Data Management & Analysis: Intermediate PASW Topics II Workshop

PASW STATISTICS V 17.0 (SPSS FOR WINDOWS)

Beginning, Intermediate & Advanced Applied Statistics

Zayed University Office of Research SPSS for Windows®
Workshop Series

Presented by

Dr. Maher Khelifa
Associate Professor
Department of Humanities and Social Sciences
College of Arts and Sciences

PASW Statistics 17 (2).lnk

Data Analysis Using SPSS

MEASURING RELATIONSHIPS BETWEEN VARIABLES (CORRELATIONS)

Topics to be Covered

- Correlational vs. Regression model.
- Correlational model
  - Bivariate Correlations
    - Pearson Product Moment Correlation Coefficient
    - Spearman Rank Order Correlations
    - Kendall’s tau
  - Partial correlation
    - Partial correlations & Zero-order correlations

Correlational Analysis

- Correlational analysis: Any statistical procedures used for assessing the extent of a relationship between 2 variables.
- A correlational index is typically used to identify the extent to which two variables correlate or co-vary.
- The correlational method can be applied when data have been collected on two different variables X & Y.
- Each case has 2 scores, one on X and one on Y.

Correlational Analysis

- The correlational model is different from the regression model.
  - Correlational model
    - Seeks to establish relationships.
    - Does not distinguish between variables X & Y.
  - Regression model
    - Seeks prediction and causality
      - Causal relationships are investigated using experimental research.
    - Distinguishes between X (predictor) and Y (criterion) variables.
      - Changes in X are used to predict changes in Y.

BIVARIATE CORRELATIONS

Bivariate Correlations

- Bivariate correlations explores pairwise relationships for a set of scale or ordinal variables.

- In Bivariate Correlations the test measures if variables or rank orders are related. Results are displayed in a correlation matrix.

- In SPSS, depending on the level of measurement, three different tests of significance can be computed under Bivariate Correlations procedure:
  - Pearson Product Moment Correlation Coefficient,
  - Spearman Rank Order Correlations, and
  - Kendall's tau.

- Bivariate tests of significance are very useful in determining the strength and direction of associations between variables (for scale and ordinal variables only).

THE PEARSON PRODUCT MOMENT CORRELATION COEFFICIENT

The Pearson Product-Moment Correlation (*r*)

- The Pearson Product-Moment Correlation Coefficient (*r*) assesses the degree that two quantitative variables are linearly related.

- The *r* provides information on the direction and magnitude of an observed correlation between two variables (X and Y).

- $rxy = \frac{\sum zxi\ zyi}{n-1}$
- This formula requires transforming X and Y to Z scores, then summing the cross-products over all individuals in the sample. The sum is divided by n-1.
- $z = \frac{x - \bar{x}}{s}$

The Pearson Product-Moment Correlation (*r*)

- The correlation coefficient, *r*, ranges from -1 to +1.

- **Value of r Interpretation**

$r= 0$	The two variables do not vary together at all.
$0 > r > 1$	The two variables tend to increase or decrease together.
$r = 1.0$	Perfect correlation.
$-1 > r > 0$	One variable increases as the other decreases.
$r = -1.0$	Perfect negative or inverse correlation.

The Pearson Product-Moment Correlation (*r*)

- Guidelines for interpreting *r*

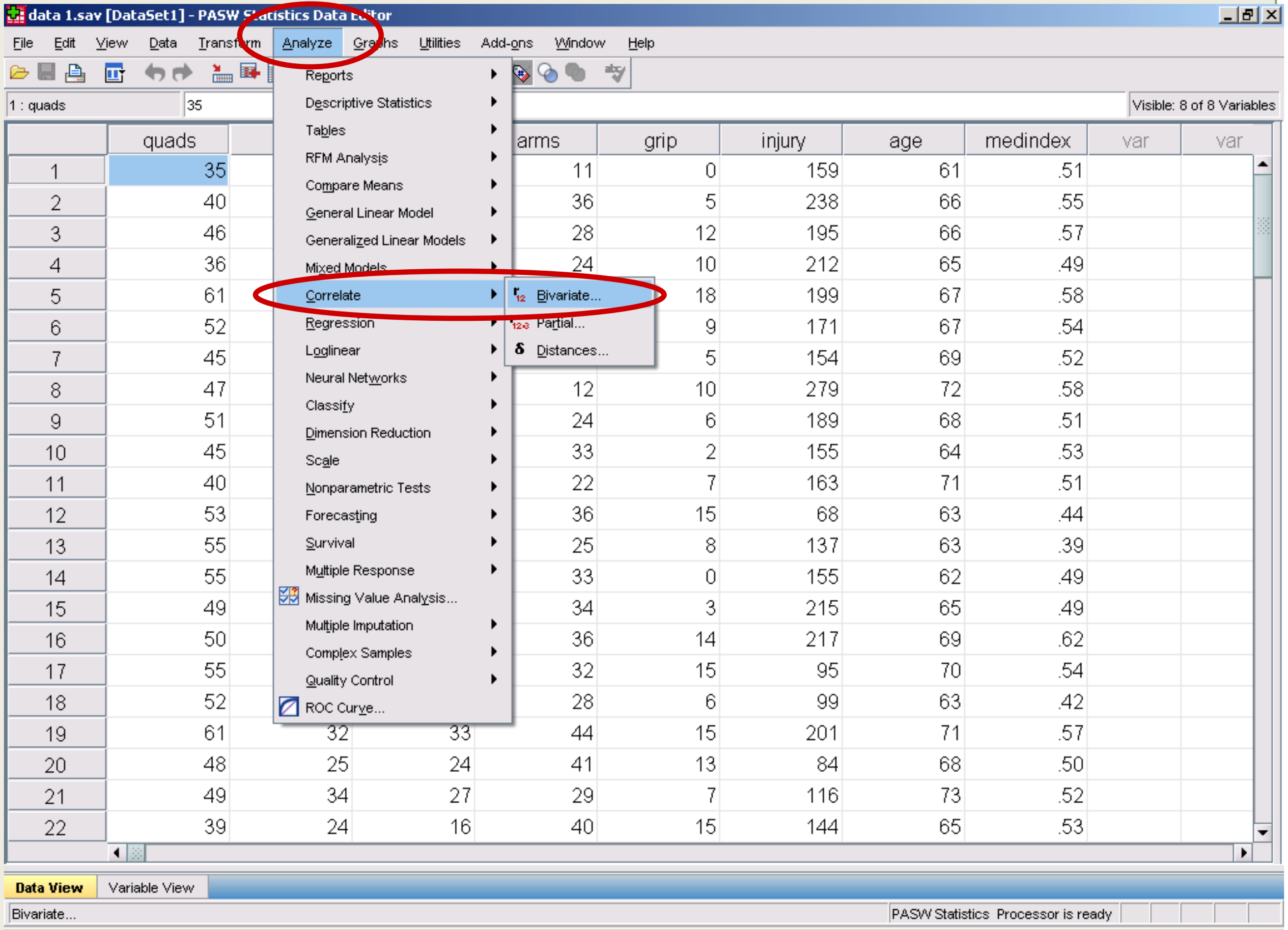
Range of Positive Coefficient	Type of Relationship	Range of Negative Coefficient
.00 to .20	low	.00 to -.20
.21 to .40	Low to moderate	-.21 to -.40
.41 to .60	Moderate	-.41 to -.60
.61 to .80	Moderate to high	-.61 to -.80
.81 to 1.00	High	-.81 to -1.00

•What is large or small relationship depends on the researcher's discipline.

•In Social Sciences correlation coefficients of .10, .30, and .50 (irrespective of sign) are interpreted as small, medium, and large relationships, respectively.

How to Obtain a Person *r*

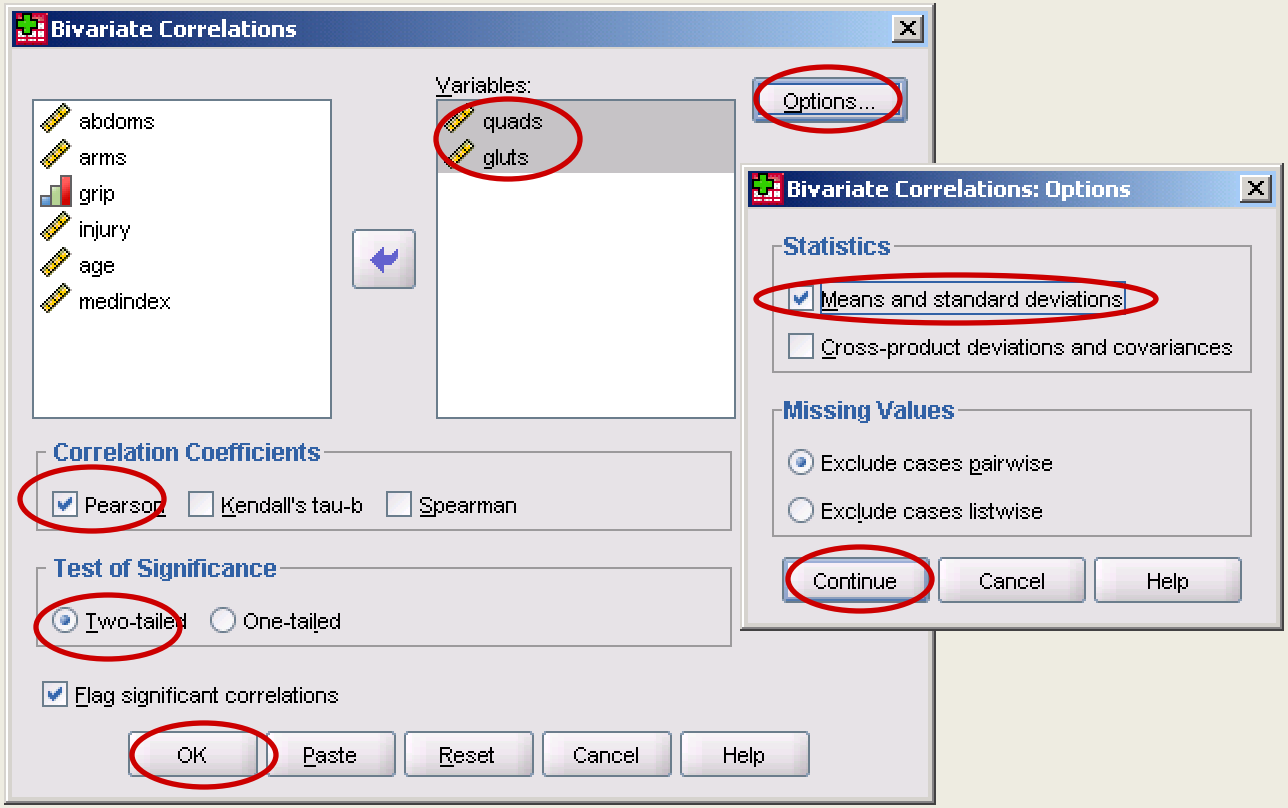
- Go to Analyze, Correlate, Bivariate.
- A correlation dialogue box will appear.



	quads			arms	grip	injury	age	medindex	var	var
1	35			11	0	159	61	.51		
2	40			36	5	238	66	.55		
3	46			28	12	195	66	.57		
4	36			24	10	212	65	.49		
5	61				18	199	67	.58		
6	52				9	171	67	.54		
7	45				5	154	69	.52		
8	47			12	10	279	72	.58		
9	51			24	6	189	68	.51		
10	45			33	2	155	64	.53		
11	40			22	7	163	71	.51		
12	53			36	15	68	63	.44		
13	55			25	8	137	63	.39		
14	55			33	0	155	62	.49		
15	49			34	3	215	65	.49		
16	50			36	14	217	69	.62		
17	55			32	15	95	70	.54		
18	52			28	6	99	63	.42		
19	61	32	33	44	15	201	71	.57		
20	48	25	24	41	13	84	68	.50		
21	49	34	27	29	7	116	73	.52		
22	39	24	16	40	15	144	65	.53		

How to Obtain a Person *r*

- Move the variables you want to correlate to the variable box. Then select Pearson.
- Select options and select means and SD and press continue.
- Press OK.



Bivariate Correlations

abdoms
arms
grip
injury
age
medindex

Variables:

quads
gluts

Options...

Correlation Coefficients

☑ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance

● Two-tailed ○ One-tailed

☑ Flag significant correlations

OK | Paste | Reset | Cancel | Help

Bivariate Correlations: Options

Statistics

☑ Means and standard deviations

☐ Cross-product deviations and covariances

Missing Values

● Exclude cases pairwise

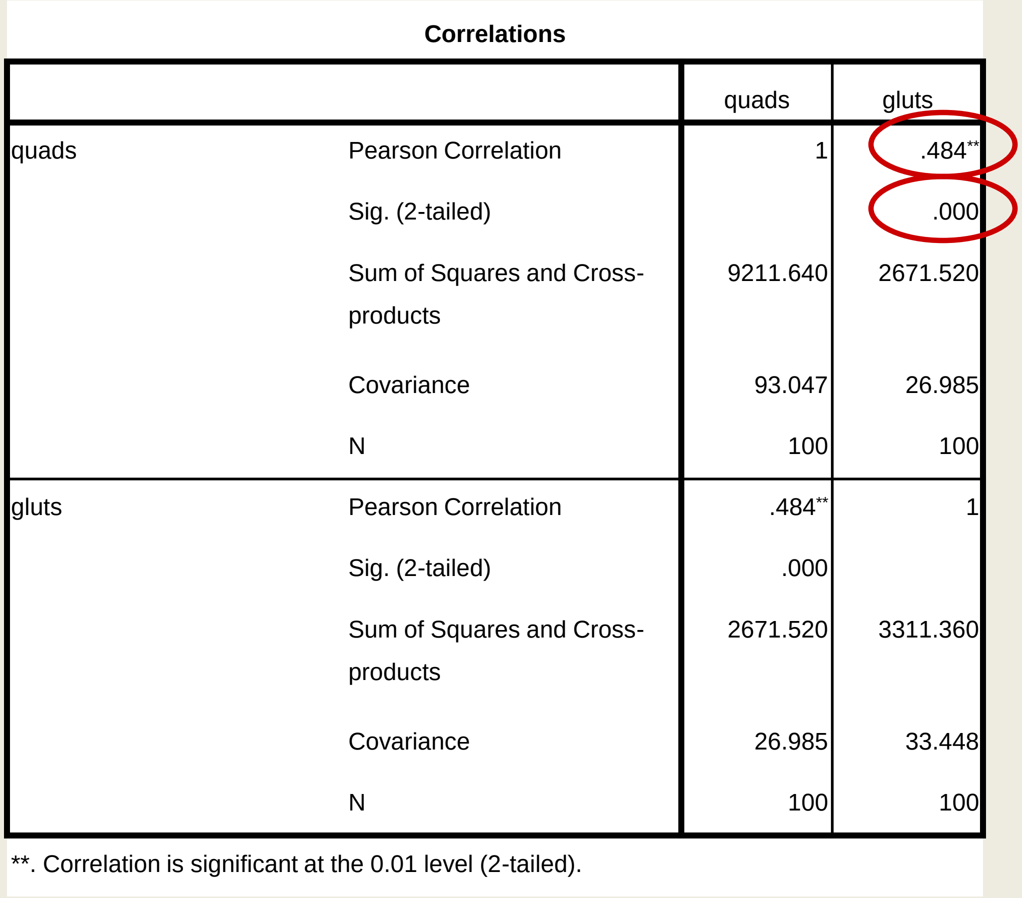
○ Exclude cases listwise

Continue | Cancel | Help

How to Obtain a Person *r*

A correlation matrix will appear in the Output window.

Check the Pearson correlation coefficient and the level of significance.



Correlations

		quads	gluts
quads	Pearson Correlation	1	.484**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	9211.640	2671.520
	Covariance	93.047	26.985
	N	100	100
gluts	Pearson Correlation	.484**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	2671.520	3311.360
	Covariance	26.985	33.448
	N	100	100

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

Effect Size

- To properly interpret the value of the Pearson *r, the r* is squared to calculate r^2 (*r* squared) also called the *coefficient of determination*.

- *r^2 ranges from 0 to 1.*

- *r^2 is the* variance shared between X and Y.
  - For example, if r^2 =0.59, then 59% of the variance in X can be explained by variation in Y, and 59% of the variance in Y can be explained by variation in X.

  - So *r^2 is the amount of variance accounted for by the linear relationship of the 2 variables.*

Effect Size

- The Pearson *r* scale values are not of equal intervals.
- The correlation scale is ordinal not interval or ratio.
- For example:
  - A value of r = .80 does not represent twice the relationship as r = .40.
  - The difference in a relationship of .40 to .50 does not represent the same magnitude of relationship difference as .70 to .80.
- If we use r^2 descriptive comparative statements can be made.
- The r^2 provides an added dimension for interpreting and understanding the magnitude of a linear relationship.

Assumptions of the Pearson *r*

- There are 2 assumptions underlying the Pearson *r* test of significance.
- 1. Assumption of Independence:
  - Cases represent a random sample from the population and the scores on variables for one case are independent of scores of other cases.
    - The test is not robust to violations of this assumption and therefore the Pearson *r* test of significance should not be computed.

Assumptions of the Pearson *r*

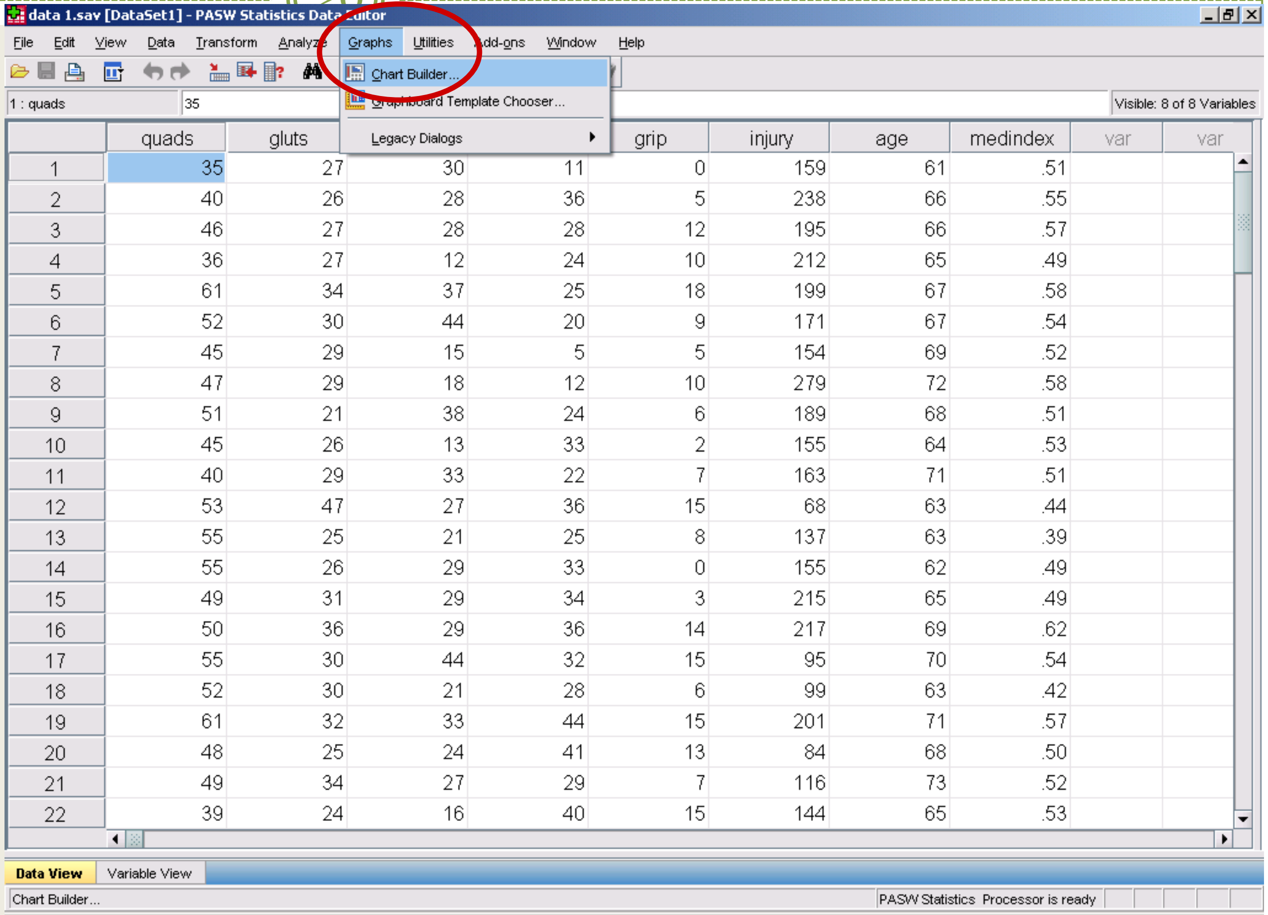
- 2. Normality Assumptions:
- The variables are bivariately normally distributed.
  - Pearson correlation calculations are based on the assumption that both X and Y values are sampled from populations that follow a normal distribution, at least approximately.
  - the Pearson correlation coefficient works best when the variables are approximately normally distributed and have no outliers.
  - With large samples, the test is robust to violations of this assumption.
  - If the normality assumption is met, the only type of statistical relationship that can exist between 2 variables is a linear association.

Assumptions of the Pearson *r*

- 2. Normality Assumptions: (Continued)
- The Pearson's *r* is a measure of linear association only. However, two variables can be perfectly related but not in a linear fashion. The Pearson's correlation coefficient is not an appropriate statistic for measuring non-linear associations.
- Before calculating *r*, data should be screened for outliers and evidence of a linear relationship. A scatterplot can reveal possible problems.
- Outliers may cause misleading results.
- If the normality assumption is not made or met, alternate distribution-free tests should be used instead.

How to Obtain A Scatter Plot

- Go to Graphs,
- Chart builder.



1 : quads 35

Visible: 8 of 8 Variables

	quads	gluts			grip	injury	age	medindex	var	var
1	35	27	30	11	0	159	61	.51		
2	40	26	28	36	5	238	66	.55		
3	46	27	28	28	12	195	66	.57		
4	36	27	12	24	10	212	65	.49		
5	61	34	37	25	18	199	67	.58		
6	52	30	44	20	9	171	67	.54		
7	45	29	15	5	5	154	69	.52		
8	47	29	18	12	10	279	72	.58		
9	51	21	38	24	6	189	68	.51		
10	45	26	13	33	2	155	64	.53		
11	40	29	33	22	7	163	71	.51		
12	53	47	27	36	15	68	63	.44		
13	55	25	21	25	8	137	63	.39		
14	55	26	29	33	0	155	62	.49		
15	49	31	29	34	3	215	65	.49		
16	50	36	29	36	14	217	69	.62		
17	55	30	44	32	15	95	70	.54		
18	52	30	21	28	6	99	63	.42		
19	61	32	33	44	15	201	71	.57		
20	48	25	24	41	13	84	68	.50		
21	49	34	27	29	7	116	73	.52		
22	39	24	16	40	15	144	65	.53		

Chart Builder...

PASW Statistics Processor is ready

How to Obtain A Scatter Plot

- From Gallery, chose Scatter/dot, then drag and drop the type of graph you want into the Chart preview

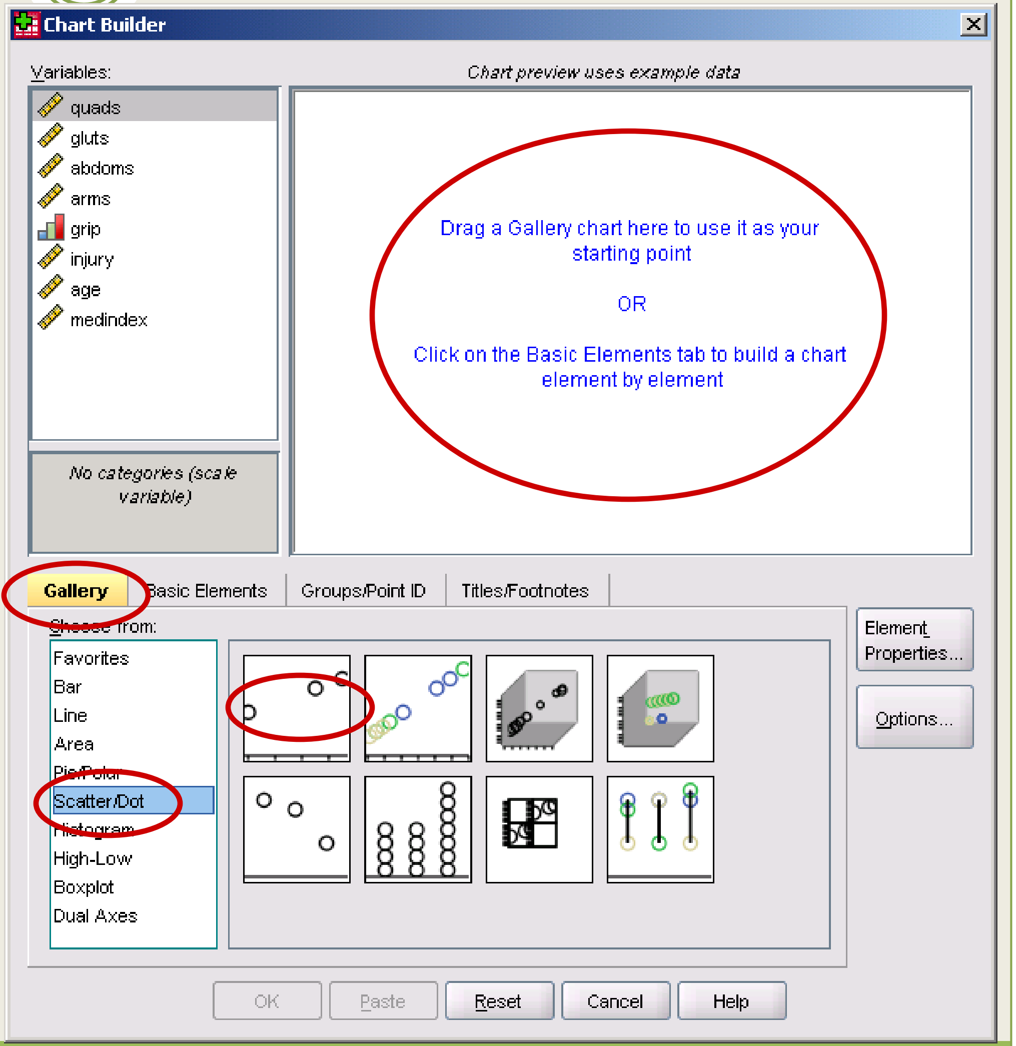


Chart Builder

Variables:

- quads
- gluts
- abdoms
- arms
- grip
- injury
- age
- medindex

No categories (scale variable)

Chart preview uses example data

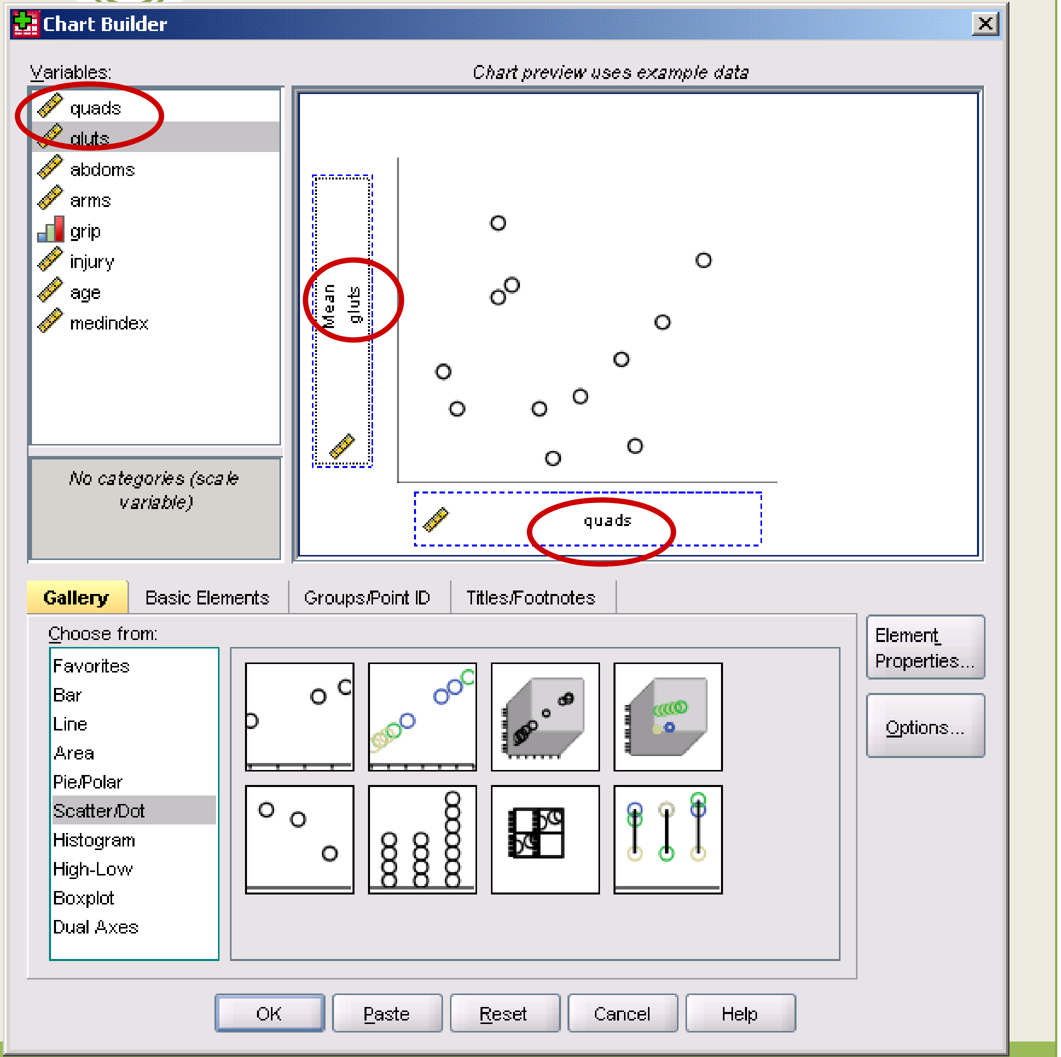
Drag a Gallery chart here to use it as your starting point

OR

Click on the Basic Elements tab to build a chart element by element

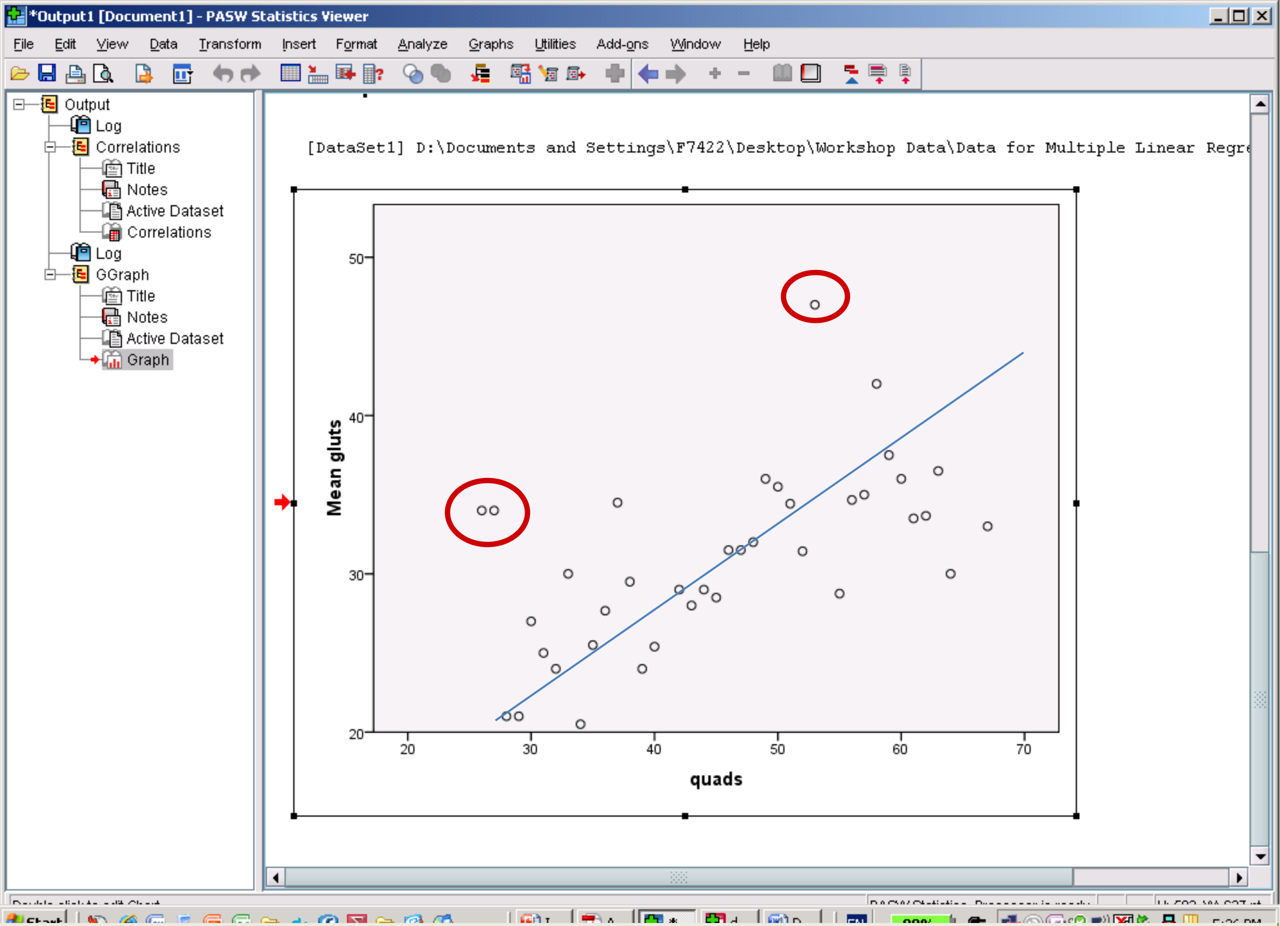
How to Obtain A Scatter Plot

- Drag and drop the selected variables to correlate into the X and Y axes. Then Press OK.



How to Obtain A Scatter Plot

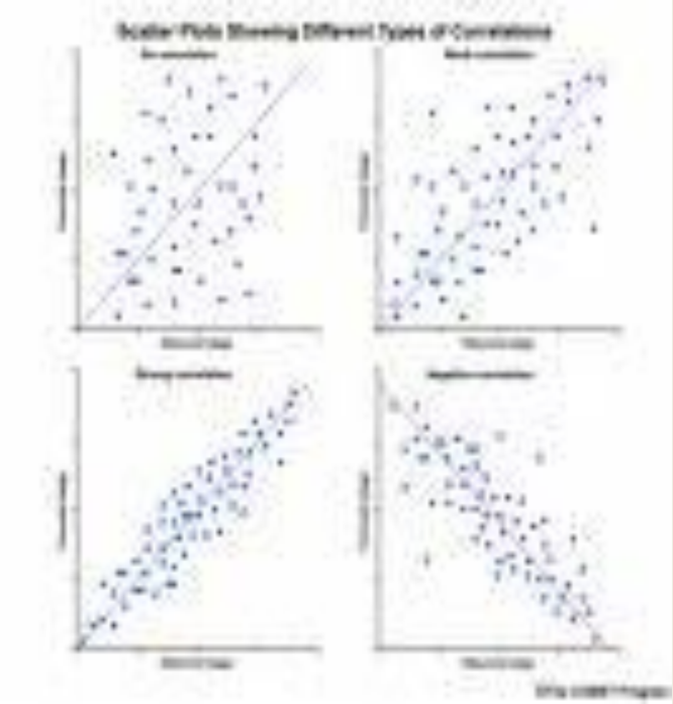
- A scatter plot is obtained in the Output window.
- Look for outliers and serious deviations from a linear relationship pattern.

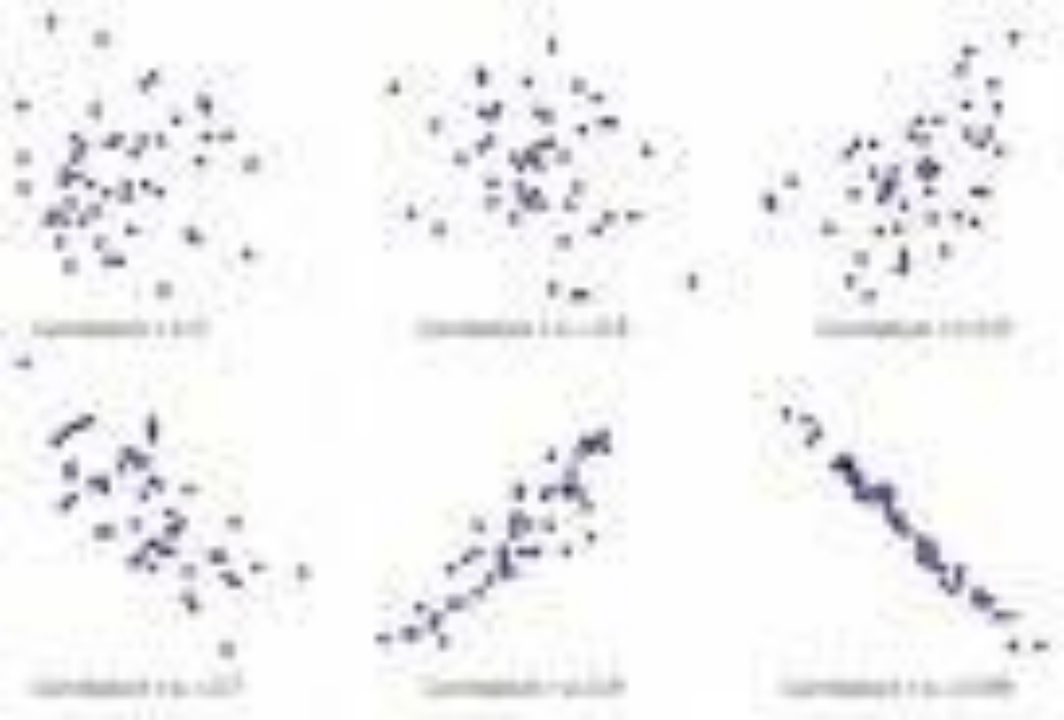


*Output1 [Document1] - PASW Statistics Viewer

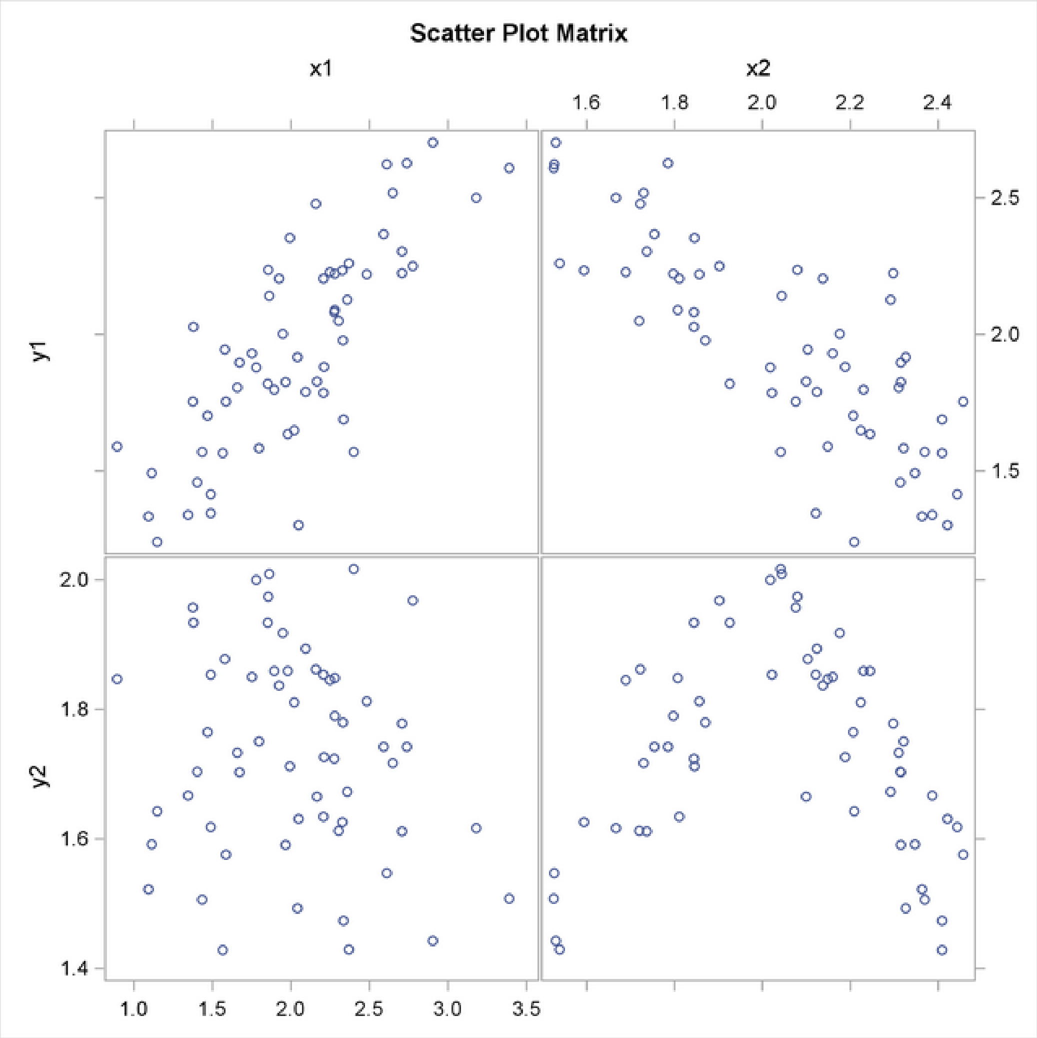
[DataSet1] D:\Documents and Settings\F7422\Desktop\Workshop Data\Data for Multiple Linear Regre

Example of Relationship Scatterplots

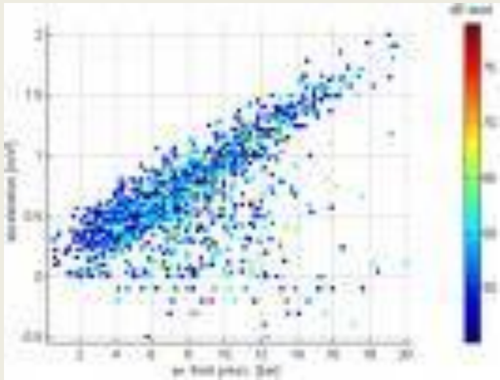




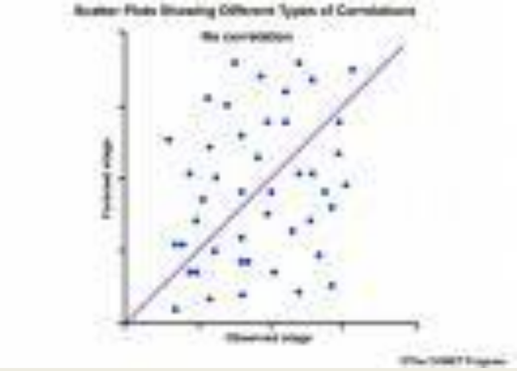
Example of Relationship Scatterplots



Example of Relationship Scatterplots



•If a relationship is less than perfect, the scatterplot of data points will deviate from a perfect straight line.



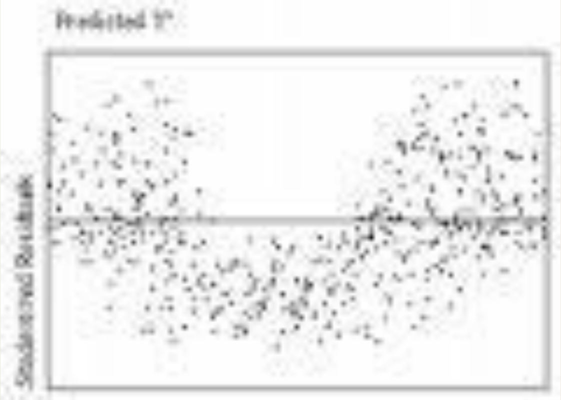
Non Linear Relationships

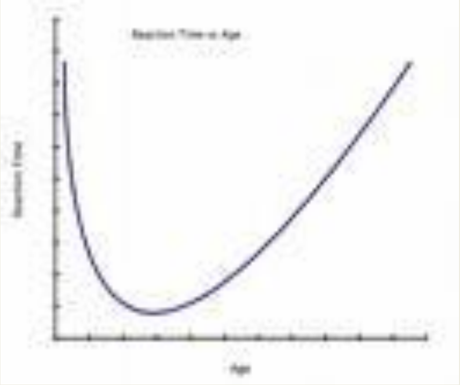
- Two variables can be perfectly related, but in a non-linear fashion.

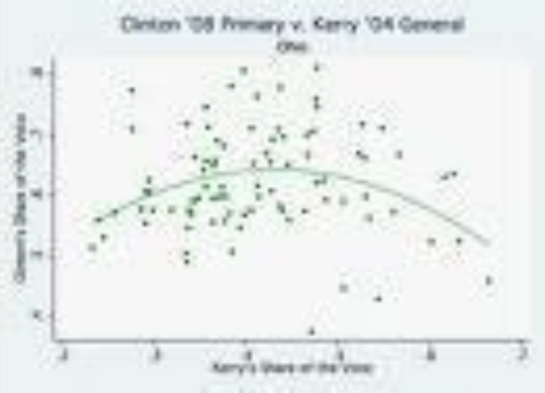
- When this is the case, the Pearson's correlation coefficient is not an appropriate statistic for measuring their association.

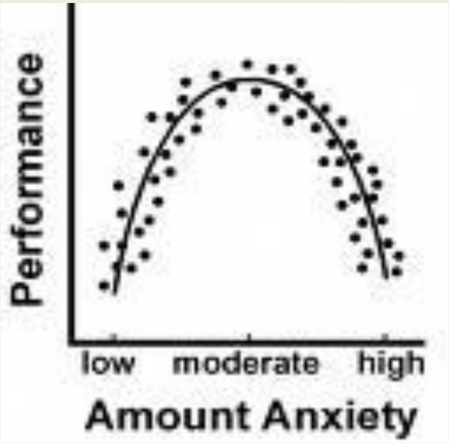
- Scan the plots for relationships other than linear.

- The attached plots indicate non-linear relationships (curvilinear).









Bivariate Correlation

THE SPEARMAN RANK CORRELATION COEFFICIENT & THE KENDALL'S RANK CORRELATION COEFFICIENT

The Spearman's rho & Kendall's tau-*b*

- If the data are not normally distributed, or have ordered categories, the Spearman's rho and Kendall's tau-*b* statistics may offer good alternatives to the Pearson *r*.

- The Spearman's rho and Kendall's tau-b measure the association between rank-orders (ranked variables).

- The Spearman's rho and Kendall's tau-b work regardless of the distributions of the variables. They are considered distribution-free or non-parametric tests.

- Spearman's rho and Kendall's tau-b can be computed on quantitative variables or variables with ordered categories, i.e.:
  - on 2 scale variables,
  - on 2 ordinal variables, on
  - on one scale and one ordinal variable.

The Spearman's rho (r_s)

- The Spearman correlation coefficient is a correlation coefficient between 2 ranked variables.

- Raw scores in the 2 variables are converted into ranks and differences between the ranks for each observation on the 2 variables are calculated using the following formula.

$$(R) = 1 - \frac{6\sum d^2}{n^3 - n}$$

Σ

Spearman rho: Example

$$(R) = 1 - \frac{6\sum d^2}{n^3 - n}$$

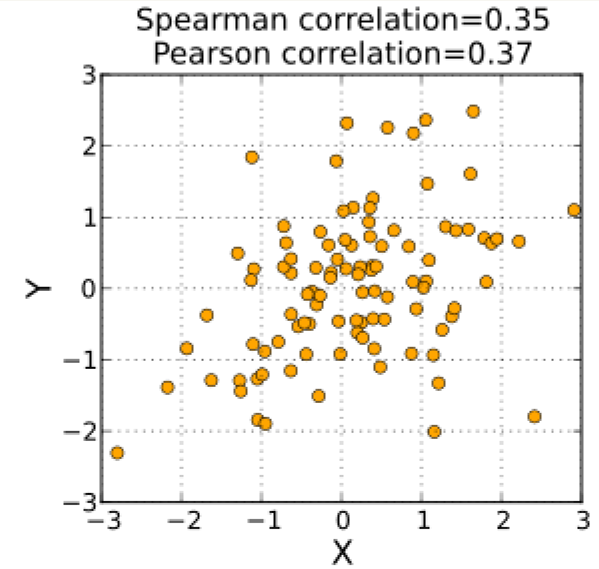
Convenience Store	Distance from CAM (m)	Rank variable distance	Price of 50cl bottle (€)	Rank variable price	Difference between ranks (d)	d²
1	50	10	1.80	2	8	64
2	175	9	1.20	3.5	5.5	30.25
3	270	8	2.00	1	7	49
4	375	7	1.00	6	1	1
5	425	6	1.00	6	0	0
6	580	5	1.20	3.5	1.5	2.25
7	710	4	0.80	9	-5	25
8	790	3	0.60	10	-7	49
9	890	2	1.00	6	-4	16
10	980	1	0.85	8	-7	49
					d^2 = 285.5	

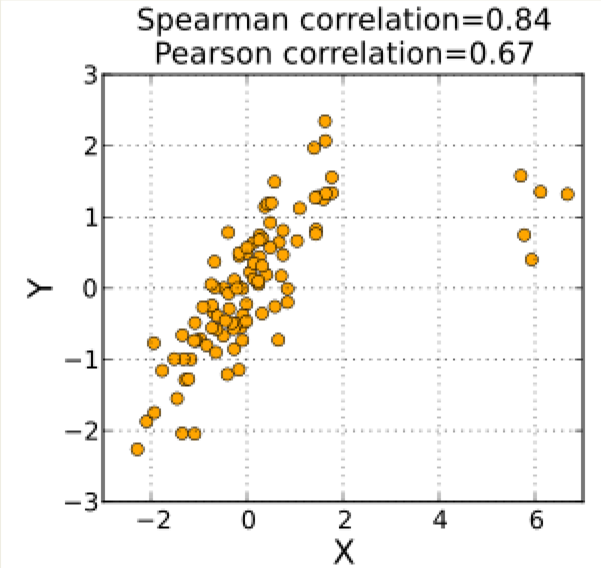
Interpreting the Spearman rho

- The Spearman rho varies between -1 and +1 and is interpreted in the same fashion as the Pearson *r*.

Spearman rho & Outliers

- Outliers have less of an effect on Spearman's rho.





The Kendall’s tau-*b*

- Like the Spearman rho, the Kendall's tau (τ) coefficient is a distribution-free statistic used to measure the statistical association between two variables.
- The Kendall's tau (τ) is a measure of rank correlation.
- The test examines the similarity of the orderings of the ranked data.
  - If the two variable rankings are identical, the coefficient has value 1.
  - If one ranking is the perfect reverse of the other, the coefficient has value –1.
  - Increasing coefficient values indicate more concordance between the rankings.
  - If the coefficient has value of 0, this implies that the rankings are completely independent.

Kendall's tau-*b*

- $\tau = \dfrac{n_c - n_d}{n(n-1)/2}$

- *Where* n_c is the number of concordant pairs, and

n_d is the number of discordant pairs in the data set.

- The range of the Kendall's tau is -1 to +1, and the tau coefficient is interpreted the same way as the Spearman rho and Pearson r.

How to obtain the Spearman's rho and Kendall's tau-*b*

- In PASW, Spearman's rho and Kendall's tau-b are obtained the same way as the Pearson *r*. Just select the test you are interested in (Pearson, rho, and tau-*b*).

- PASW allows you to select all tests at the same time. The three tests generally yield similar alphas (not identical values though).

Comapring the Bivariate Tests

- The Pearson *r* is more powerful if the data satisfies the test assumptions (normality and independence).

- More predictive models are available for linear relationships, and the linear models are generally easier to implement and interpret.

- The Pearson *r* is also used extensively in measurement, in test construction, and in research to determine instrument reliability:
  - test-retest reliability,
  - equivalence reliability, and
  - internal consistency reliability.

- The distribution-free measures of associations are handy for discovering whether there is any kind of association between two variables with ranked data.

- They distribution-free measures of associations are especially useful if the data is not normally distributed or there are outliers.

Alternate tests to Bivariate Correlations

- The Bivariate Correlation procedures discussed above are useful for studying the pairwise associations for a set of scale or ordinal variables.

- If the variables are nominal, use the Crosstabs procedure to obtain measures of association including Chi square, contingency coefficient, and Phi and Cramer’s V.

- If you want to predict the value of a scale variable based on its linear relationship to other variable(s), use Linear Regression procedure.

- If you want to explore the variation in your data to look for underlying patterns, use Factor Analysis.

Partial Correlations

Partial Correlations (r_p)

- Any time the relationship between 2 variables are examined, one must be concerned with the effects of other variables on the relationship of interest.
- Example: Education and income (controlling age and work experience).

- Partial correlation coefficients is a procedure that describes the linear relationship between two variables while controlling for the effects of one or more additional variables.

- r_p is an effect size index which indicates the degree that two variables are linearly related in a sample, partialling out the effects of one or more variables.

Partial Correlations (r_p)

- The partial correlation attempts to estimate the correlation between 2 variables holding constant the values of the control variable(s).

- The significance test evaluates whether in the population the partial correlation is equal to zero.

- To interpret properly the partial correlations results, zero-order correlations must be considered.

- The data file for a partial correlation should contain scores on at least 3 variables. All the variables should be at the scale level of measurement.

- Results of the partial correlation, like the Pearson r, range in value between -1 and +1 and are interpreted in a similar manner.

Usefulness of Partial Correlation

- Partial correlation is very useful for:
  - uncovering spurious relationships, and
  - Detecting hidden relationships

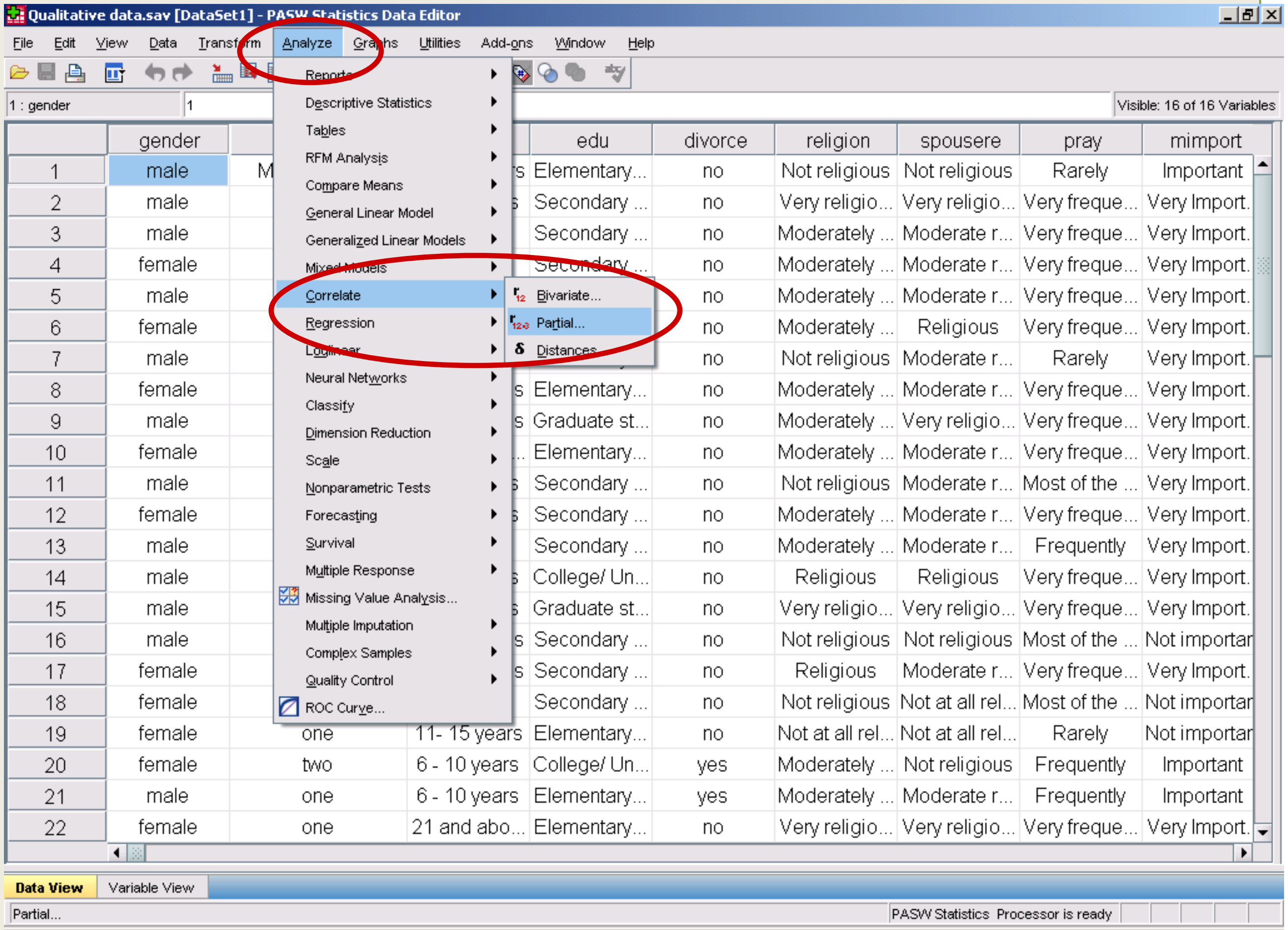
- Spurious correlation: When two or more variables are statistically related but are not in fact causally linked—usually because the statistical relation is caused by a third variable (usually the true predictor). When the effects of the third variable are removed, they are said to have been partialed out.

- Spurious correlations mean that 2 things look like they cause each other, but in reality they don’t (a correlation between A & B is actually caused by a third variable C, usually referred to as lurking variable, confounding and moderator variable).

- Hidden correlation: Theory, intuition and common sense may suggest that there should be a relationship between two variables even though the data suggests no correlation. It may be the case that one or more variables are suppressing the expected relationship.

How to obtain a Partial Correlation

- Select Analyze, Correlate, Partial



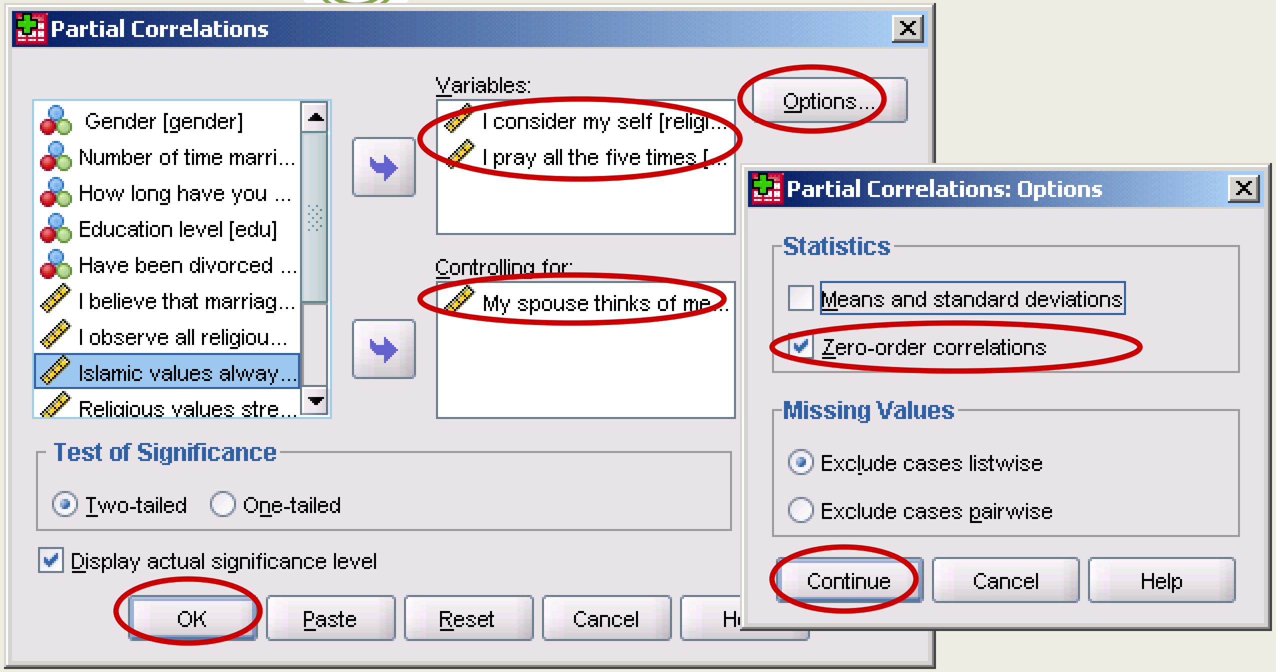
1 : gender | 1 | Visible: 16 of 16 Variables

	gender			edu	divorce	religion	spousere	pray	mimport
1	male	M		Elementary...	no	Not religious	Not religious	Rarely	Important
2	male			Secondary ...	no	Very religio...	Very religio...	Very freque...	Very Import.
3	male			Secondary ...	no	Moderately ...	Moderate r...	Very freque...	Very Import.
4	female			Secondary ...	no	Moderately ...	Moderate r...	Very freque...	Very Import.
5	male				no	Moderately ...	Moderate r...	Very freque...	Very Import.
6	female				no	Moderately ...	Religious	Very freque...	Very Import.
7	male				no	Not religious	Moderate r...	Rarely	Very Import.
8	female			Elementary...	no	Moderately ...	Moderate r...	Very freque...	Very Import.
9	male			Graduate st...	no	Moderately ...	Very religio...	Very freque...	Very Import.
10	female			Elementary...	no	Moderately ...	Moderate r...	Very freque...	Very Import.
11	male			Secondary ...	no	Not religious	Moderate r...	Most of the ...	Very Import.
12	female			Secondary ...	no	Moderately ...	Moderate r...	Very freque...	Very Import.
13	male			Secondary ...	no	Moderately ...	Moderate r...	Frequently	Very Import.
14	male			College/ Un...	no	Religious	Religious	Very freque...	Very Import.
15	male			Graduate st...	no	Very religio...	Very religio...	Very freque...	Very Import.
16	male			Secondary ...	no	Not religious	Not religious	Most of the ...	Not importar
17	female			Secondary ...	no	Religious	Moderate r...	Very freque...	Very Import.
18	female			Secondary ...	no	Not religious	Not at all rel...	Most of the ...	Not importar
19	female	one	11- 15 years	Elementary...	no	Not at all rel...	Not at all rel...	Rarely	Not importar
20	female	two	6 - 10 years	College/ Un...	yes	Moderately ...	Not religious	Frequently	Important
21	male	one	6 - 10 years	Elementary...	yes	Moderately ...	Moderate r...	Frequently	Important
22	female	one	21 and abo...	Elementary...	no	Very religio...	Very religio...	Very freque...	Very Import.

Partial... | PASW Statistics Processor is ready

How to obtain a Partial Correlation

- Move the 2 variables to correlate to the variable box.
- Select the control variable.
- Select Options, and select zero-order correlations



Partial Correlations

- Gender [gender]
- Number of time marri...
- How long have you ...
- Education level [edu]
- Have been divorced ...
- I believe that marriag...
- I observe all religiou...
- Islamic values alway...
- Religious values stre...

Variables:

- I consider my self [religi...
- I pray all the five times [...

Controlling for:

- My spouse thinks of me...

Test of Significance

● Two-tailed ○ One-tailed

☑ Display actual significance level

Partial Correlations: Options

Statistics

□ Means and standard deviations

☑ Zero-order correlations

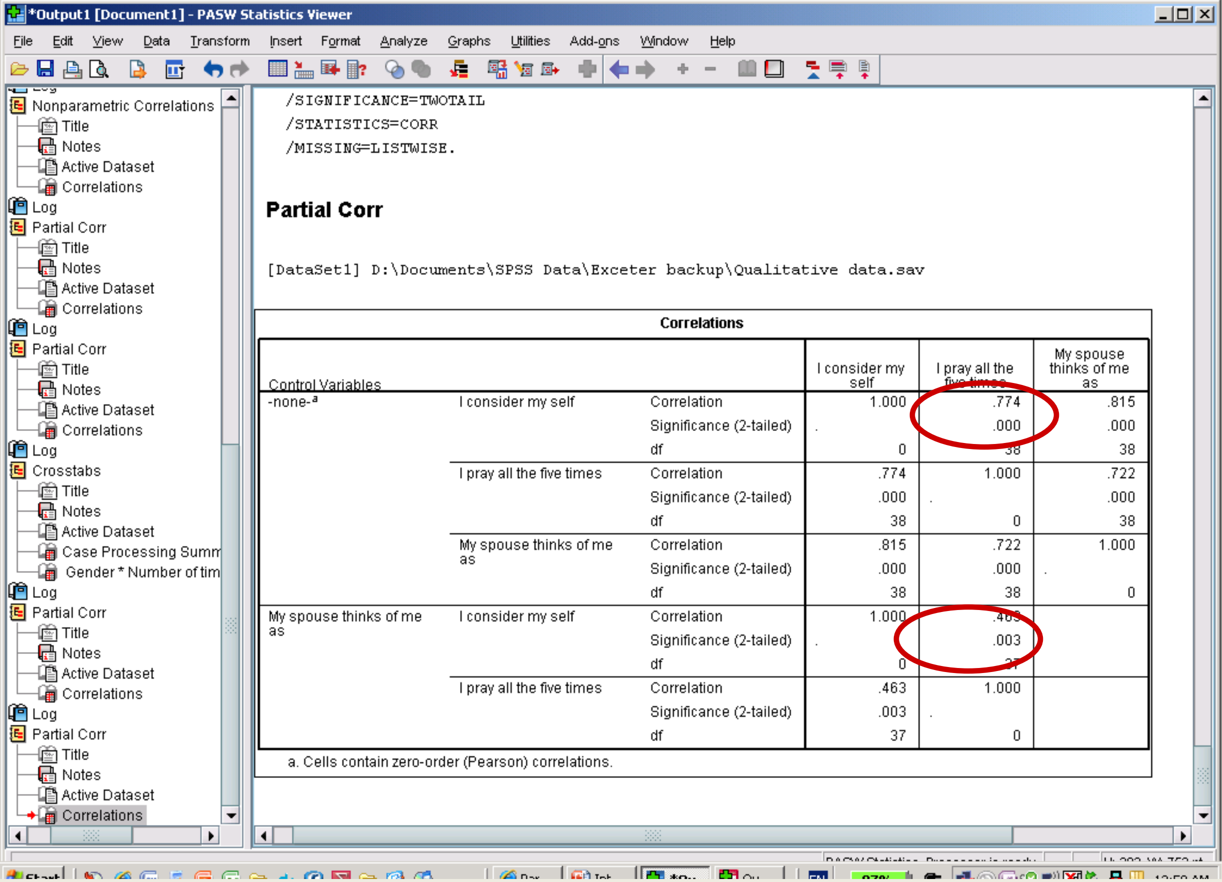
Missing Values

● Exclude cases listwise

○ Exclude cases pairwise

How to obtain a Partial Correlation

- A Partial correlation matrix appears in the Output Window.
- Look for zero-order correlations and compare them to the values of the partial correlation.



```
  /SIGNIFICANCE=TWOTAIL
  /STATISTICS=CORR
  /MISSING=LISTWISE.
```

Partial Corr

```
[DataSet1] D:\Documents\SPSS Data\Exceter backup\Qualitative data.sav
```

Correlations

Control Variables			I consider my self	I pray all the five times	My spouse thinks of me as
-none-[a]	I consider my self	Correlation	1.000	.774	.815
		Significance (2-tailed)	.	.000	.000
		df	0	38	38
	I pray all the five times	Correlation	.774	1.000	.722
		Significance (2-tailed)	.000	.	.000
		df	38	0	38
	My spouse thinks of me as	Correlation	.815	.722	1.000
		Significance (2-tailed)	.000	.000	.
		df	38	38	0
My spouse thinks of me as	I consider my self	Correlation	1.000	.463	
		Significance (2-tailed)	.	.003	
		df	0	37	
	I pray all the five times	Correlation	.463	1.000	
		Significance (2-tailed)	.003	.	
		df	37	0	

a. Cells contain zero-order (Pearson) correlations.

Alternate tests

- The Partial Correlations procedure is only appropriate for scale variables.

- If you have categorical (nominal or ordinal) data, use the Crosstabs procedure. Layer variables in Crosstabs are similar to control variables in Partial Correlations.

- If you want to predict the value of a scale variable based on its linear relationship to other variables, try Linear Regression.

Thank you for your Attention