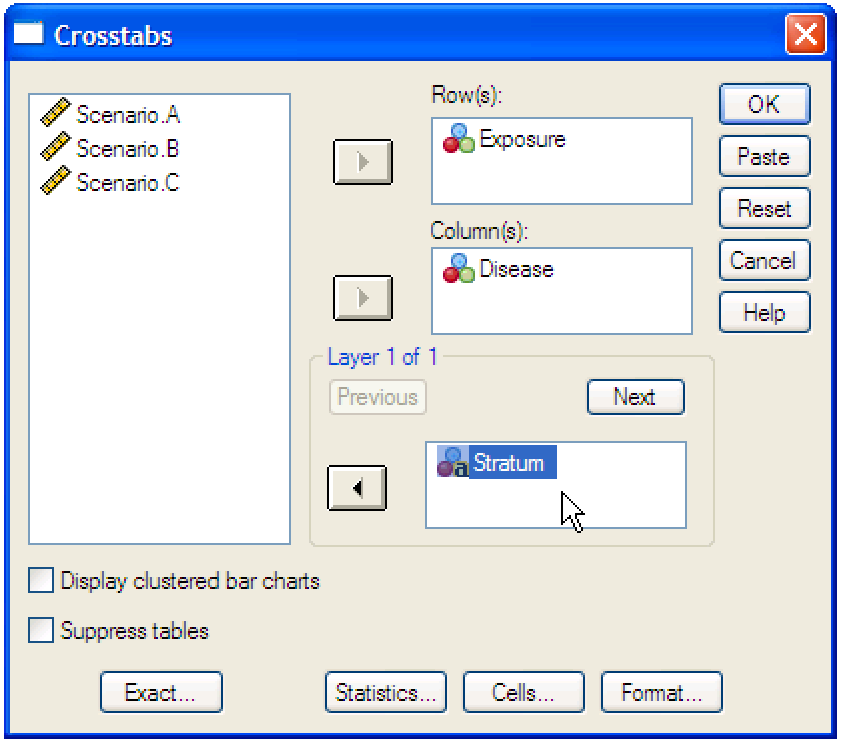
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	102.857[b]	1	.000		
Continuity Correction[a]	101.490	1	.000		
Likelihood Ratio	109.245	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	102.806	1	.000		
N of Valid Cases	2000				

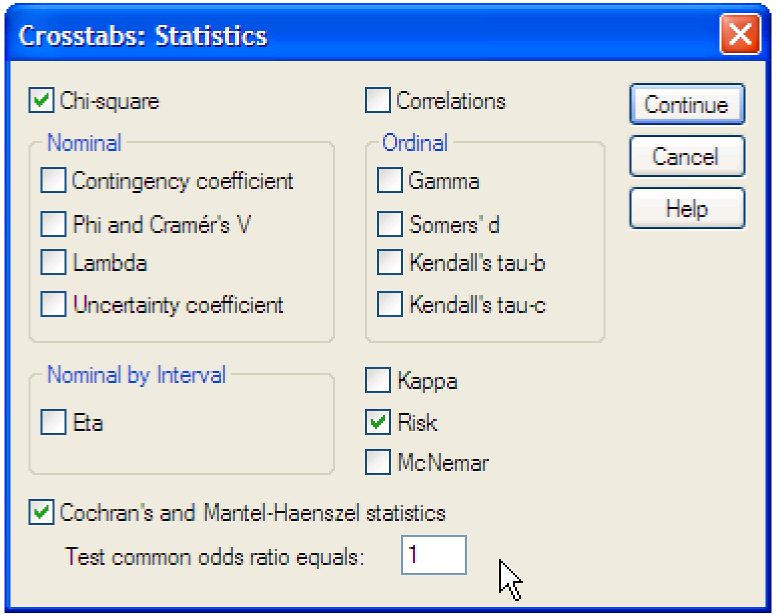
a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 125.00.

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Exposure (E+ / E-)	4.750	3.436	6.567
For cohort Disease = D+	4.000	2.971	5.385
For cohort Disease = D-	.842	.814	.871
N of Valid Cases	2000		


Crosstabs
Scenario.A
Scenario.B
Scenario.C
Row(s):
Exposure
Column(s):
Disease
OK
Paste
Reset
Cancel
Help
Layer 1 of 1
Previous
Next
Stratum
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
Format...


Crosstabs: Statistics
Chi-square
Correlations
Continue
Cancel
Help
Nominal
Contingency coefficient
Phi and Cramér's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals:
1

Exposure * Disease * Stratum Crosstabulation

Stratum				Disease D+	Disease D-	Total
S1	Exposure	E+	Count	160	240	400
			% within Exposure	40.0%	60.0%	100.0%
		E-	Count	40	360	400
			% within Exposure	10.0%	90.0%	100.0%
	Total		Count	200	600	800
			% within Exposure	25.0%	75.0%	100.0%
S2	Exposure	E+	Count	40	560	600
			% within Exposure	6.7%	93.3%	100.0%
		E-	Count	10	590	600
			% within Exposure	1.7%	98.3%	100.0%
	Total		Count	50	1150	1200
			% within Exposure	4.2%	95.8%	100.0%

Chi-Square Tests

Stratum		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
S1	Pearson Chi-Square	96.000[b]	1	.000		
	Continuity Correction[a]	94.407	1	.000		
	Likelihood Ratio	101.261	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	95.880	1	.000		
	N of Valid Cases	800				
S2	Pearson Chi-Square	18.783[c]	1	.000		
	Continuity Correction[a]	17.551	1	.000		
	Likelihood Ratio	20.057	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	18.767	1	.000		
	N of Valid Cases	1200				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 100.00.

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 25.00.

Risk Estimate

Stratum		Value	95% Confidence Interval Lower	95% Confidence Interval Upper
S1	Odds Ratio for Exposure (E+ / E-)	6.000	4.091	8.800
	For cohort Disease = D+	4.000	2.912	5.495
	For cohort Disease = D-	.667	.611	.727
	N of Valid Cases	800		
S2	Odds Ratio for Exposure (E+ / E-)	4.214	2.087	8.508
	For cohort Disease = D+	4.000	2.019	7.924
	For cohort Disease = D-	.949	.927	.972
	N of Valid Cases	1200		

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	.753	1	.385
Tarone's	.748	1	.387

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	113.684	1	.000
Mantel-Haenszel	112.045	1	.000

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			5.500
ln(Estimate)			1.705
Std. Error of ln(Estimate)			.172
Asymp. Sig. (2-sided)			.000
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	3.926
		Upper Bound	7.706
	ln(Common Odds Ratio)	Lower Bound	1.367
		Upper Bound	2.042

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

Exercise 1

1. Use file Data1_3rdfactor.sav, weigh for Scenario.B & check for 3rd factor
2. Repeat Q1 above but weigh for Scenario.C

Answers – Scenario B

Risk Estimate

Stratum		Value	95% Confidence Interval Lower	Upper
S1	Odds Ratio for Exposure (E+ / E-)	1.014	.623	1.649
	For cohort Disease = D+	1.010	.698	1.462
	For cohort Disease = D-	.997	.887	1.120
	N of Valid Cases	900		
S2	Odds Ratio for Exposure (E+ / E-)	1.040	.422	2.560
	For cohort Disease = D+	1.038	.433	2.490
	For cohort Disease = D-	.999	.972	1.026
	N of Valid Cases	1100		

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	.008	1	.930
Mantel-Haenszel	.000	1	.983

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	.002	1	.961
Tarone's	.002	1	.961

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			1.019
ln(Estimate)			.019
Std. Error of ln(Estimate)			.219
Asymp. Sig. (2-sided)			.930
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	.664
		Upper Bound	1.565
	ln(Common Odds Ratio)	Lower Bound	-.409
		Upper Bound	.448

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

Answer – Scenario C

Risk Estimate

Stratum		Value	95% Confidence Interval Lower	95% Confidence Interval Upper
S1	Odds Ratio for Exposure (E+ / E-)	1.000	.521	1.920
	For cohort Disease = D+	1.000	.542	1.847
	For cohort Disease = D-	1.000	.962	1.040
	N of Valid Cases	1000		
S2	Odds Ratio for Exposure (E+ / E-)	30.412	7.481	123.627
	For cohort Disease = D+	23.500	5.885	93.842
	For cohort Disease = D-	.773	.742	.805
	N of Valid Cases	1000		

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	38.516	1	.000
Mantel-Haenszel	37.416	1	.000

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	33.299	1	.000
Tarone's	30.611	1	.000

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			4.513
ln(Estimate)			1.507
Std. Error of ln(Estimate)			.230
Asymp. Sig. (2-sided)			.000
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	2.876
		Upper Bound	7.081
	ln(Common Odds Ratio)	Lower Bound	1.056
		Upper Bound	1.957

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

Exercise

- Use mets.sav
- A study of MetS among 205 people
- Study relationship between Age & Gender with Mets
- Proof that Age & Gender are INDEPENDENTLY not related with MetS

Gender vs Mets controlled by PA

Gender * Metabolic Syndrome * Physical Activity Crosstabulation

Physical Activity			
High	Gender	Male	Count
			% with
		Female	Count
			% with
	Total		Count
			% with
Low	Gender	Male	Count
			% with
		Female	Count
			% with
	Total		Count
			% with
Total	Gender	Male	Count
			% with
		Female	Count
			% with
	Total		Count
			% with

Chi-Square Tests

Physical Activity		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
High	Pearson Chi-Square	2.879[c]	1	.090		
	Continuity Correction[b]	.342	1	.559		
	Likelihood Ratio	1.844	1	.174		
	Fisher's Exact Test				.225	.225
	Linear-by-Linear Association	2.830	1	.092		
	N of Valid Cases	59				
Low	Pearson Chi-Square	.011[d]	1	.915		
	Continuity Correction[b]	.000	1	1.000		
	Likelihood Ratio	.011	1	.915		
	Fisher's Exact Test				1.000	.546
	Linear-by-Linear Association	.011	1	.915		
	N of Valid Cases	146				
Total	Pearson Chi-Square	22.833[a]	1	.000		
	Continuity Correction[b]	21.504	1	.000		
	Likelihood Ratio	23.359	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	22.721	1	.000		
	N of Valid Cases	205				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 42.93.

b. Computed only for a 2x2 table

c. 2 cells (50.0%) have expected count less than 5. The minimum expected count is .24.

d. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 11.25.

Risk Estimate

Physical Activity		Value	95% Confidence Interval Lower	Upper
High	Odds Ratio for Gender (Male / Female)	8.500	.469	154.174
	For cohort Metabolic Syndrome = Yes	1.144	.844	1.552
	For cohort Metabolic Syndrome = No	.135	.009	1.919
	N of Valid Cases	59		
Low	Odds Ratio for Gender (Male / Female)	.956	.418	2.188
	For cohort Metabolic Syndrome = Yes	.965	.502	1.856
	For cohort Metabolic Syndrome = No	1.010	.848	1.202
	N of Valid Cases	146		
Total	Odds Ratio for Gender (Male / Female)	4.104	2.269	7.421
	For cohort Metabolic Syndrome = Yes	2.271	1.573	3.278
		.553	.428	.716
		205		

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	2.686	1	.101
Tarone's	2.686	1	.101

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	.044	1	.833
Mantel-Haenszel	.000	1	.997

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			1.087
ln(Estimate)			.083
Std. Error of ln(Estimate)			.403
Asymp. Sig. (2-sided)			.836
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	.493
		Upper Bound	2.396
	ln(Common Odds Ratio)	Lower Bound	-.707
		Upper Bound	.874

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

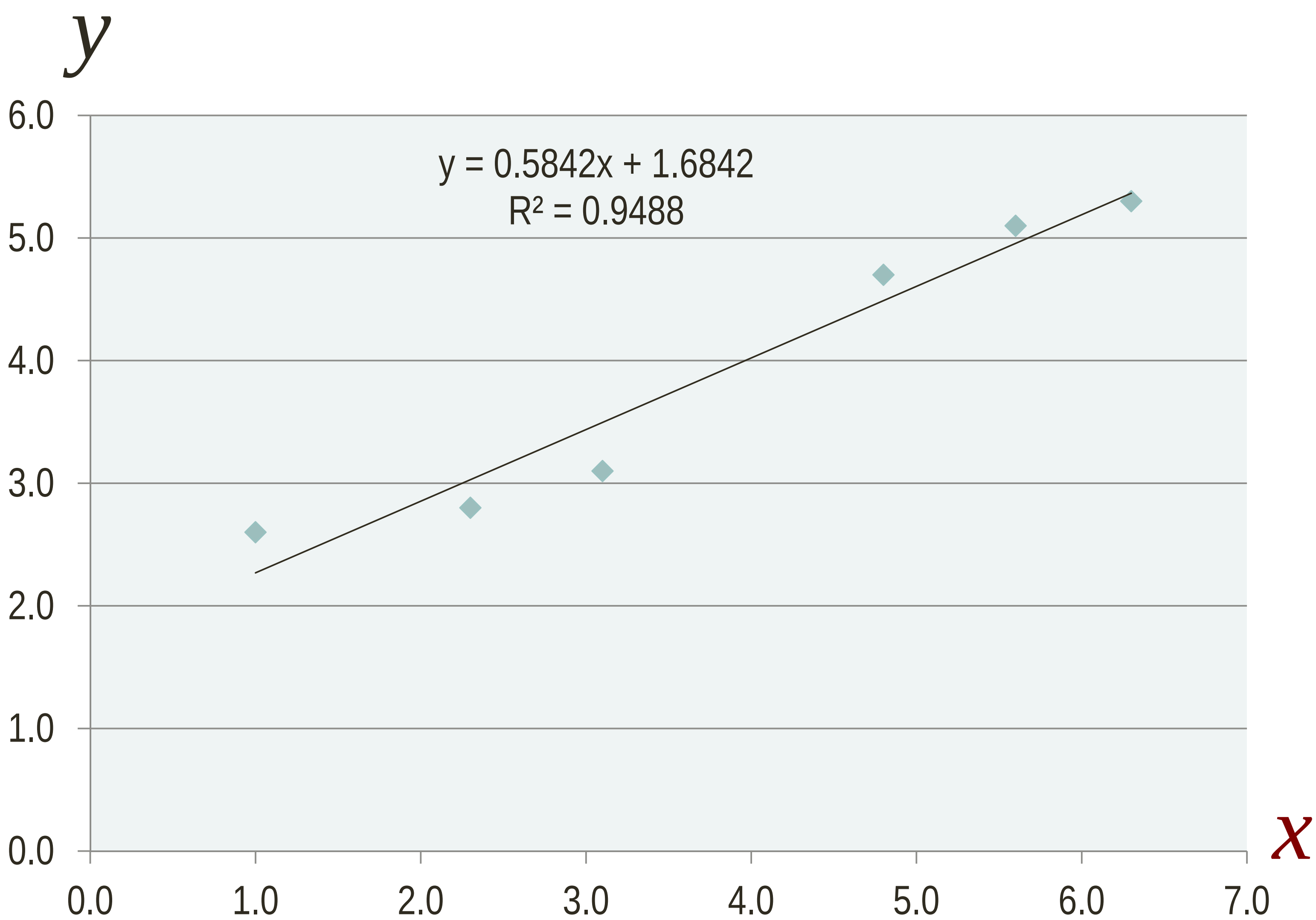
Regression Analysis

Regression Analysis

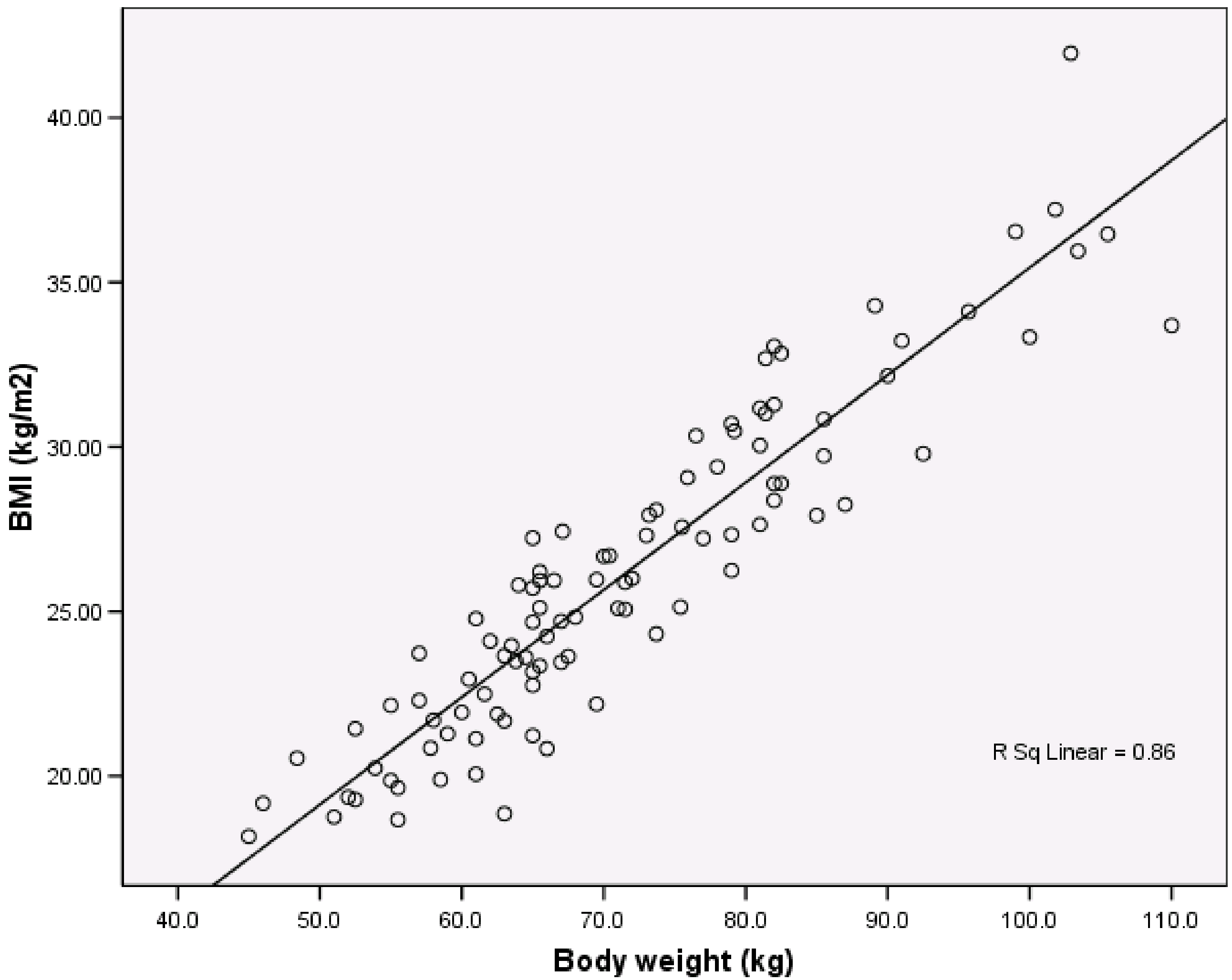
- To model relationships between variables, determine **the magnitude of the relationships between variables**, and can be used to make predictions based on the models (Wikipedia 2006)

Linear Regression

- The process of fitting **the best possible straight line** through a series of points

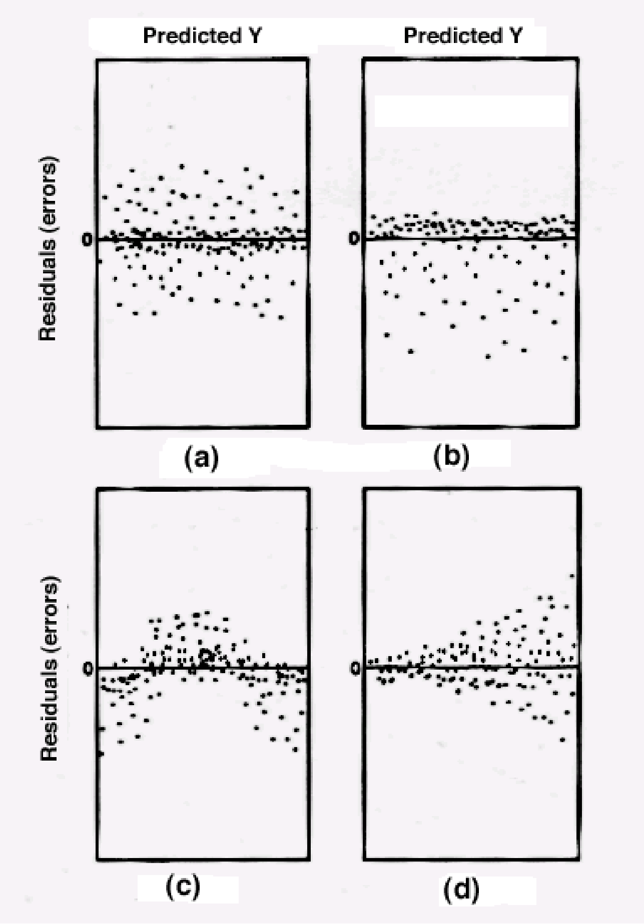


y
y = 0.5842x + 1.6842
R² = 0.9488
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
x



BMI (kg/m2)
40.00
35.00
30.00
25.00
20.00
40.0
50.0
60.0
70.0
80.0
90.0
100.0
110.0
Body weight (kg)
R Sq Linear = 0.86

Residual



- The difference between the measured and predicted values of some quantity
- Residual vs. Predicted – test linearity, normality & heteroscedasticity

Type of Linear Regression

Dependent Variable *(Yi)*	Independent Variable *(Xi)*	Test
1 – Cont	≥ 2 – All Cont	Linear Regression
1 – Cont	≥ 2 – All Cat	ANOVA
1 – Cont	≥ 2 – Cont + Cat	ANCOVA

Biostatistics for Medical Researchers (Advanced)

LINEAR REGRESSION

Simple Linear Regression

- How good age predicts arterial blood pressure?
- Use file Data2_abp.sav

Assumptions

1. Multivariate **normal** – for all IV, check residuals
2. **Linearity** – scatterplot for each IV, check residuals
3. No **multicollinearity** – correlation matric between IVs, high tolerance, low VIF
4. **Homoscedasticity** of residuals – residuals should be normally distributed (zresid vs zpred plot)

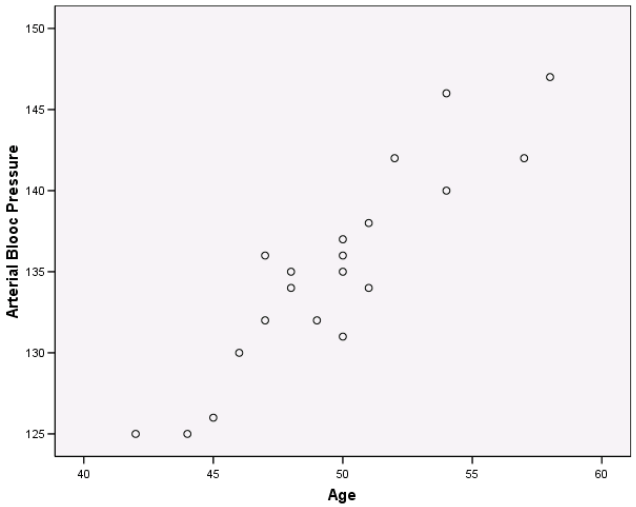
Testing Assumptions

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Arterial Blood Pressure	.096	20	.200*	.966	20	.663
Age	.120	20	.200*	.978	20	.898

*. This is a lower bound of the true significance.

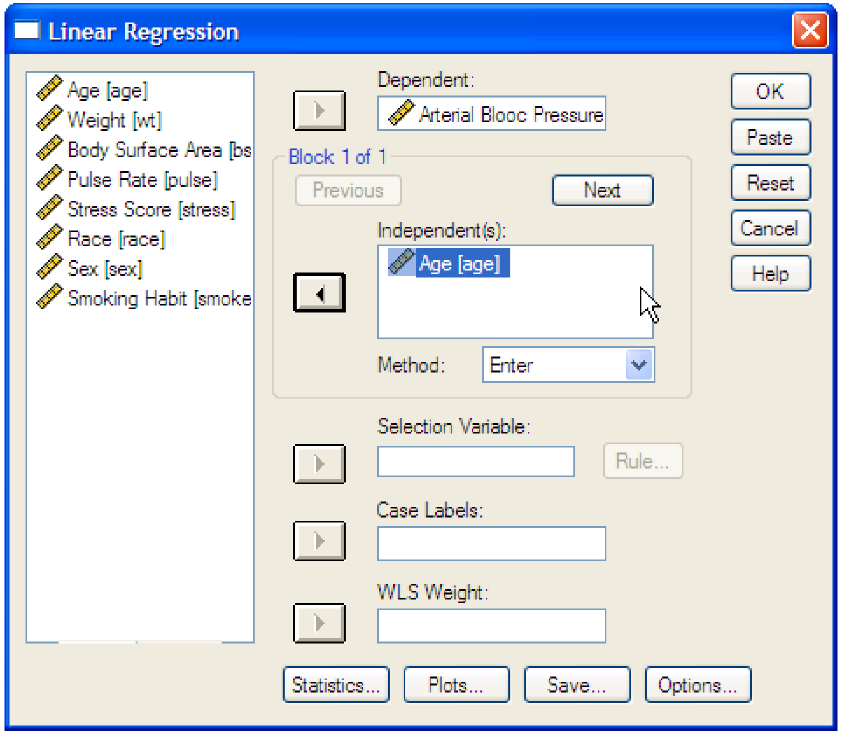
a. Lilliefors Significance Correction

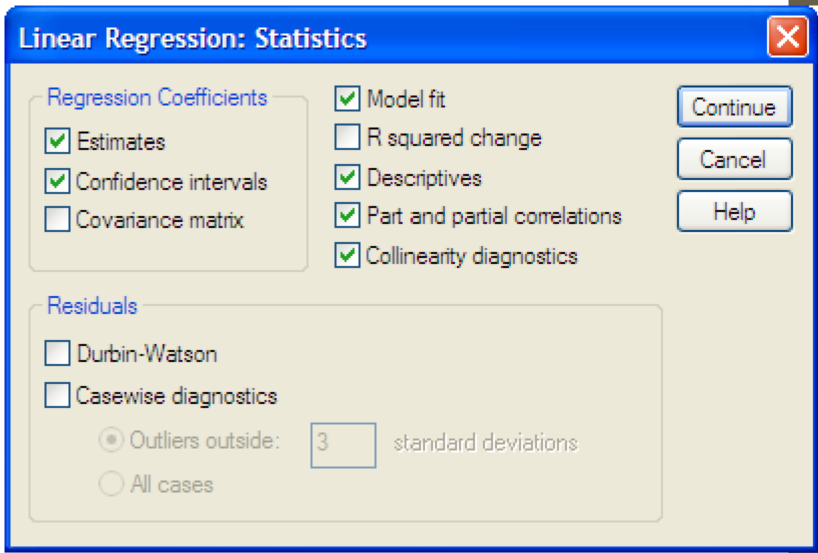


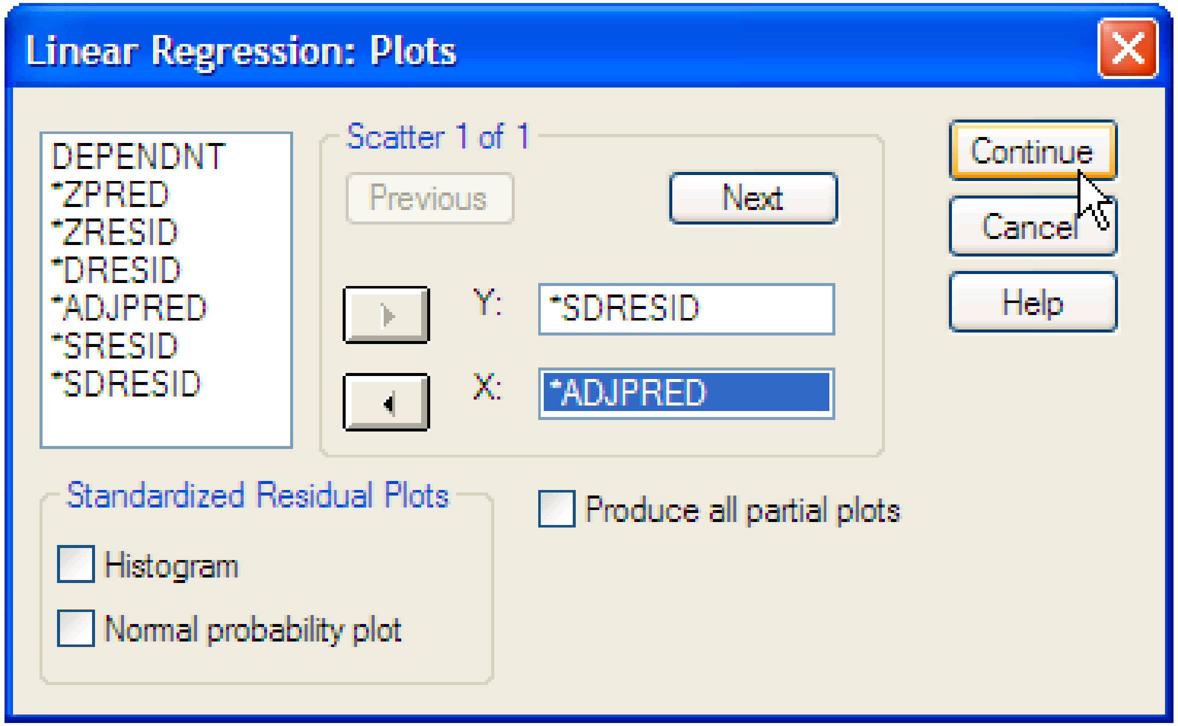
Correlations

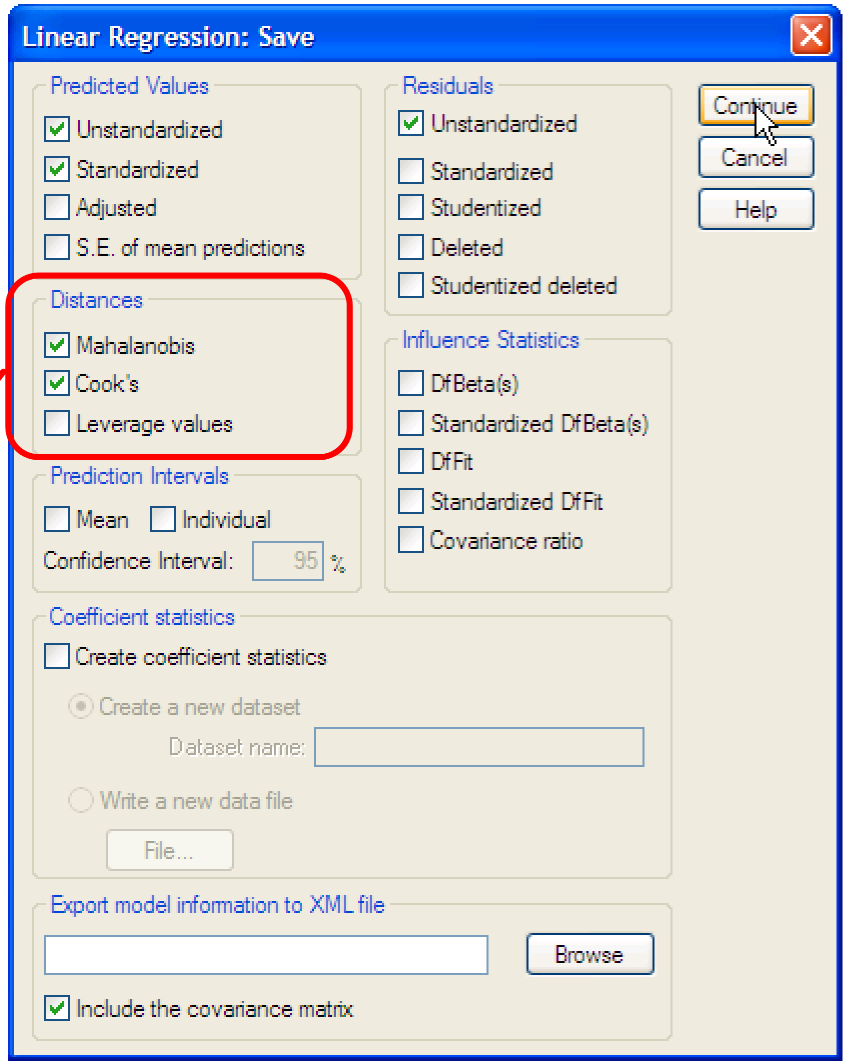
		Arterial Blooc Pressure	Age
Arterial Blooc Pressure	Pearson Correlation	1	.911**
	Sig. (2-tailed)		.000
	N	20	20
Age	Pearson Correlation	.911**	1
	Sig. (2-tailed)	.000	
	N	20	20

**. Correlation is significant at the 0.01 level (2-tailed).


Linear Regression
Age [age]
Weight [wt]
Body Surface Area [bs
Pulse Rate [pulse]
Stress Score [stress]
Race [race]
Sex [sex]
Smoking Habit [smoke
Dependent:
Arterial Blooc Pressure
Block 1 of 1
Previous
Next
Independent(s):
Age [age]
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
OK
Paste
Reset
Cancel
Help
Statistics...
Plots...
Save...
Options...


Linear Regression: Statistics
Regression Coefficients
Estimates
Confidence intervals
Covariance matrix
Model fit
R squared change
Descriptives
Part and partial correlations
Collinearity diagnostics
Continue
Cancel
Help
Residuals
Durbin-Watson
Casewise diagnostics
Outliers outside:
3
standard deviations
All cases


Linear Regression: Plots
DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID
Scatter 1 of 1
Previous
Next
Y:
*SDRESID
X:
*ADJPRED
Continue
Cancel
Help
Standardized Residual Plots
Histogram
Normal probability plot
Produce all partial plots



Checking properties of variance

Interpret Output

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	65.816	7.443		8.842	.000	50.178	81.454		
	Age	1.396	.149	.911	9.345	.000	1.083	1.710	1.000	1.000

a. Dependent Variable: Arterial Blood Pressure

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.911[a]	.829	.820	2.659	1.790

a. Predictors: (Constant), Age

b. Dependent Variable: Arterial Blood Pressure

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	617.306	1	617.306	87.324	.000[a]
	Residual	127.244	18	7.069		
	Total	744.550	19			

a. Predictors: (Constant), Age

b. Dependent Variable: Arterial Blood Pressure

Checking Assumptions

B	Collinearity Statistics	
nd	Tolerance	VIF
54		
10	1.000	1.000

Coefficient Correlations[a]

Model			Age
1	Correlations	Age	1.000
	Covariances	Age	.022

a. Dependent Variable: Arterial Blood Pressure

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	Age
1	1	1.997	1.000	.00	.00
	2	.003	25.000	1.00	1.00

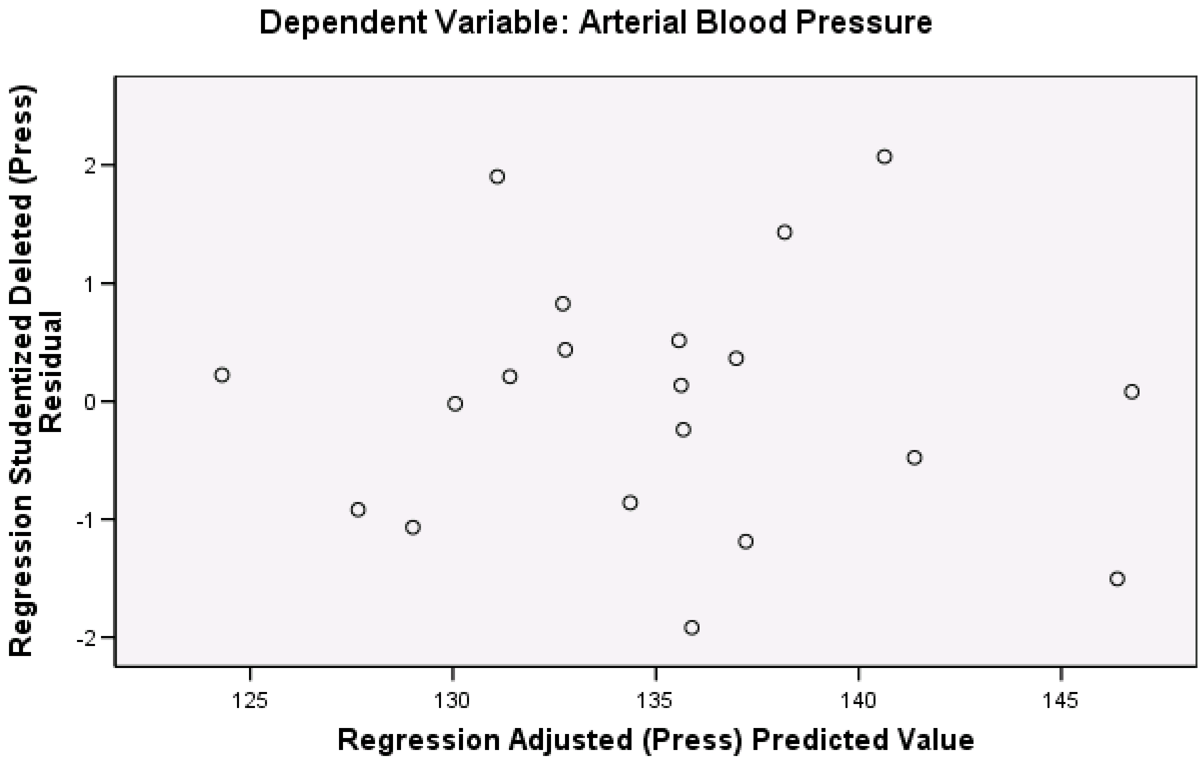
a. Dependent Variable: Arterial Blood Pressure

Residuals Statistics[a]

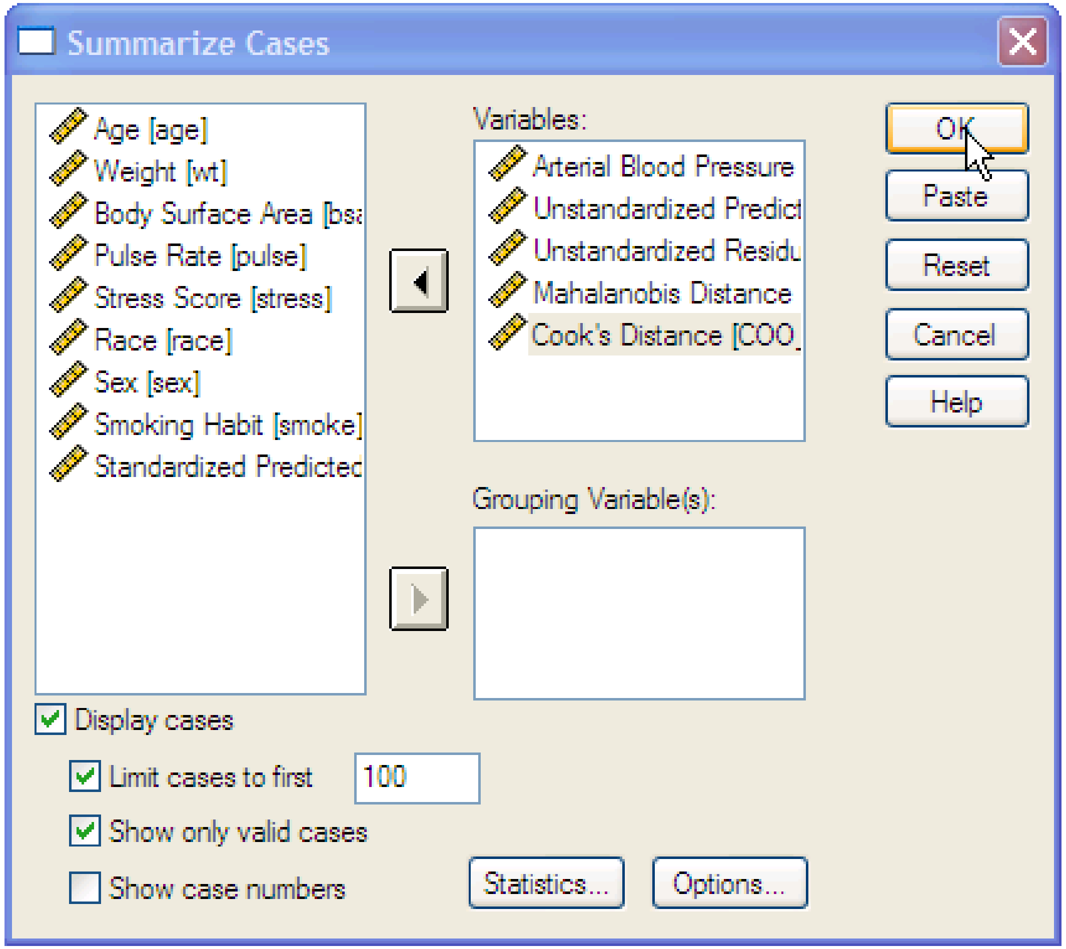
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	124.47	146.81	135.15	5.700	20
Std. Predicted Value	-1.874	2.046	.000	1.000	20
Standard Error of Predicted Value	.597	1.382	.804	.251	20
Adjusted Predicted Value	124.30	146.74	135.18	5.746	20
Residual	-4.639	4.775	.000	2.588	20
Std. Residual	-1.745	1.796	.000	.973	20
Stud. Residual	-1.790	1.904	-.005	1.023	20
Deleted Residual	-4.885	5.364	-.031	2.864	20
Stud. Deleted Residual	-1.919	2.070	.000	1.069	20
Mahal. Distance	.007	4.185	.950	1.286	20
Cook's Distance	.000	.300	.054	.080	20
Centered Leverage Value	.000	.220	.050	.068	20

a. Dependent Variable: Arterial Blood Pressure

Scatterplot



Checking the prediction



Case Summaries[a]

		Arterial Blood Pressure	Unstandardized Predicted Value	Unstandardized Residual	Mahalanobis Distance	Cook's Distance
1		125	127.25999	-2.25999	1.91606	.07557
2		136	135.63876	.36124	.00735	.00052
3		137	135.63876	1.36124	.00735	.00732
4		138	137.03522	.96478	.10939	.00412
5		132	134.24230	-2.24230	.02536	.02029
6		147	146.81046	.18954	4.18489	.00129
7		142	138.43169	3.56831	.33147	.06985
8		130	130.05291	-.05291	.79964	.00002
9		131	135.63876	-4.63876	.00735	.08504
10		134	132.84584	1.15416	.16341	.00623
11		140	141.22461	-1.22461	1.13577	.01469
12		135	135.63876	-.63876	.00735	.00161
13		134	137.03522	-3.03522	.10939	.04075
14		126	128.65645	-2.65645	1.29783	.07596
15		146	141.22461	4.77539	1.13577	.22343
16		135	132.84584	2.15416	.16341	.02170
17		125	124.46707	.53293	3.51264	.00806
18		136	131.44938	4.55062	.42151	.12282
19		132	131.44938	.55062	.42151	.00180
20		142	145.41399	-3.41399	3.24254	.29950
Total	N	20	20	20	20	20

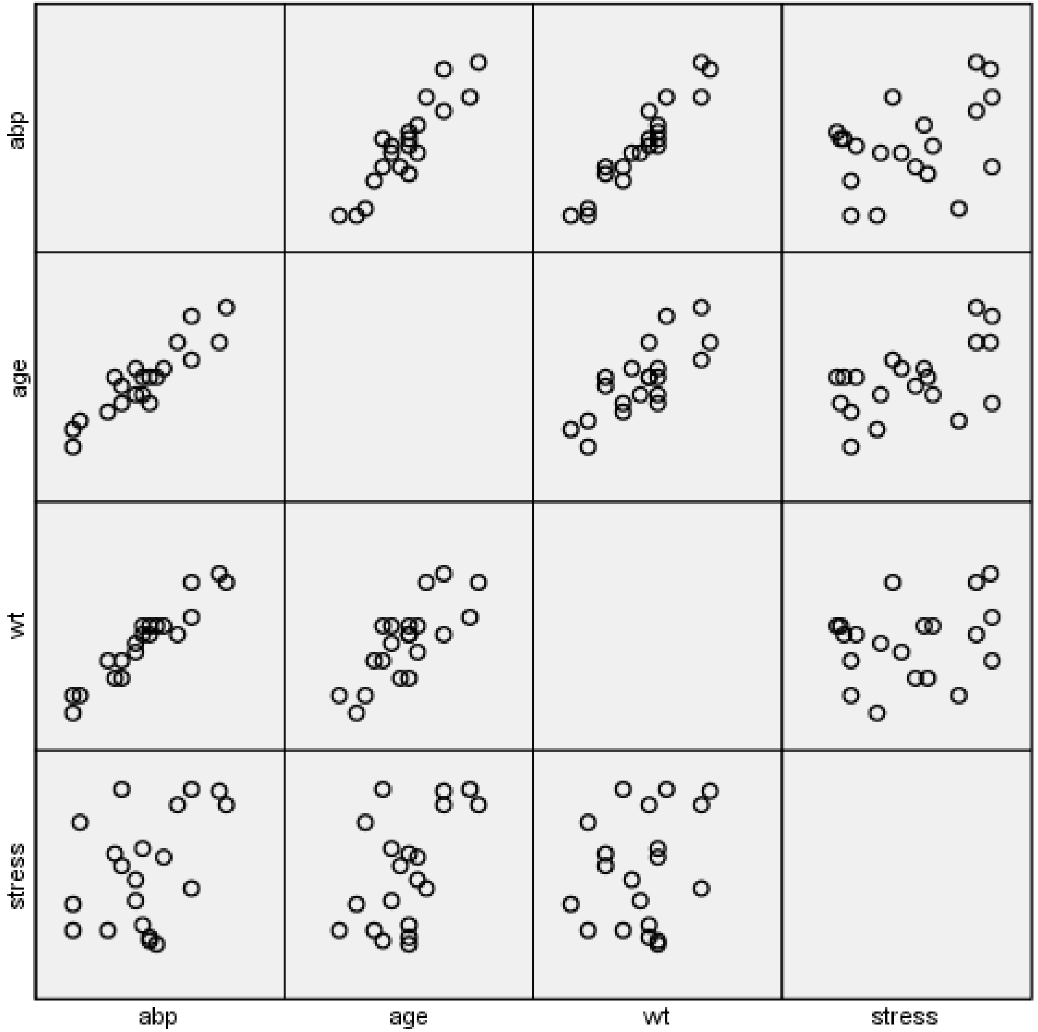
a. Limited to first 100 cases.

Multiple Regression

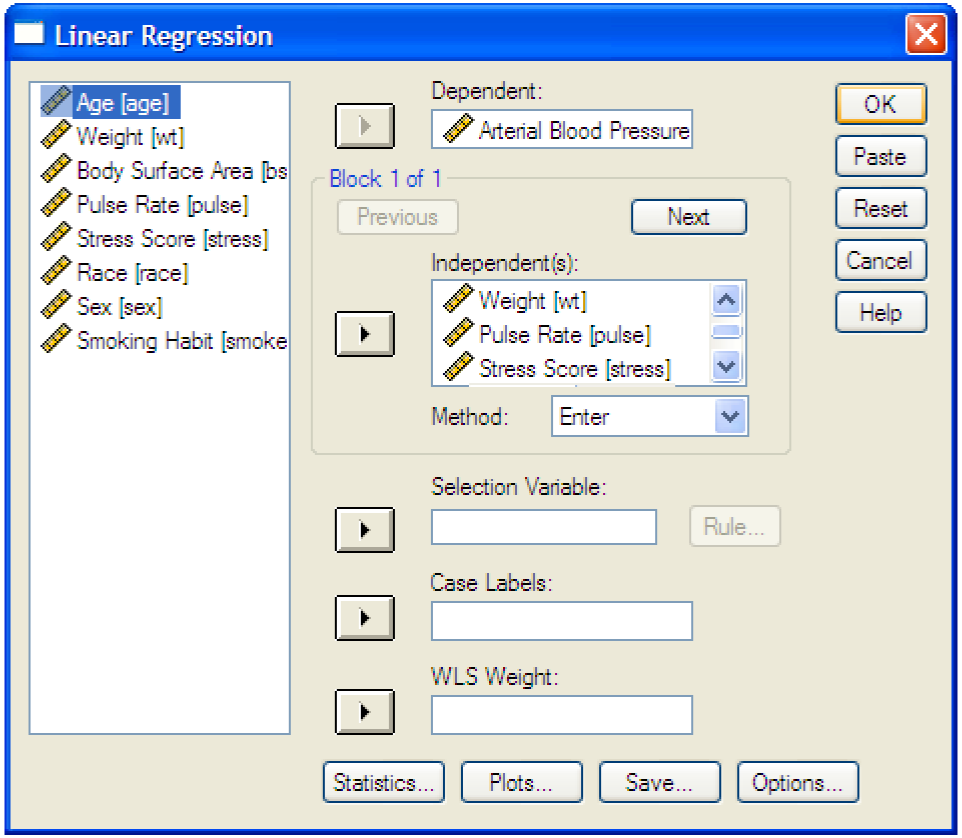
$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

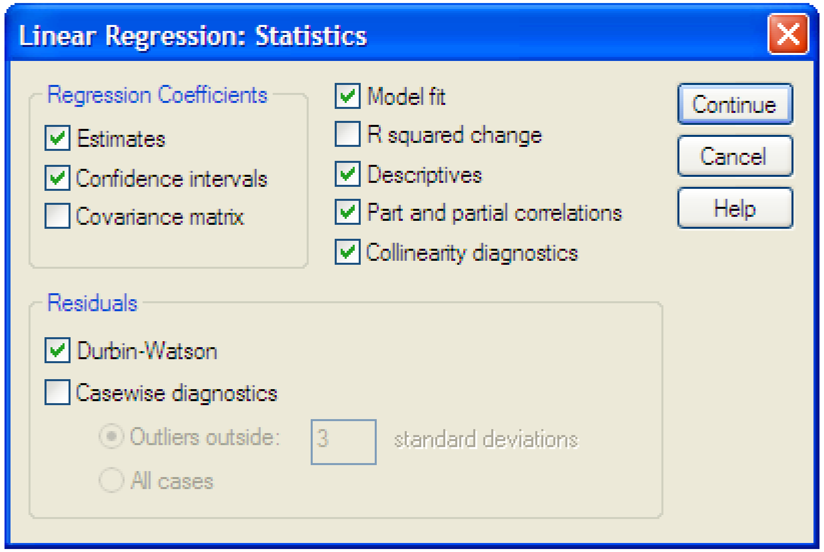
- Arterial BP
- Age
- Body weight
- Pulse rate
- Stress

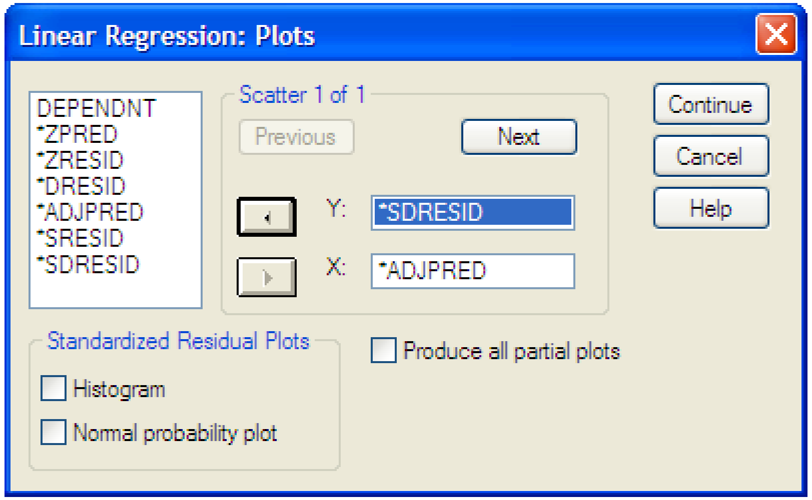
- You do it first.
- Run multiple regression.
- Describe the coefficients.
- Check the model fit.
- Check assumptions.

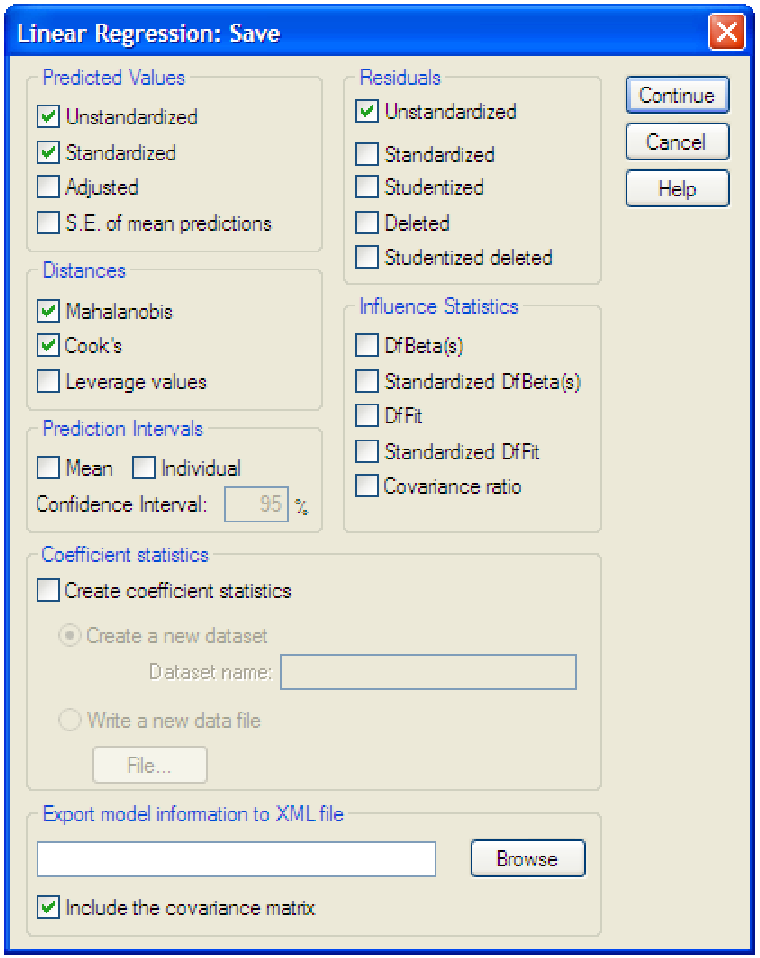

abp
age
wt
stress
abp
age
wt
stress

The steps


Linear Regression
Age [age]
Weight [wt]
Body Surface Area [bs
Pulse Rate [pulse]
Stress Score [stress]
Race [race]
Sex [sex]
Smoking Habit [smoke
Dependent:
Arterial Blood Pressure
Block 1 of 1
Previous
Next
Independent(s):
Weight [wt]
Pulse Rate [pulse]
Stress Score [stress]
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
OK
Paste
Reset
Cancel
Help
Statistics...
Plots...
Save...
Options...


Linear Regression: Statistics
Regression Coefficients
Estimates
Confidence intervals
Covariance matrix
Model fit
R squared change
Descriptives
Part and partial correlations
Collinearity diagnostics
Continue
Cancel
Help
Residuals
Durbin-Watson
Casewise diagnostics
Outliers outside:
3
standard deviations
All cases


Linear Regression: Plots
DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID
Scatter 1 of 1
Previous
Next
Y:
*SDRESID
X:
*ADJPRED
Continue
Cancel
Help
Standardized Residual Plots
Histogram
Normal probability plot
Produce all partial plots


Linear Regression: Save
Predicted Values
Unstandardized
Standardized
Adjusted
S.E. of mean predictions
Distances
Mahalanobis
Cook's
Leverage values
Prediction Intervals
Mean
Individual
Confidence Interval:
95
%
Residuals
Unstandardized
Standardized
Studentized
Deleted
Studentized deleted
Influence Statistics
DfBeta(s)
Standardized DfBeta(s)
DfFit
Standardized DfFit
Covariance ratio
Continue
Cancel
Help
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

Interpret Output

Descriptive Statistics

	Mean	Std. Deviation	N
Arterial Blood Pressure	135.15	6.260	20
Age	49.65	4.082	20
Weight	93.15	4.392	20
Pulse Rate	69.60	3.803	20
Stress Score	52.40	31.955	20

Correlations

		Arterial Blood Pressure	Age	Weight	Pulse Rate	Stress Score
Pearson Correlation	Arterial Blood Pressure	1.000	.911	.954	.752	.408
	Age	.911	1.000	.772	.764	.533
	Weight	.954	.772	1.000	.640	.250
	Pulse Rate	.752	.764	.640	1.000	.529
	Stress Score	.408	.533	.250	.529	1.000
Sig. (1-tailed)	Arterial Blood Pressure	.	.000	.000	.000	.037
	Age	.000	.	.000	.000	.008
	Weight	.000	.000	.	.001	.144
	Pulse Rate	.000	.000	.001	.	.008
	Stress Score	.037	.008	.144	.008	.
N	Arterial Blood Pressure	20	20	20	20	20
	Age	20	20	20	20	20
	Weight	20	20	20	20	20
	Pulse Rate	20	20	20	20	20
	Stress Score	20	20	20	20	20

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.994[a]	.987	.984	.793	1.744

a. Predictors: (Constant), Stress Score, Weight, Pulse Rate, Age

b. Dependent Variable: Arterial Blood Pressure

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	735.108	4	183.777	291.948	.000[a]
	Residual	9.442	15	.629		
	Total	744.550	19			

a. Predictors: (Constant), Stress Score, Weight, Pulse Rate, Age

b. Dependent Variable: Arterial Blood Pressure

R^2 = 0.986, meaning 99% of variation in ABP is explained by Age, Body Weight, Pulse Rate & Stress ($F(4,15)=291.948$), $P<0.001$)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	17.308	4.623		3.744	.002	7.454	27.163
	Age	.573	.091	.374	6.285	.000	.379	.767
	Weight	.892	.070	.626	12.756	.000	.743	1.040
	Pulse Rate	.087	.078	.053	1.126	.278	-.078	.253
	Stress Score	.005	.007	.025	.662	.518	-.011	.021

a. Dependent Variable: Arterial Blood Pressure

How to report

	Zero-Order r						
Variable	Age	Weight	Pulse Rate	Stress Score	ABP	β	B
Age		0.772*	0.764*	0.533*	0.911*	0.374*	0.573
Weight			0.640*	0.25	0.954*	0.626*	0.892
Pulse Rate				0.529*	0.752*	0.053	0.087
Stress Score					0.408*	0.025	0.005
					Intercept	=	17.308
Mean	49.65	93.15	69.6	52.4	135.15		
SD	4.08	4.4	3.8	31.95	6.26	R^2 =	0.994*

* P<0.05

Checking assumptions

Correlations			Collinearity Statistics	
Zero-order	Partial	Part	Tolerance	VIF
.911	.851	.183	.239	4.181
.954	.957	.371	.352	2.844
.752	.279	.033	.379	2.636
.408	.169	.019	.601	1.663

Collinearity Diagnostics[a]

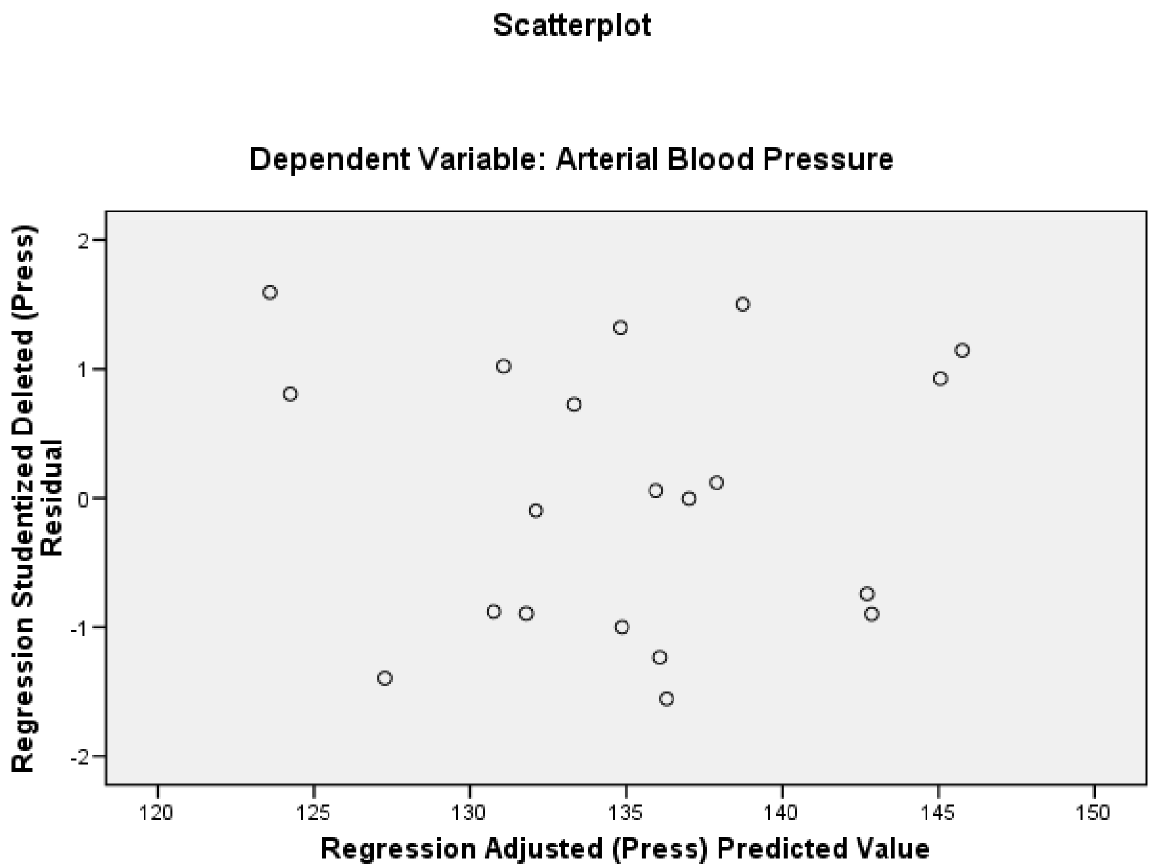
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions				
				(Constant)	Age	Weight	Pulse Rate	Stress Score
1	1	4.792	1.000	.00	.00	.00	.00	.01
	2	.204	4.841	.00	.00	.00	.00	.63
	3	.002	44.207	.27	.33	.00	.00	.19
	4	.001	76.454	.14	.05	.14	.97	.11
	5	.001	97.639	.59	.62	.86	.03	.06

a. Dependent Variable: Arterial Blood Pressure

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	123.97	146.35	135.15	6.220	20
Std. Predicted Value	-1.797	1.800	.000	1.000	20
Standard Error of Predicted Value	.262	.546	.390	.072	20
Adjusted Predicted Value	123.59	145.76	135.11	6.194	20
Residual	-1.078	1.031	.000	.705	20
Std. Residual	-1.359	1.300	.000	.889	20
Stud. Residual	-1.487	1.518	.023	1.025	20
Deleted Residual	-1.290	1.414	.044	.944	20
Stud. Deleted Residual	-1.555	1.594	.026	1.045	20
Mahal. Distance	1.128	8.045	3.800	1.741	20
Cook's Distance	.000	.231	.070	.062	20
Centered Leverage Value	.059	.423	.200	.092	20

a. Dependent Variable: Arterial Blood Pressure


Scatterplot
Dependent Variable: Arterial Blood Pressure
Regression Studentized Deleted (Press) Residual
2
1
0
-1
-2
120
125
130
135
140
145
150
Regression Adjusted (Press) Predicted Value

Exercise 3

- Please use pwv.sav
- Fit a multiple linear regression model for predicting the Pulse Wave Velocity

[Multivariate Analysis]

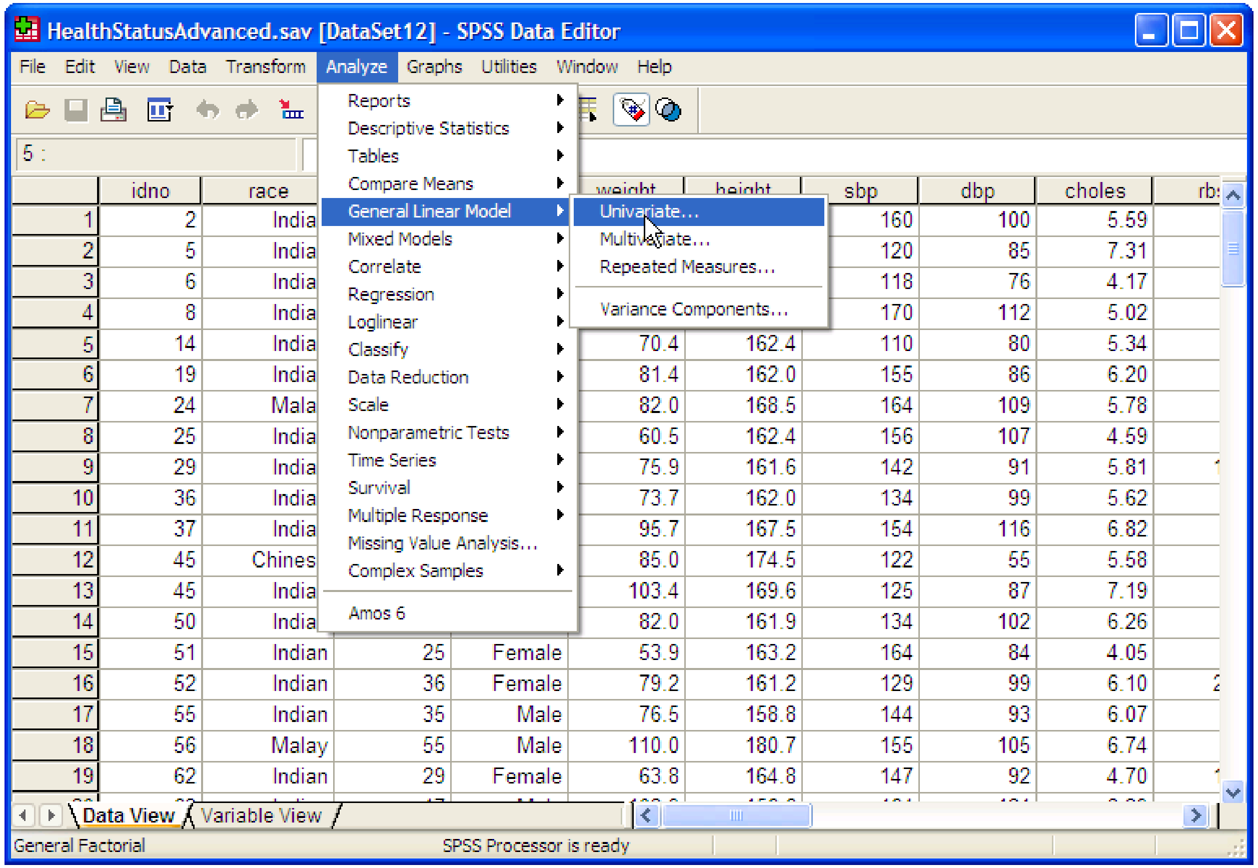
Univariate GLM

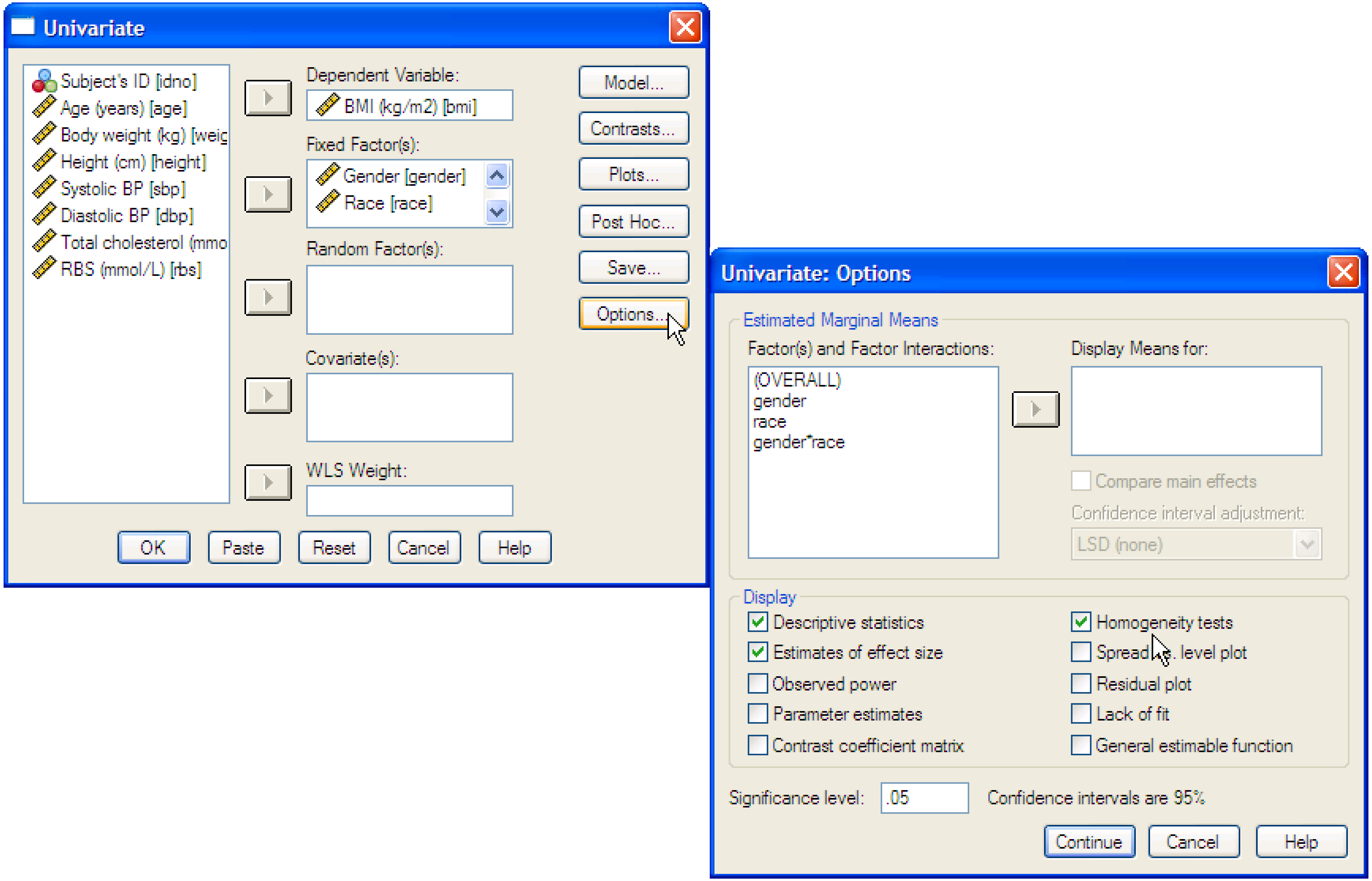
Jamalludin Ab Rahman MD MPH
Department of Community Medicine
International Islamic University Malaysia

When to use Univariate GLM

- When we want to analyze the relationship of a continuous dependant variable and one or more continuous or/and categorical independent variables
- When we want to control the effect of confounders

Two-way ANOVA




Univariate
Subject's ID [idno]
Age (years) [age]
Body weight (kg) [weig
Height (cm) [height]
Systolic BP [sbp]
Diastolic BP [dbp]
Total cholesterol (mmo
RBS (mmol/L) [rbs]
Dependent Variable:
BMI (kg/m2) [bmi]
Fixed Factor(s):
Gender [gender]
Race [race]
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
OK
Paste
Reset
Cancel
Help
Univariate: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
gender
race
gender*race
Display Means for:
Compare main effects
Confidence interval adjustment:
LSD (none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
Contrast coefficient matrix
Homogeneity tests
Spread vs. level plot
Residual plot
Lack of fit
General estimable function
Significance level:
.05
Confidence intervals are 95%
Continue
Cancel
Help

Descriptive Statistics

Dependent Variable: BMI (kg/m2)

Gender	Race	Mean	Std. Deviation	N
Male	Malay	24.1883	4.96657	16
	Chinese	24.7853	3.13839	16
	Indian	29.2637	4.77315	21
	Total	26.3795	4.93336	53
Female	Malay	24.9971	4.43537	16
	Chinese	24.7793	4.09399	17
	Indian	28.1758	5.87370	14
	Total	25.8651	4.93129	47
Total	Malay	24.5927	4.65009	32
	Chinese	24.7822	3.60518	33
	Indian	28.8285	5.18512	35
	Total	26.1378	4.91419	100

Levene's Test of Equality of Error Variances[a]

Dependent Variable: BMI (kg/m2)

F	df1	df2	Sig.
1.985	5	94	.088

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

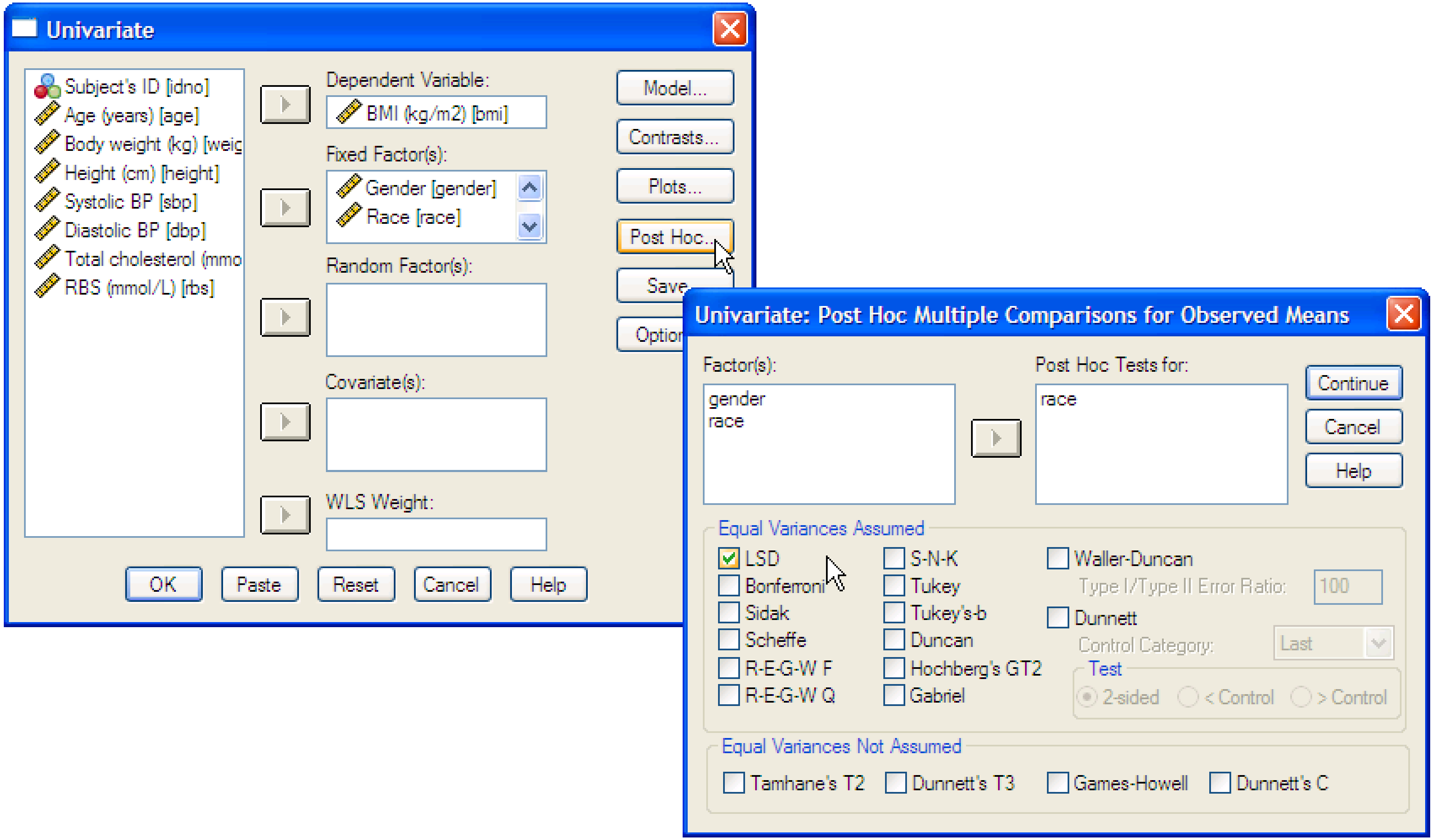
a. Design: Intercept+gender+race+gender * race

Tests of Between-Subjects Effects

Dependent Variable: BMI (kg/m2)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	405.609[a]	5	81.122	3.841	.003	.170
Intercept	66768.090	1	66768.090	3161.544	.000	.971
gender	.223	1	.223	.011	.918	.000
race	360.409	2	180.204	8.533	.000	.154
gender * race	14.875	2	7.437	.352	.704	.007
Error	1985.169	94	21.119			
Total	70709.067	100				
Corrected Total	2390.778	99				

a. R Squared = .170 (Adjusted R Squared = .125)



Univariate
Subject's ID [idno]
Age (years) [age]
Body weight (kg) [weig
Height (cm) [height]
Systolic BP [sbp]
Diastolic BP [dbp]
Total cholesterol (mmo
RBS (mmol/L) [rbs]
Dependent Variable:
BMI (kg/m2) [bmi]
Fixed Factor(s):
Gender [gender]
Race [race]
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save
OK
Paste
Reset
Cancel
Help
Univariate: Post Hoc Multiple Comparisons for Observed Means
Factor(s):
gender
race
Post Hoc Tests for:
race
Continue
Cancel
Help
Equal Variances Assumed
LSD
Bonferroni
Sidak
Scheffe
R-E-G-W F
R-E-G-W Q
S-N-K
Tukey
Tukey's-b
Duncan
Hochberg's GT2
Gabriel
Waller-Duncan
Type I/Type II Error Ratio:
100
Dunnett
Control Category:
Last
Test
2-sided
< Control
> Control
Equal Variances Not Assumed
Tamhane's T2
Dunnett's T3
Games-Howell
Dunnett's C

Multiple Comparisons

Dependent Variable: BMI (kg/m2)

LSD

(I) Race	(J) Race	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Malay	Chinese	-.1895	1.14014	.868	-2.4533	2.0743
	Indian	-4.2358*	1.12399	.000	-6.4675	-2.0041
Chinese	Malay	.1895	1.14014	.868	-2.0743	2.4533
	Indian	-4.0463*	1.11506	.000	-6.2603	-1.8323
Indian	Malay	4.2358*	1.12399	.000	2.0041	6.4675
	Chinese	4.0463*	1.11506	.000	1.8323	6.2603

Based on observed means.

*. The mean difference is significant at the .05 level.

Exercise 2

1. Using HealthStatusAdvanced.sav file, determine the effect of Race and Gender towards glucose, cholesterol and hypertension
2. Using Ex2-2WayANOVA.sav file, study the effect of density and intersec toward number of accident

[Multivariate Analysis]

Multivariate GLM

Jamalludin Ab Rahman MD MPH
Department of Community Medicine
International Islamic University Malaysia

MANOVA

- Multivariate Analysis of Variance
- Involve more than one dependent ~ multivariate
- Dependents are interval
- All independents are categorical

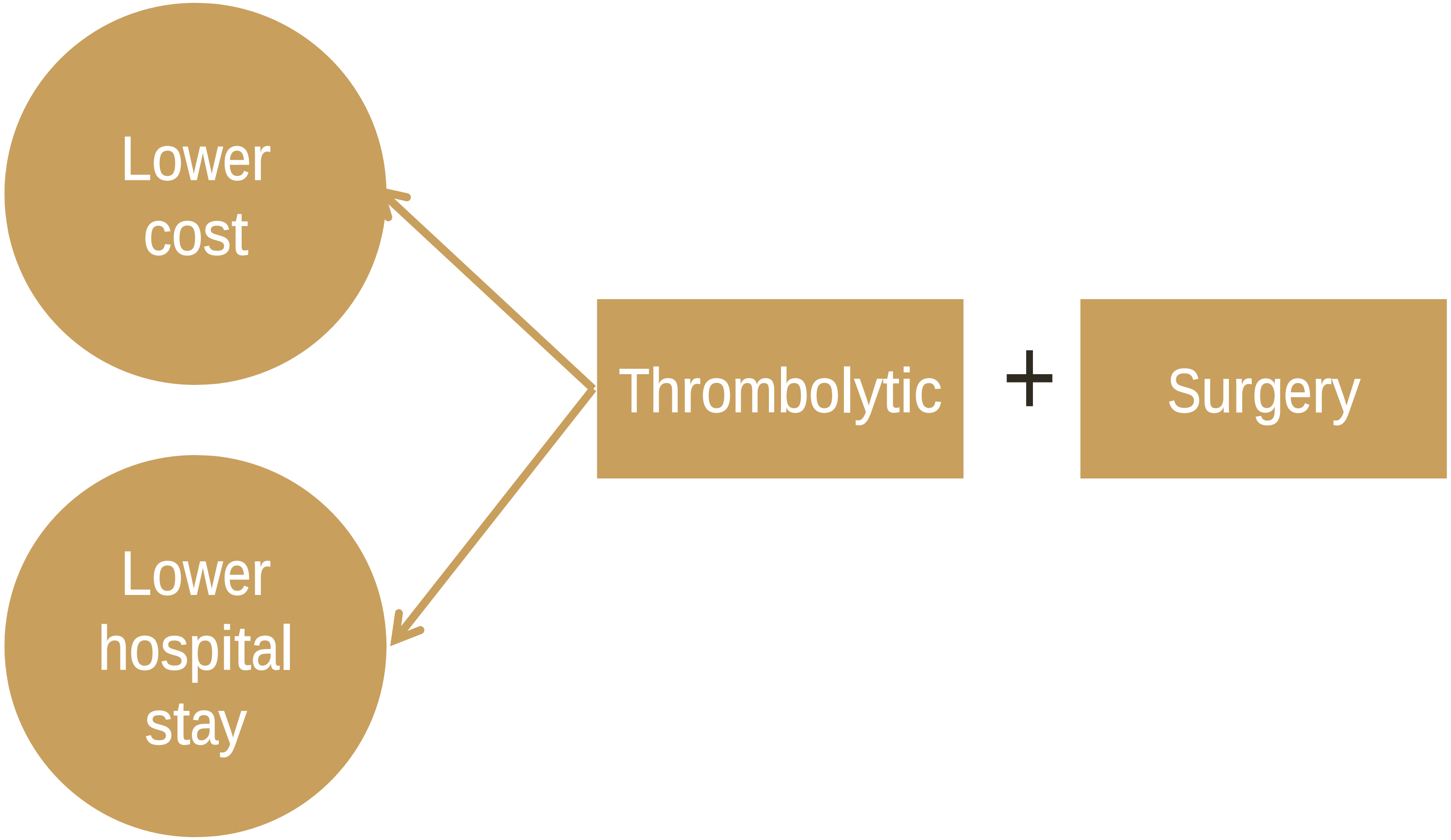
MANOVA answers the followings

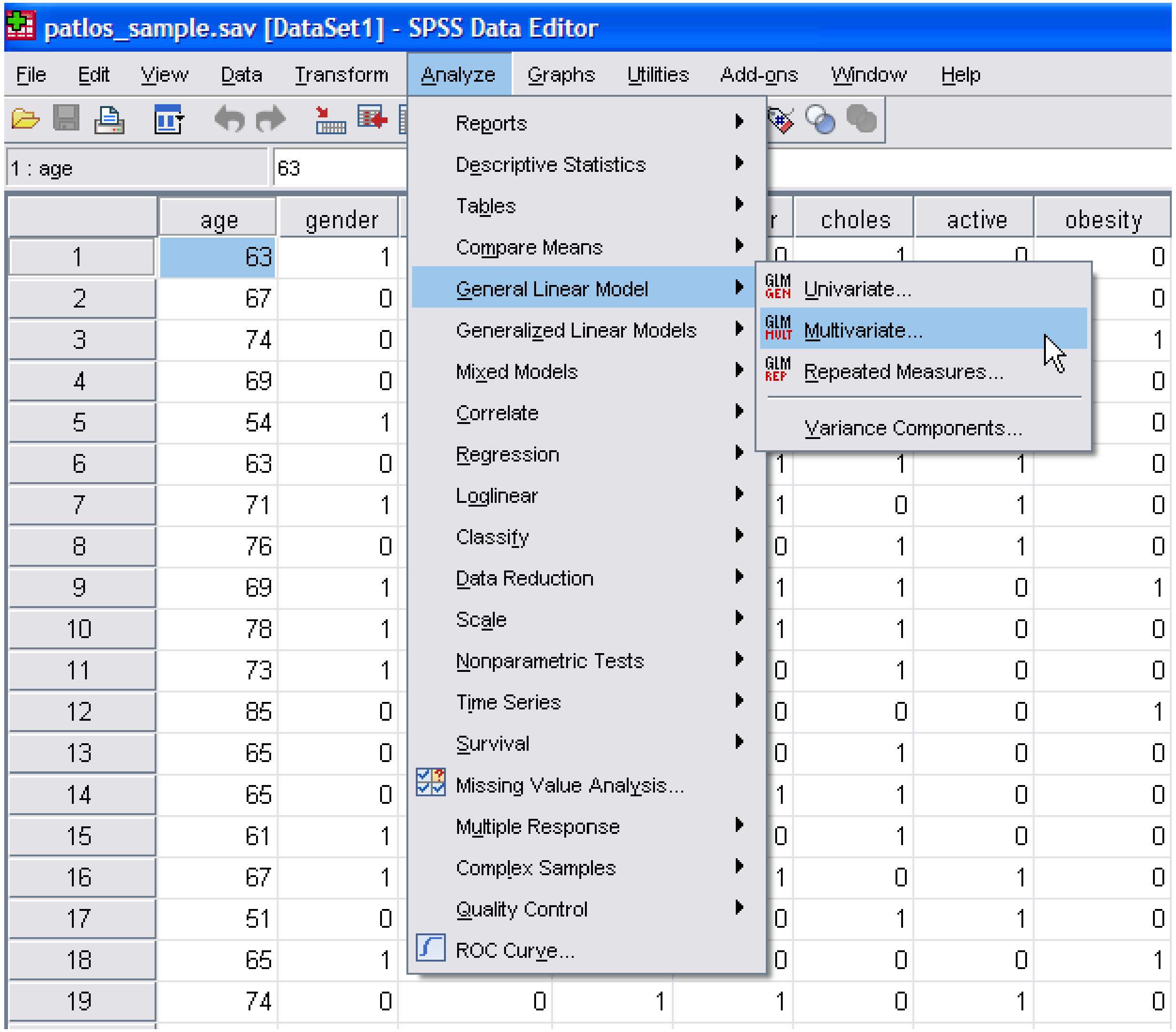
1. What are the main effects of the independent variables?
2. What are the interactions among the independent variables?
3. What is the importance of the dependent variables?
4. What is the strength of association between dependent variables?
5. What are the effects of covariates? How may they be utilized?

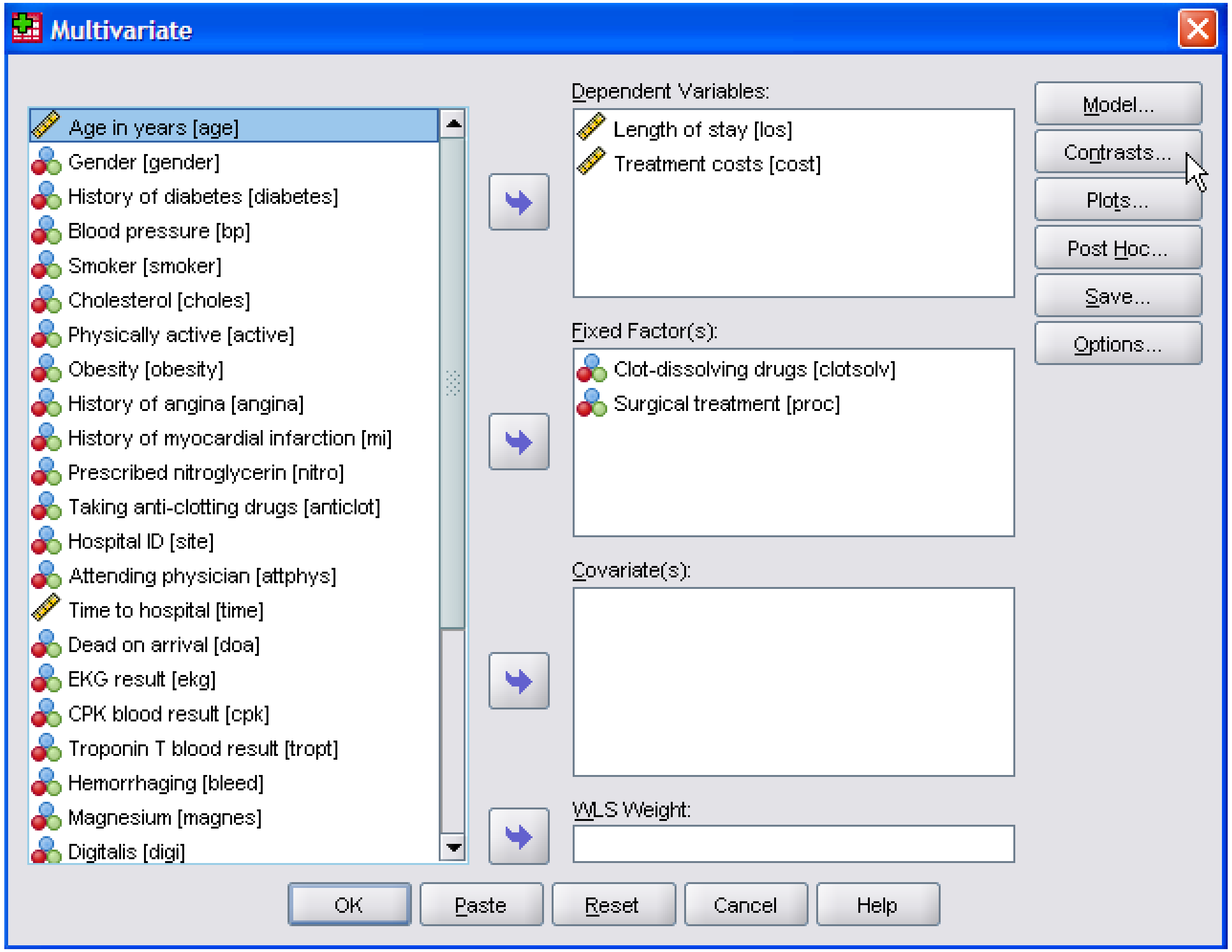
Exercise

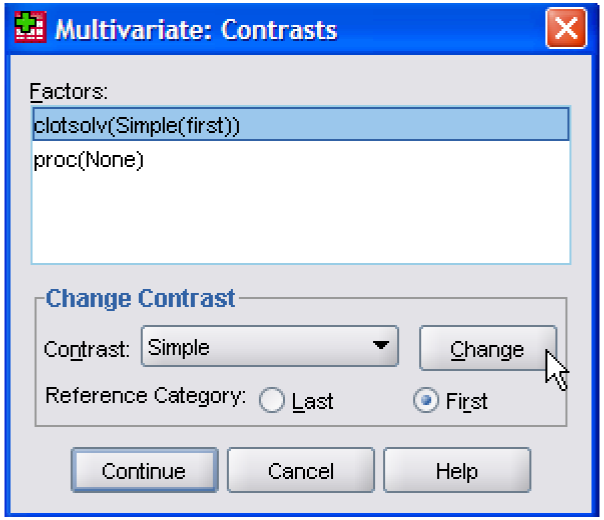
- Which newer thrombolytic agent (between alteplase and reteplase) can reduce length of hospital stay but cost-effective enough to replace currently popular drug (streptokinase)
- Surgery (PTCA & CABG) is the ultimate treatment for atherosclerosis (blocked artery), therefore should be ‘controlled’

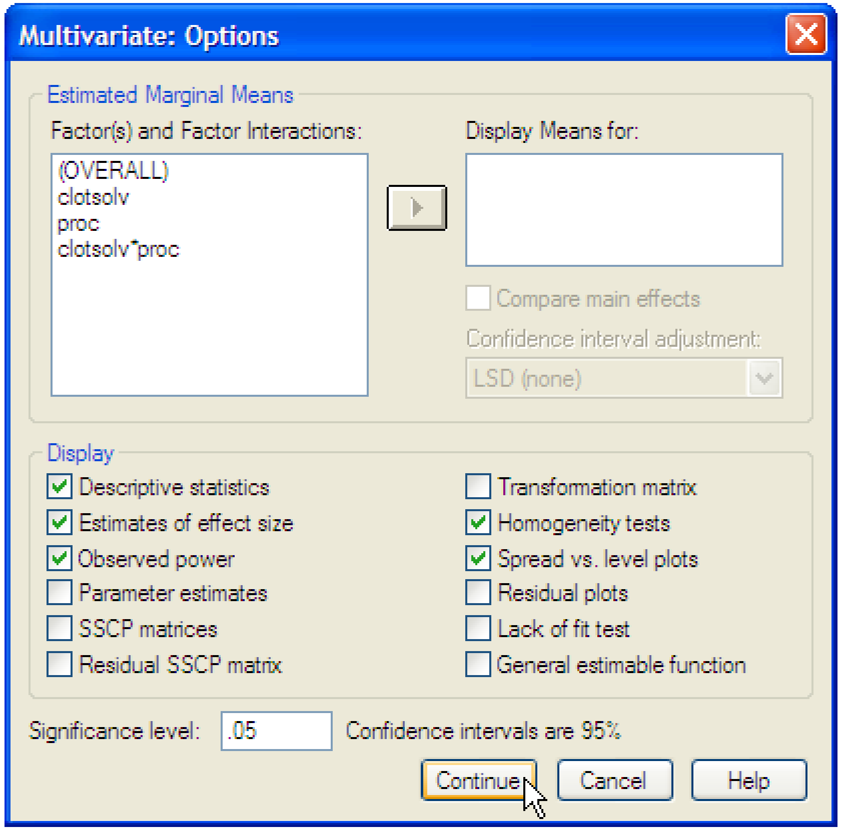
Simplified Model




patlos_sample.sav [DataSet1] - SPSS Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Add-ons
Window
Help
1 : age
63
age
gender
choles
active
obesity
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Missing Value Analysis...
Multiple Response
Complex Samples
Quality Control
ROC Curve...
Univariate...
Multivariate...
Repeated Measures...
Variance Components...


Multivariate
Age in years [age]
Gender [gender]
History of diabetes [diabetes]
Blood pressure [bp]
Smoker [smoker]
Cholesterol [choles]
Physically active [active]
Obesity [obesity]
History of angina [angina]
History of myocardial infarction [mi]
Prescribed nitroglycerin [nitro]
Taking anti-clotting drugs [anticlot]
Hospital ID [site]
Attending physician [attphys]
Time to hospital [time]
Dead on arrival [doa]
EKG result [ekg]
CPK blood result [cpk]
Troponin T blood result [tropt]
Hemorrhaging [bleed]
Magnesium [magnes]
Digitalis [digi]
Dependent Variables:
Length of stay [los]
Treatment costs [cost]
Fixed Factor(s):
Clot-dissolving drugs [clotsolv]
Surgical treatment [proc]
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
OK
Paste
Reset
Cancel
Help


Multivariate: Contrasts
Factors:
clotsolv(Simple(first))
proc(None)
Change Contrast
Contrast: Simple
Change
Reference Category: Last First
Continue
Cancel
Help


Multivariate: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
clotsolv
proc
clotsolv*proc
Display Means for:
Compare main effects
Confidence interval adjustment:
LSD (none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
SSCP matrices
Residual SSCP matrix
Transformation matrix
Homogeneity tests
Spread vs. level plots
Residual plots
Lack of fit test
General estimable function
Significance level: .05 Confidence intervals are 95%
Continue
Cancel
Help

Between-Subjects Factors

		Value Label	N
Clot-dissolving drugs	1	Streptokinase	116
	2	Reteplase	696
	3	Alteplase	669
Surgical treatment	1	PTCA	907
	2	CABG	574

Descriptive Statistics

	Clot-...	Sur...	Mean	Std. Deviation	N
Length of stay	Streptokinase	PTCA	4.94	1.105	68
		CABG	7.25	1.263	48
		Total	5.90	1.633	116
	Reteplase	PTCA	4.81	1.072	441
		CABG	6.62	1.137	255
		Total	5.47	1.399	696
	Alteplase	PTCA	4.68	1.048	398
		CABG	6.48	1.135	271
		Total	5.41	1.396	669
	Total	PTCA	4.77	1.066	907
		CABG	6.60	1.163	574
		Total	5.48	1.422	1481
Treatment costs	Streptokinase	PTCA	28.3838	3.27388	68
		CABG	44.7225	5.42780	48
		Total	35.1447	9.14344	116
	Reteplase	PTCA	29.6674	3.18096	441
		CABG	44.6251	5.22506	255
		Total	35.1476	8.27021	696
	Alteplase	PTCA	29.8073	3.60094	398
		CABG	44.7432	5.63081	271
		Total	35.8575	8.62337	669
	Total	PTCA	29.6326	3.39406	907
		CABG	44.6890	5.42789	574
		Total	35.4681	8.50314	1481

Box's Test of Equality of Covariance Matrices[a]

Box's M	270.509
F	17.908
df1	15
df2	358296.5
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+clotsolv+proc+clotsolv * proc

Multivariate F

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Intercept	Pillai's Trace	.975	2.878E4	2.000	1474.000	.000	.975	57562.560	1.000
	Wilks' Lambda	.025	2.878E4	2.000	1474.000	.000	.975	57562.560	1.000
	Hotelling's Trace	39.052	2.878E4	2.000	1474.000	.000	.975	57562.560	1.000
	Roy's Largest Root	39.052	2.878E4	2.000	1474.000	.000	.975	57562.560	1.000
clotsolv	Pillai's Trace	.026	9.833	4.000	2950.000	.000	.013	39.332	1.000
	Wilks' Lambda	.974	9.892[a]	4.000	2948.000	.000	.013	39.570	1.000
	Hotelling's Trace	.027	9.952	4.000	2946.000	.000	.013	39.807	1.000
	Roy's Largest Root	.027	19.909[c]	2.000	1475.000	.000	.026	39.818	1.000
proc	Pillai's Trace	.622	1212.157[a]	2.000	1474.000	.000	.622	2424.314	1.000
	Wilks' Lambda	.378	1212.157[a]	2.000	1474.000	.000	.622	2424.314	1.000
	Hotelling's Trace	1.645	1212.157[a]	2.000	1474.000	.000	.622	2424.314	1.000
	Roy's Largest Root	1.645	1212.157[a]	2.000	1474.000	.000	.622	2424.314	1.000
clotsolv * proc	Pillai's Trace	.004	1.508	4.000	2950.000	.197	.002	6.031	.472
	Wilks' Lambda	.996	1.508[a]	4.000	2948.000	.197	.002	6.033	.472
	Hotelling's Trace	.004	1.509	4.000	2946.000	.197	.002	6.036	.472
	Roy's Largest Root	.004	3.022[c]	2.000	1475.000	.049	.004	6.043	.586

a. Exact statistic

b. Computed using alpha = .05

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept + clotsolv + proc + clotsolv * proc

Univariate F

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Length of stay	1217.307[a]	5	243.461	202.406	.000	.407	1012.029	1.000
	Treatment costs	79811.122[c]	5	15962.224	865.665	.000	.746	4328.327	1.000
Intercept	Length of stay	25234.532	1	25234.532	20979.169	.000	.934	20979.169	1.000
	Treatment costs	1027759.201	1	1027759.201	55737.565	.000	.974	55737.565	1.000
clotsolv	Length of stay	26.650	2	13.325	11.078	.000	.015	22.156	.992
	Treatment costs	50.127	2	25.063	1.359	.257	.002	2.718	.294
proc	Length of stay	727.562	1	727.562	604.872	.000	.291	604.872	1.000
	Treatment costs	44593.620	1	44593.620	2418.407	.000	.621	2418.407	1.000
clotsolv * proc	Length of stay	6.757	2	3.379	2.809	.061	.004	5.618	.553
	Treatment costs	50.182	2	25.091	1.361	.257	.002	2.721	.294
Error	Length of stay	1774.185	1475	1.203					
	Treatment costs	27197.902	1475	18.439					
Total	Length of stay	47424.000	1481						
	Treatment costs	1970083.194	1481						
Corrected Total	Length of stay	2991.492	1480						
	Treatment costs	107009.024	1480						

a. R Squared = .407 (Adjusted R Squared = .405)

b. Computed using alpha = .05

c. R Squared = .746 (Adjusted R Squared = .745)

Contrast Result

Contrast Results (K Matrix)

Clot-dissolving drugs Simple Contrast[a]			Dependent Variable: Length of stay	Dependent Variable: Treatment costs
Level 2 vs. Level 1	Contrast Estimate		-.382	.593
	Hypothesized Value		0	0
	Difference (Estimate - Hypothesized)		-.382	.593
	Std. Error		.112	.439
	Sig.		.001	.176
	95% Confidence Interval for Difference	Lower Bound	-.602	-.267
		Upper Bound	-.162	1.453
Level 3 vs. Level 1	Contrast Estimate		-.516	.722
	Hypothesized Value		0	0
	Difference (Estimate - Hypothesized)		-.516	.722
	Std. Error		.112	.439
	Sig.		.000	.100
	95% Confidence Interval for Difference	Lower Bound	-.736	-.138
		Upper Bound	-.296	1.583

a. Reference category = 1

Diagnostic Test

Box's Test of Equality of Covariance Matrices[a]

Box's M	270.509
F	17.908
df1	15.000
df2	358296.484
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + clotsolv + proc + clotsolv * proc

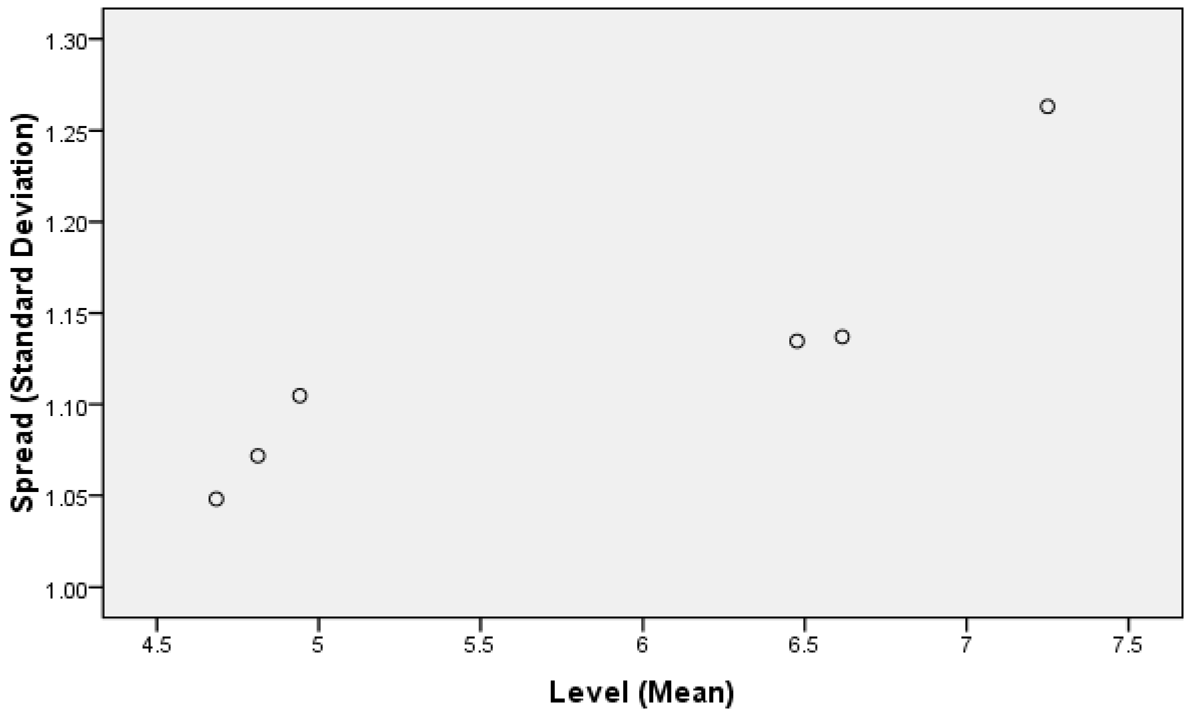
Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Length of stay	1.507	5	1475	.185
Treatment costs	10.001	5	1475	.000

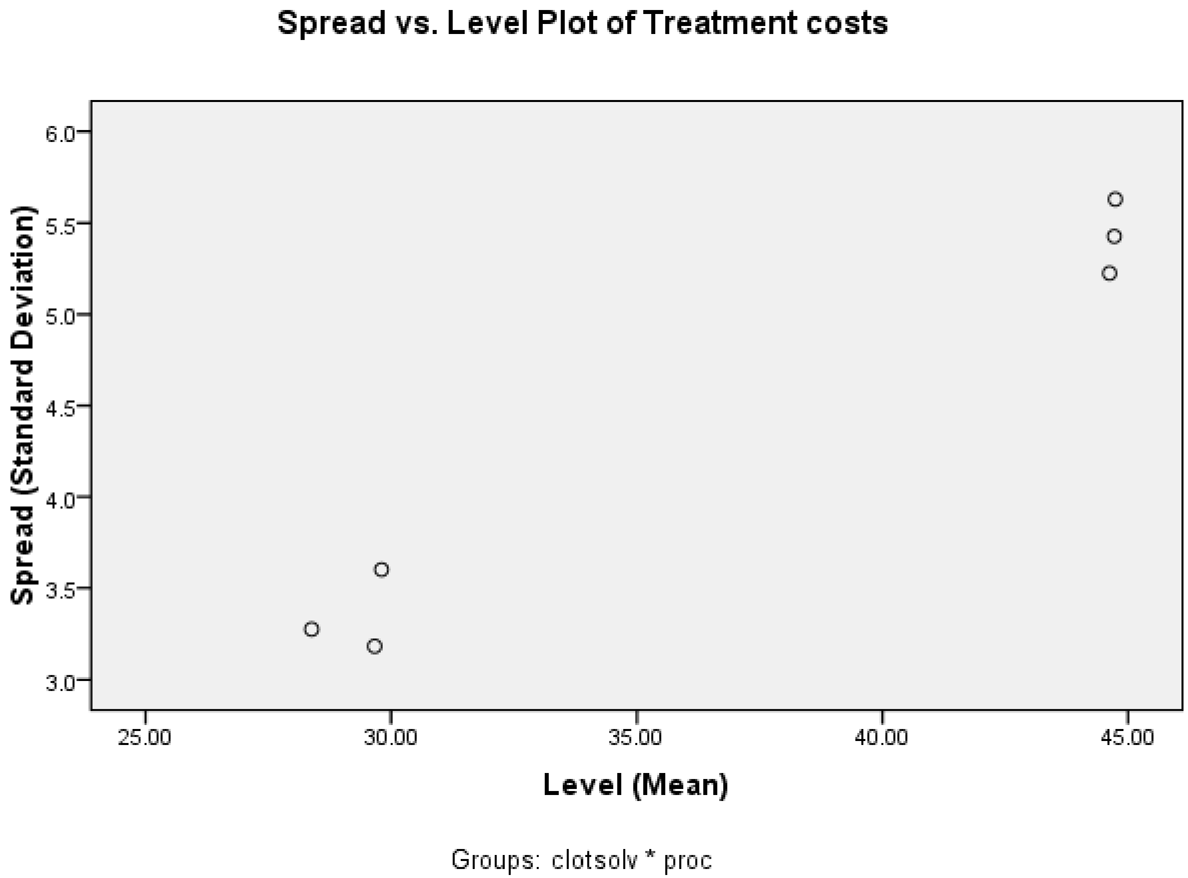
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + clotsolv + proc + clotsolv * proc

Spread vs. Level Plot of Length of stay

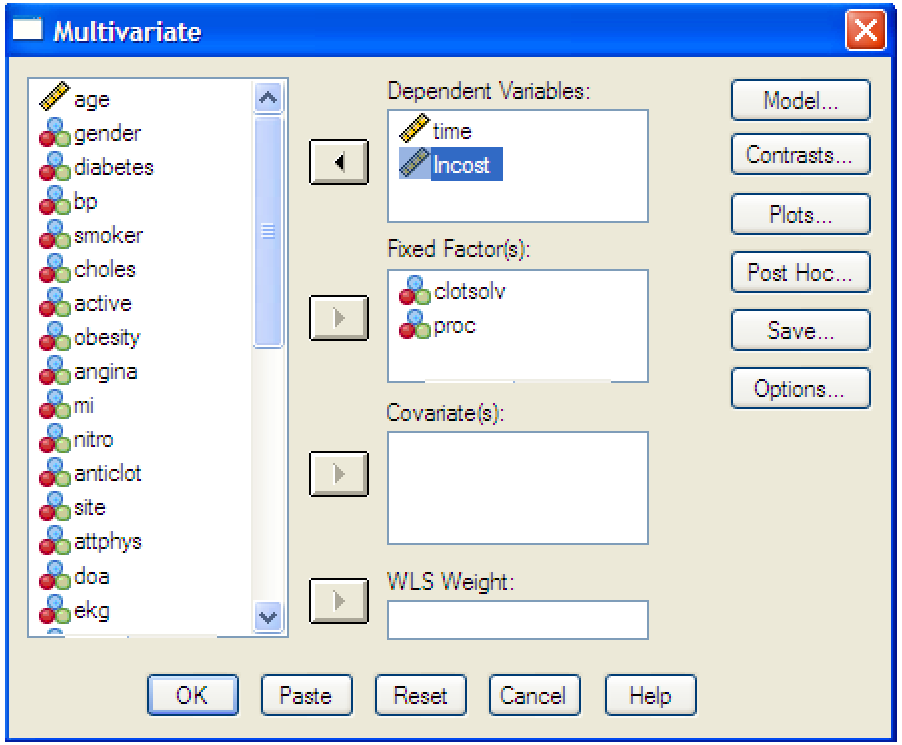


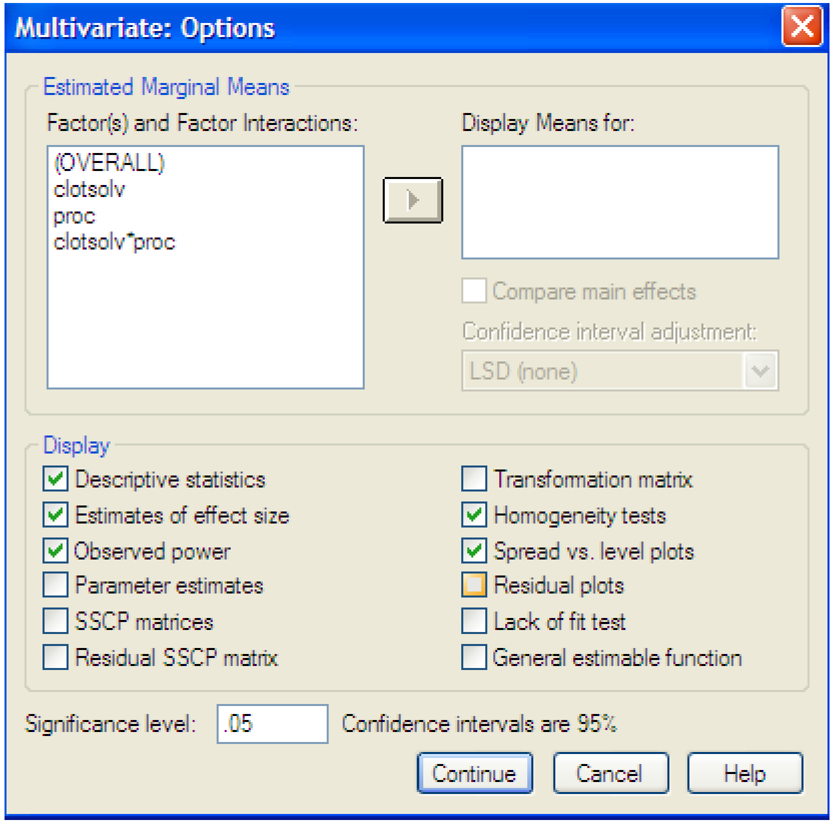
Groups: clotsolv * proc

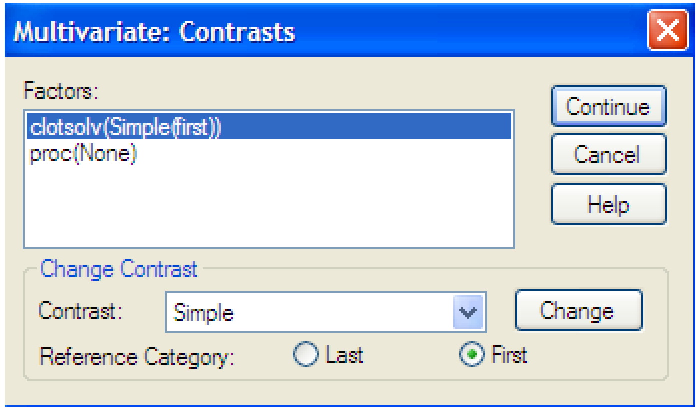


Spread vs. Level Plot of Treatment costs
Spread (Standard Deviation)
6.0
5.5
5.0
4.5
4.0
3.5
3.0
25.00
30.00
35.00
40.00
45.00
Level (Mean)
Groups: clotsolv * proc

Using transformed cost




Multivariate: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
clotsolv
proc
clotsolv*proc
Display Means for:
Compare main effects
Confidence interval adjustment:
LSD (none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
SSCP matrices
Residual SSCP matrix
Transformation matrix
Homogeneity tests
Spread vs. level plots
Residual plots
Lack of fit test
General estimable function
Significance level:
.05
Confidence intervals are 95%
Continue
Cancel
Help


Multivariate: Contrasts
Factors:
clotsolv(Simple(first))
proc(None)
Continue
Cancel
Help
Change Contrast
Contrast:
Simple
Change
Reference Category:
Last
First

Box's Test of Equality of Covariance Matrices[a]

Box's M	296.751
F	19.646
df1	15
df2	358296.5
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+clotsolv+proc+clotsolv * proc

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Time to hospital	46.140	5	1475	.000
Log-cost	1.291	5	1475	.265

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+clotsolv+proc+clotsolv * proc

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	Pillai's Trace	.998	416402.4[b]	2.000	1474.000	.000	.998	832804.790	1.000
	Wilks' Lambda	.002	416402.4[b]	2.000	1474.000	.000	.998	832804.790	1.000
	Hotelling's Trace	564.996	416402.4[b]	2.000	1474.000	.000	.998	832804.790	1.000
	Roy's Largest Root	564.996	416402.4[b]	2.000	1474.000	.000	.998	832804.790	1.000
clotsolv	Pillai's Trace	.237	99.008	4.000	2950.000	.000	.118	396.032	1.000
	Wilks' Lambda	.764	106.290[b]	4.000	2948.000	.000	.126	425.160	1.000
	Hotelling's Trace	.309	113.626	4.000	2946.000	.000	.134	454.505	1.000
	Roy's Largest Root	.306	225.921[c]	2.000	1475.000	.000	.234	451.843	1.000
proc	Pillai's Trace	.662	1442.931[b]	2.000	1474.000	.000	.662	2885.862	1.000
	Wilks' Lambda	.338	1442.931[b]	2.000	1474.000	.000	.662	2885.862	1.000
	Hotelling's Trace	1.958	1442.931[b]	2.000	1474.000	.000	.662	2885.862	1.000
	Roy's Largest Root	1.958	1442.931[b]	2.000	1474.000	.000	.662	2885.862	1.000
clotsolv * proc	Pillai's Trace	.005	1.748	4.000	2950.000	.137	.002	6.993	.539
	Wilks' Lambda	.995	1.748[b]	4.000	2948.000	.137	.002	6.992	.539
	Hotelling's Trace	.005	1.747	4.000	2946.000	.137	.002	6.990	.539
	Roy's Largest Root	.004	2.798[c]	2.000	1475.000	.061	.004	5.596	.552

a. Computed using alpha = .05

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept+clotsolv+proc+clotsolv * proc

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	Time to hospital	710.609[b]	5	142.122	96.537	.000	.247	482.683	1.000
	Log-cost	59.379[c]	5	11.876	1022.151	.000	.776	5110.753	1.000
Intercept	Time to hospital	7586.952	1	7586.952	5153.460	.000	.777	5153.460	1.000
	Log-cost	9638.025	1	9638.025	829543.8	.000	.998	829543.822	1.000
clotsolv	Time to hospital	661.902	2	330.951	224.799	.000	.234	449.599	1.000
	Log-cost	.057	2	.028	2.443	.087	.003	4.886	.493
proc	Time to hospital	.380	1	.380	.258	.612	.000	.258	.080
	Log-cost	33.552	1	33.552	2887.814	.000	.662	2887.814	1.000
clotsolv * proc	Time to hospital	2.866	2	1.433	.973	.378	.001	1.947	.220
	Log-cost	.059	2	.030	2.541	.079	.003	5.083	.510
Error	Time to hospital	2171.503	1475	1.472					
	Log-cost	17.137	1475	.012					
Total	Time to hospital	15612.000	1481						
	Log-cost	18656.492	1481						
Corrected Total	Time to hospital	2882.112	1480						
	Log-cost	76.516	1480						

a. Computed using alpha = .05

b. R Squared = .247 (Adjusted R Squared = .244)

c. R Squared = .776 (Adjusted R Squared = .775)

Contrast Results (K Matrix)

Clot-dissolving drugs Simple Contrast[a]			Dependent Variable Time to hospital	Log-cost
Level 2 vs. Level 1	Contrast Estimate		-.365	.022
	Hypothesized Value		0	0
	Difference (Estimate - Hypothesized)		-.365	.022
	Std. Error		.124	.011
	Sig.		.003	.049
	95% Confidence Interval for Difference	Lower Bound	-.608	.000
		Upper Bound	-.122	.043
Level 3 vs. Level 1	Contrast Estimate		-1.675	.024
	Hypothesized Value		0	0
	Difference (Estimate - Hypothesized)		-1.675	.024
	Std. Error		.124	.011
	Sig.		.000	.028
	95% Confidence Interval for Difference	Lower Bound	-1.919	.003
		Upper Bound	-1.432	.046

a. Reference category = 1

Reporting MANOVA

Example 1: Gender in the future

An initial MANOVA examined age and educational literacy as covariates, the three latent variables as dependent variables (DVs), and gender of baby in survey and gender of participant as independent variables (IVs). After excluding age and educational level as nonsignificant, a follow-up MANOVA examined associations between the DVs and IVs described above. It showed a significant multivariate effect for the three latent variables as a group in relation to the gender of the baby in the survey (girl versus boy: $p<.001$) and the gender of the participant completing the survey ($p<.01$). However, the interaction between gender of participant and gender of baby in survey was nonsignificant.

Univariate analyses for the effect of the baby in the survey significant predicted responses related to *consumer trends ($p<.05$), with responses significantly more positive for girl than boy babies.*

Reporting MANOVA

Example 3: ICT

A MANOVA was used to compare the current and preferred means of male and female teachers for the two dimensions of ICT use defined by the instrument, namely: (D1) ICT as a tool for the development of ICT-related skills and the enhancement of curriculum learning outcomes; and (D2) ICT as an integral component of reforms that change what students learn and how school is structured and organised.

The multivariate result was significant for gender, Pillai's Trace = .02, *F = 3.50, df = (4,924), p = .01, indicating a difference in the level of student use of ICT between male and female teachers. The univariate F tests showed there was a significant difference between males and females for D1, F = 7.73, df = (1,927), p = .01, and D2, F = 6.59, df = (1,927), p = .01, with respect to how frequently their students currently use ICT.*

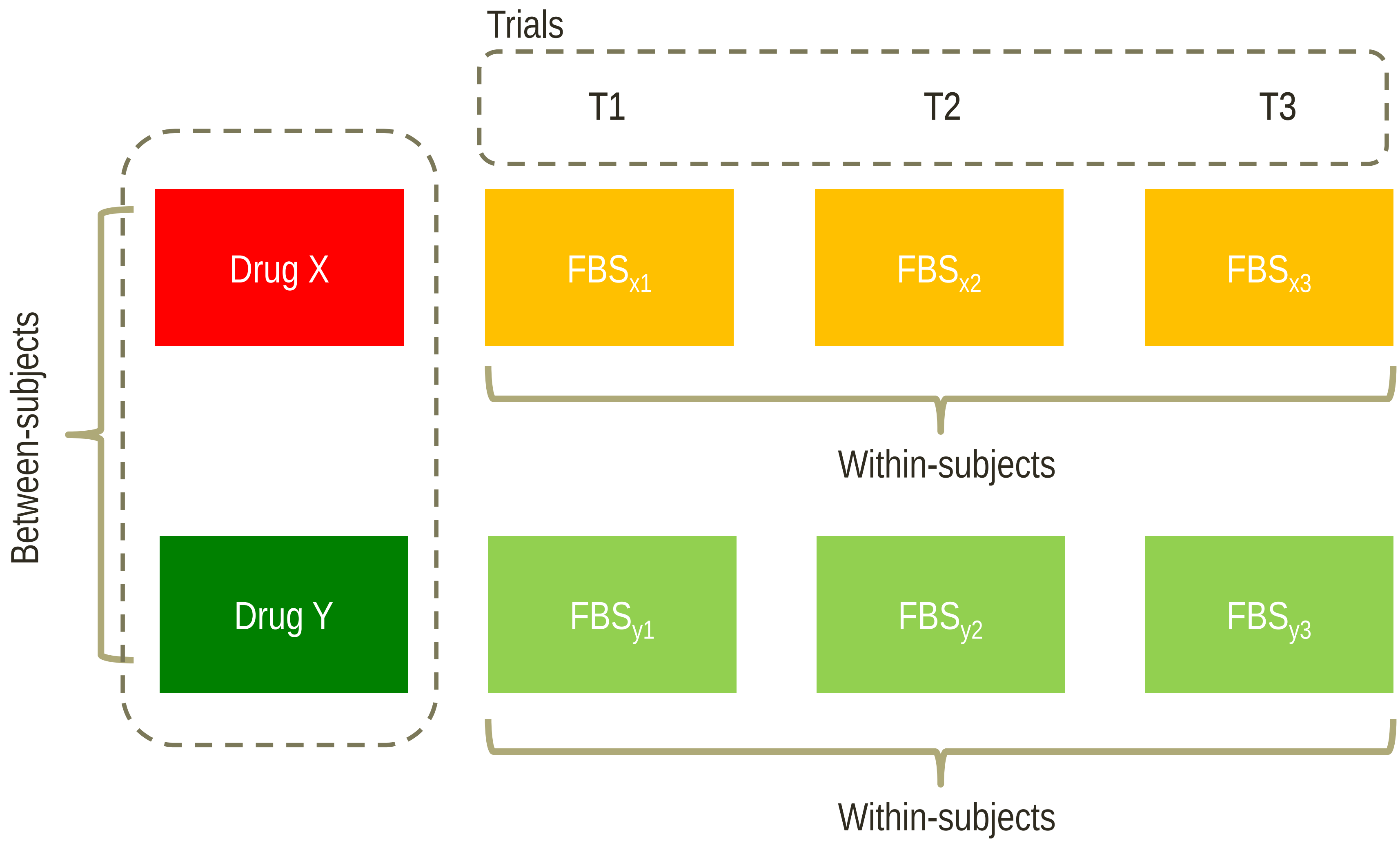
However, the *F tests for both dimensions on the preferred scale were not significant, F = 1.55, df = (1,927), p = .21 for D1, and F = .00, df = (1,927), p = .99 for D 2. Thus, male and female teachers were not significantly different in their preferred level of student use of ICT.*

Repeated Measure GLM

Jamalludin Ab Rahman *MD MPH*
Associate Professor & Head
Department of Community Medicine
Kulliyyah (Faculty) of Medicine

Indication

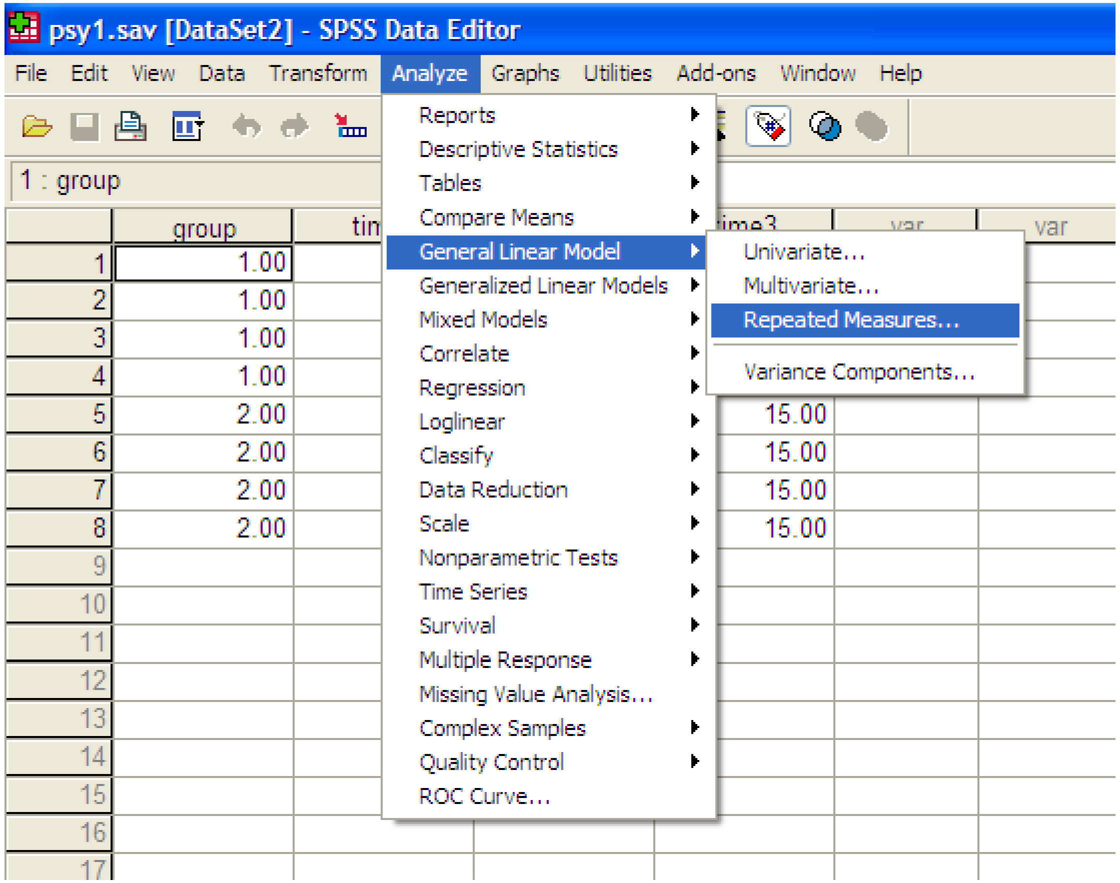
- Multiple related interval dependents
- One or multiple independent (categorical or numerical)


Trials
T1
T2
T3
Between-subjects
Drug X
FBSx1
FBSx2
FBSx3
Within-subjects
Drug Y
FBSy1
FBSy2
FBSy3
Within-subjects

Important terminologies

- **Within-subjects factor** – a dependent measured repeatedly over a set conditions (or *trial*)
- **Between-subjects factor** – a dependent measured on independent groups, & each *group* is exposed to a different conditions

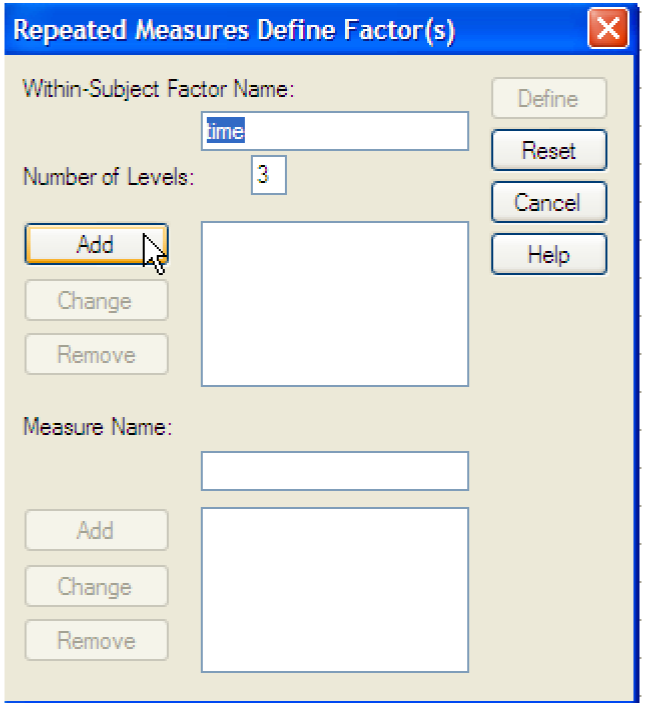
Psy1

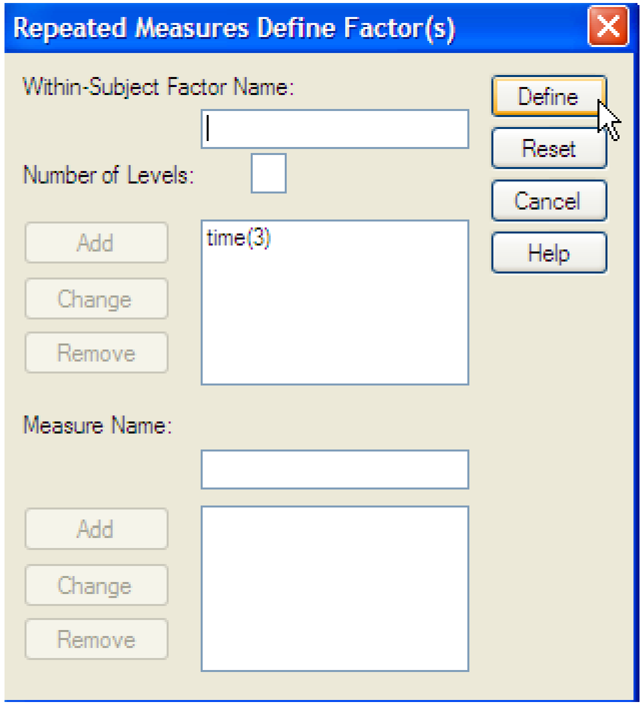


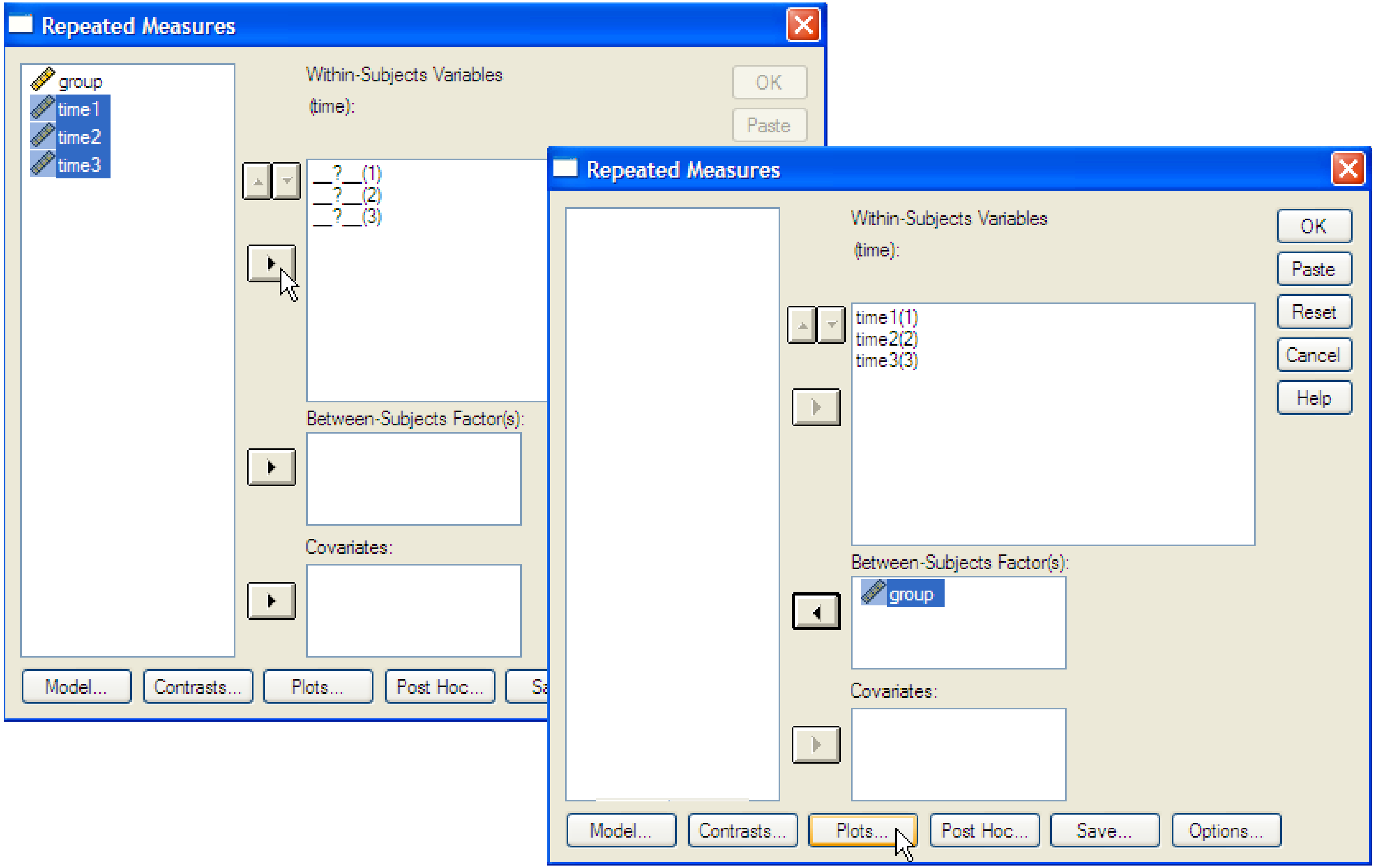
psy1.sav [DataSet2] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : group
group
var
var
1.00
1.00
1.00
1.00
2.00
2.00
2.00
2.00
15.00
15.00
15.00
15.00
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Univariate...
Multivariate...
Repeated Measures...
Variance Components...

Exercise

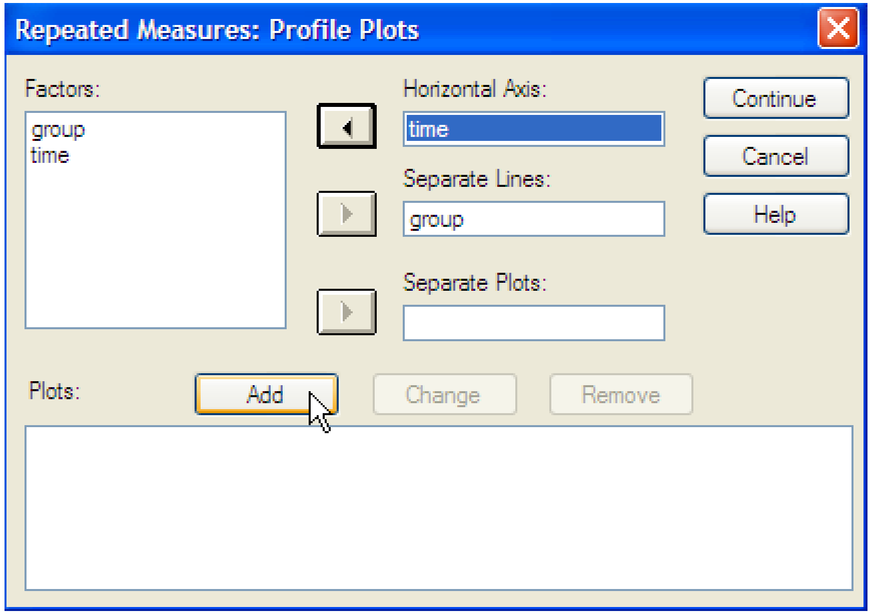
- We want to know whether a new anti-psychosis drug (e.g. seroquel) is better than the old drug (e.g. haloperidol), evaluated on patients for 3 consecutive visits (at 1, 3 & 12 months)

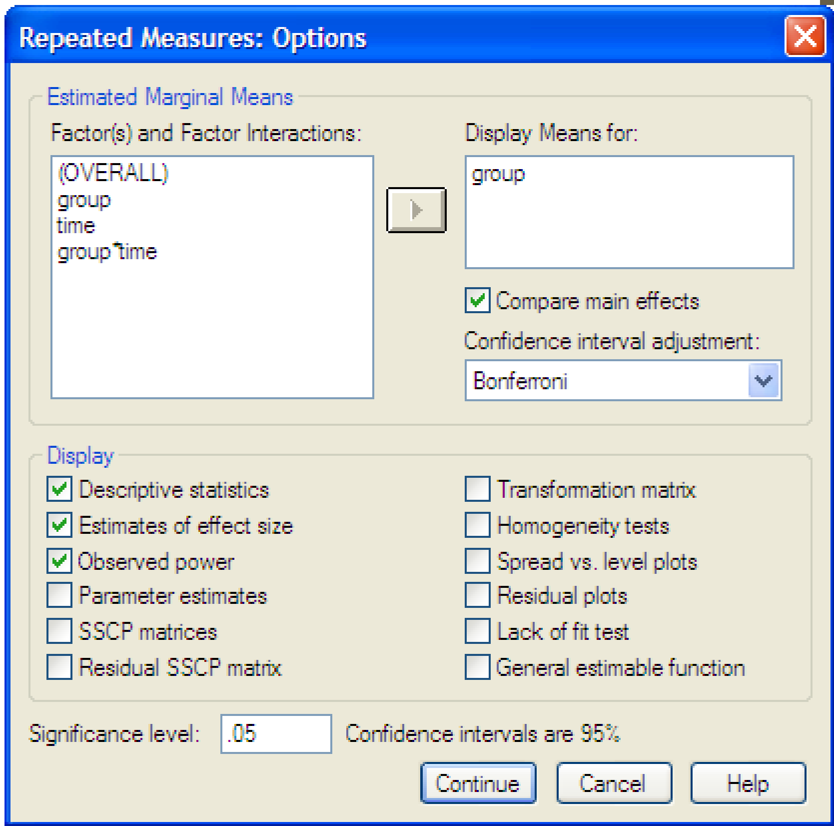

Repeated Measures Define Factor(s)
Within-Subject Factor Name:
time
Number of Levels:
3
Add
Change
Remove
Measure Name:
Add
Change
Remove
Define
Reset
Cancel
Help


Repeated Measures Define Factor(s)
Within-Subject Factor Name:
Number of Levels:
Add
Change
Remove
time(3)
Measure Name:
Add
Change
Remove
Define
Reset
Cancel
Help

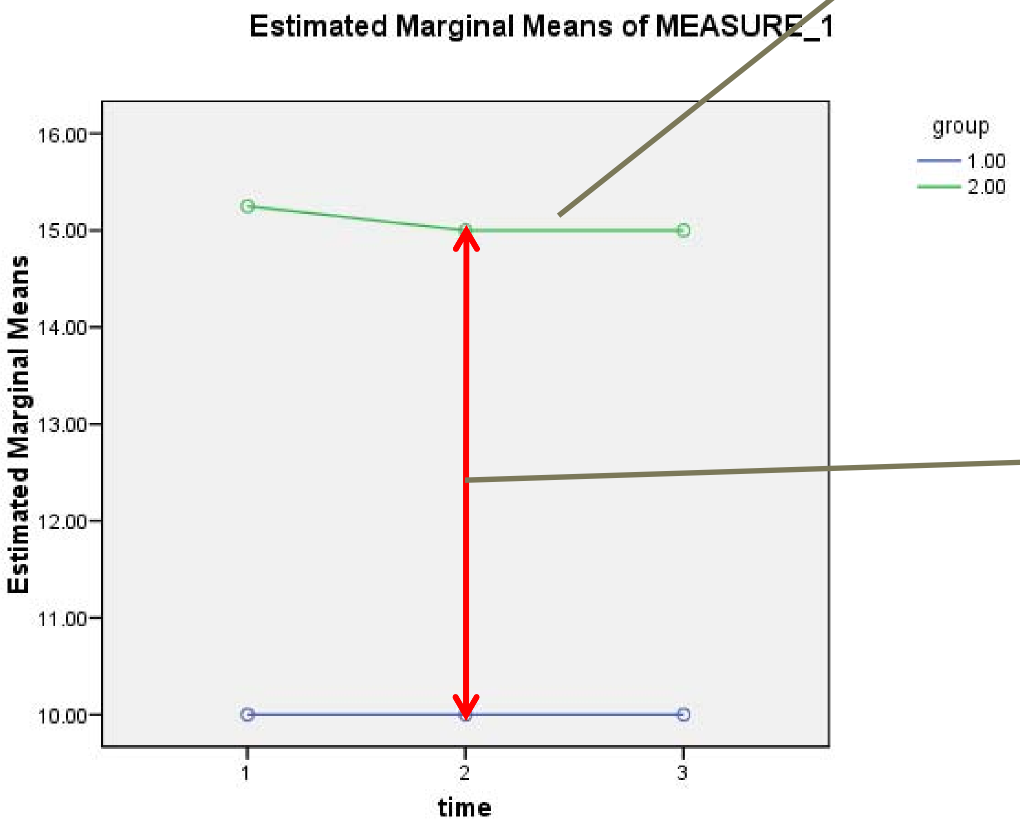


Repeated Measures
group
time1
time2
time3
Within-Subjects Variables
(time):
OK
Paste
_?_(1)
_?_(2)
_?_(3)
Between-Subjects Factor(s):
Covariates:
Model...
Contrasts...
Plots...
Post Hoc...
Repeated Measures
Within-Subjects Variables
(time):
OK
Paste
Reset
Cancel
Help
time1(1)
time2(2)
time3(3)
Between-Subjects Factor(s):
group
Covariates:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...


Repeated Measures: Profile Plots
Factors:
group
time
Horizontal Axis:
time
Separate Lines:
group
Separate Plots:
Continue
Cancel
Help
Plots:
Add
Change
Remove


Repeated Measures: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
group
time
group*time
Display Means for:
group
Compare main effects
Confidence interval adjustment:
Bonferroni
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
SSCP matrices
Residual SSCP matrix
Transformation matrix
Homogeneity tests
Spread vs. level plots
Residual plots
Lack of fit test
General estimable function
Significance level:
.05
Confidence intervals are 95%
Continue
Cancel
Help

Results



The straight lines indicates no significant difference over time

The wide gap indicates significant difference

Between-subjects

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	3775.042	1	3775.042	90601.000	.000	1.000	90601.000	1.000
group	155.042	1	155.042	3721.000	.000	.998	3721.000	1.000
Error	.250	6	.042					

a. Computed using alpha = .05

Estimates

Measure: MEASURE_1

group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	10.000	.059	9.856	10.144
2.00	15.083	.059	14.939	15.228

Within-subjects

Descriptive Statistics

	group	Mean	Std. Deviation	N
time1	1.00	10.0000	.00000	4
	2.00	15.2500	.50000	4
	Total	12.6250	2.82527	8
time2	1.00	10.0000	.00000	4
	2.00	15.0000	.00000	4
	Total	12.5000	2.67261	8
time3	1.00	10.0000	.00000	4
	2.00	15.0000	.00000	4
	Total	12.5000	2.67261	8

Univariate F

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
time	Sphericity Assumed	.083	2	.042	1.000	.397	.143	2.000	.184
	Greenhouse-Geisser	.083	1.000	.083	1.000	.356	.143	1.000	.136
	Huynh-Feldt	.083	1.200	.069	1.000	.368	.143	1.200	.146
	Lower-bound	.083	1.000	.083	1.000	.356	.143	1.000	.136
time * group	Sphericity Assumed	.083	2	.042	1.000	.397	.143	2.000	.184
	Greenhouse-Geisser	.083	1.000	.083	1.000	.356	.143	1.000	.136
	Huynh-Feldt	.083	1.200	.069	1.000	.368	.143	1.200	.146
	Lower-bound	.083	1.000	.083	1.000	.356	.143	1.000	.136
Error(time)	Sphericity Assumed	.500	12	.042					
	Greenhouse-Geisser	.500	6.000	.083					
	Huynh-Feldt	.500	7.200	.069					
	Lower-bound	.500	6.000	.083					

a. Computed using alpha = .05

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.000	.	2	.	.500	.600	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b.
Design: Intercept+group
Within Subjects Design: time

Multivariate F

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
time	Pillai's Trace	.143	1.000[b]	1.000	6.000	.356	.143	1.000	.136
	Wilks' Lambda	.857	1.000[b]	1.000	6.000	.356	.143	1.000	.136
	Hotelling's Trace	.167	1.000[b]	1.000	6.000	.356	.143	1.000	.136
	Roy's Largest Root	.167	1.000[b]	1.000	6.000	.356	.143	1.000	.136
time * group	Pillai's Trace	.143	1.000[b]	1.000	6.000	.356	.143	1.000	.136
	Wilks' Lambda	.857	1.000[b]	1.000	6.000	.356	.143	1.000	.136
	Hotelling's Trace	.167	1.000[b]	1.000	6.000	.356	.143	1.000	.136
	Roy's Largest Root	.167	1.000[b]	1.000	6.000	.356	.143	1.000	.136

a. Computed using alpha = .05

b. Exact statistic

c.
Design: Intercept+group
Within Subjects Design: time

Reporting Repeated Measurement

- A repeated measures one-way ANOVA revealed that there were no significant differences in depression score between the three times of measurement, $F(1,227) = 58.48$, $p < .001$, though this was a relatively small effect size (eta-squared = .21). LSD comparisons revealed that all three means were significantly different from each other. Mean leadership performance was significantly higher immediately after the the leadership training workshop ($M = 6.96$) than before the workshop ($M = 5.92$). The mean performance score three months later was significantly lower ($M = 6.38$) than that immediately after the workshop, but significantly higher than the mean performance before the workshop.

Exercise

- Please repeat the same exercise for psy2, psy3 & psy 4

Exercise 1

- Granumas™ is the new invention of IIUM and used as bone substitution material.
- Use granumas.sav
- Which is better? Granumas alone or granumas with platelet when compared to autograft?

Exercise 2

- A health educationist is interested to introduce a new type (Type 3) of physical exercise to the public. However he knows that the fitness is not only affected by different exercise regime but also different diet.
- Use fitness.sav & prove that his new exercise regime (Type 3) is better than the previous types of exercise.

Exercise 3

- The systemic inflammatory response following cardiopulmonary bypass (CABG) is well known and has been described extensively in adults but not in children. Porcine model, 30 neonatal pigs (aged 17-19 days) were randomized to either sternotomy alone (group 2: sham group, n=15) or to sternotomy and CABG (group 1: CPB group, n=15). Platelet - , Leukocyte count and EVF (Erythrocyte Volume Fraction) were measured at:
  - t=1 Baseline, immediately after anesthesia
  - t=2 Pre-CPB, after sternotomy
  - t=3 5 min after aortic cross clamp release
  - t=4 5 min after weaning off CPB
  - t=5 2 hours post CPB (-platelet)
  - t=6 4 hours post CPB(-platelet)
- Summarize and analyze the data? So does EVF increased significantly following CABG?

Logistic Regression

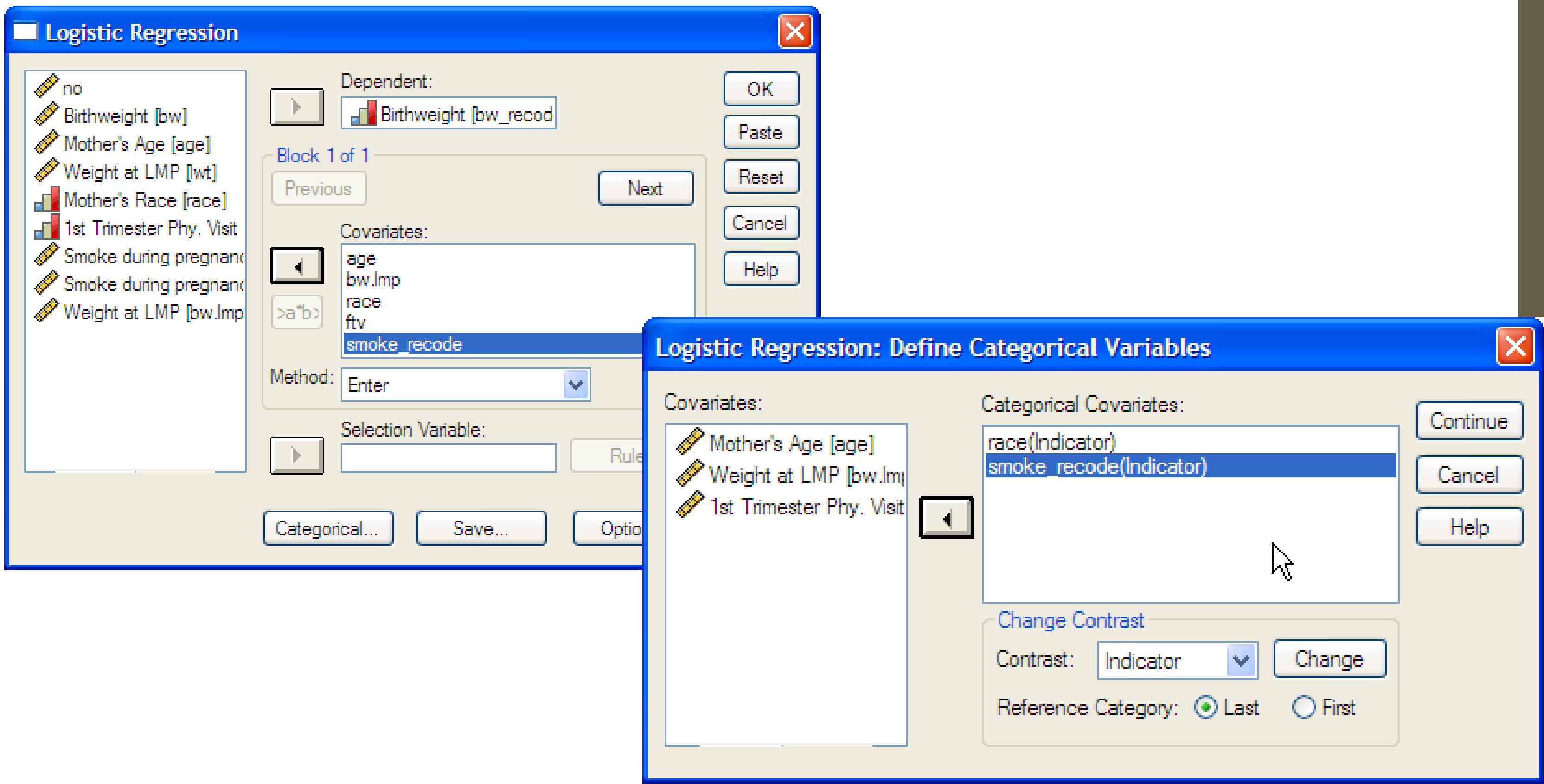
Jamalludin Ab Rahman MD MPH
Department of Community Medicine
International Islamic University Malaysia

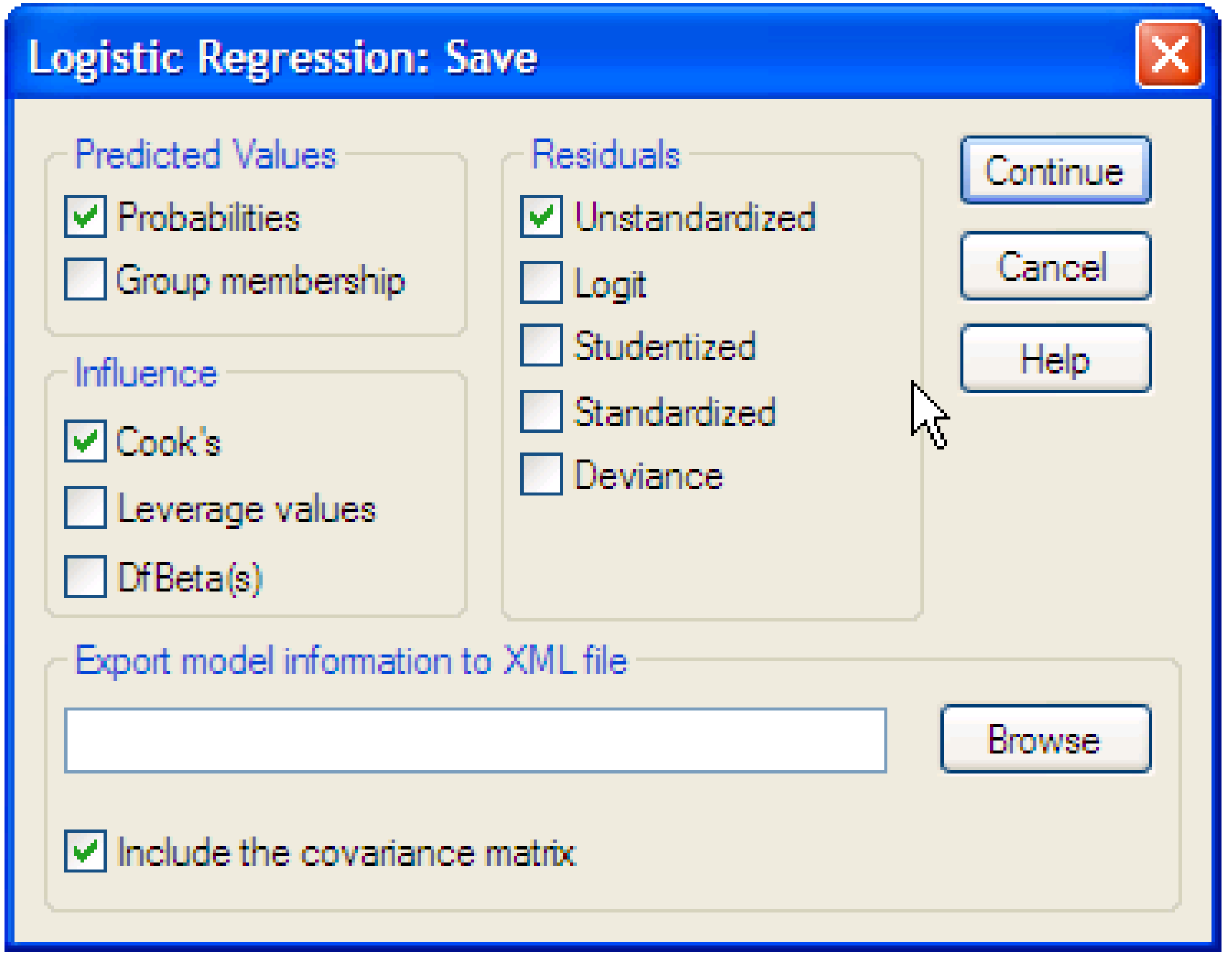
Logistic Regression

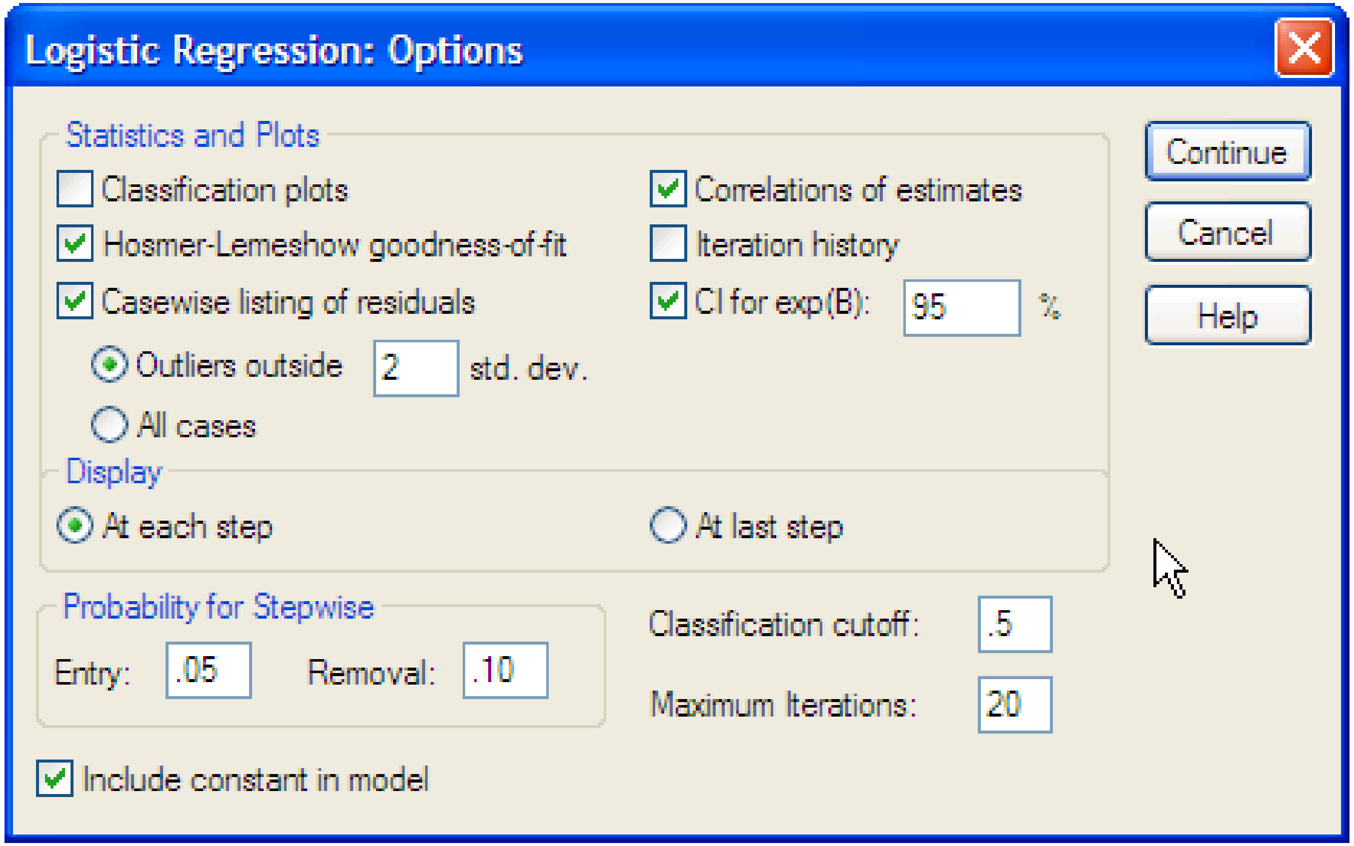
- When the outcome is binary
  e.g. What are the significant predictors for low birth weight baby
- Possible factors
  1. Mother’s age
  2. Mother’s body weight at LMP
  3. Mother’s race
  4. First trimester antenatal booking
  5. Smoking

- Open LowBirthWeight.sav
- In this study we wish to study factors that associated with low birth weight

Binary Logistic Regression




Logistic Regression: Save
Predicted Values
Probabilities
Group membership
Influence
Cook's
Leverage values
DfBeta(s)
Residuals
Unstandardized
Logit
Studentized
Standardized
Deviance
Continue
Cancel
Help
Export model information to XML file
Browse
Include the covariance matrix


Logistic Regression: Options
Statistics and Plots
Classification plots
Hosmer-Lemeshow goodness-of-fit
Casewise listing of residuals
Outliers outside
2
std. dev.
All cases
Correlations of estimates
Iteration history
CI for exp(B):
95
%
Display
At each step
At last step
Probability for Stepwise
Entry:
.05
Removal:
.10
Classification cutoff:
.5
Maximum Iterations:
20
Include constant in model
Continue
Cancel
Help

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	189	100.0
	Missing Cases	0	.0
	Total	189	100.0
Unselected Cases		0	.0
Total		189	100.0

a. If weight is in effect, see classification table for the total number of cases.

Categorical Variables Codings

		Frequency	Parameter coding	
			(1)	(2)
Mother's Race	Malay	96	1.000	.000
	Chinese	26	.000	1.000
	Indian	67	.000	.000
Smoke during pregnancy	No	115	1.000	
	Yes	74	.000	

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	20.097	6	.003
	Block	20.097	6	.003
	Model	20.097	6	.003

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	214.575[a]	.101	.142

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11.826	8	.159

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	age	-.022	.035	.410	1	.522	.978	.914	1.047
	bw.Imp	-.027	.014	3.764	1	.052	.973	.946	1.000
	race			7.789	2	.020			
	race(1)	-.942	.418	5.074	1	.024	.390	.172	.885
	race(2)	.289	.527	.301	1	.583	1.336	.475	3.755
	ftv	-.008	.164	.002	1	.963	.992	.719	1.369
	smoke_recode(1)	-1.053	.381	7.642	1	.006	.349	.165	.736
	Constant	2.324	1.053	4.869	1	.027	10.212		

a. Variable(s) entered on step 1: age, bw.Imp, race, ftv, smoke_recode.

Correlation Matrix

		Constant	age	bw.Imp	race(1)	race(2)	ftv	smoke_ recode(1)
Step 1	Constant	1.000	-.612	-.593	-.160	-.022	.133	-.258
	age	-.612	1.000	-.142	-.070	.093	-.167	.002
	bw.Imp	-.593	-.142	1.000	-.106	-.300	-.121	-.044
	race(1)	-.160	-.070	-.106	1.000	.420	-.088	.483
	race(2)	-.022	.093	-.300	.420	1.000	-.041	.179
	ftv	.133	-.167	-.121	-.088	-.041	1.000	-.068
	smoke_recode(1)	-.258	.002	-.044	.483	.179	-.068	1.000

Exercise

- What factors significantly associated with cancer?
- Use cancer.sav