# SPSS: Pearson's Correlation

Application: To test for a linear relationship between two quantitative variables.

Research Hypothesis: Knowing that store owners are often over-worked, the researcher hypothesized that stores with fewer fish would have healthier fish (thus predicting a negative or inverse relationship between these variables in this population).

H0: There is no linear relationship between the number of fish displayed in pet stores and the quality rating of the fish in the population of pet stores represented by this sample.

Analyze ➔ Correlate ➔ Bivariate
- Highlight each of the desired variables and click the arrow button
- Be sure "Pearson" and "Two-tailed" buttons are checked
- "Options" — check that you want "Means and standard deviations"
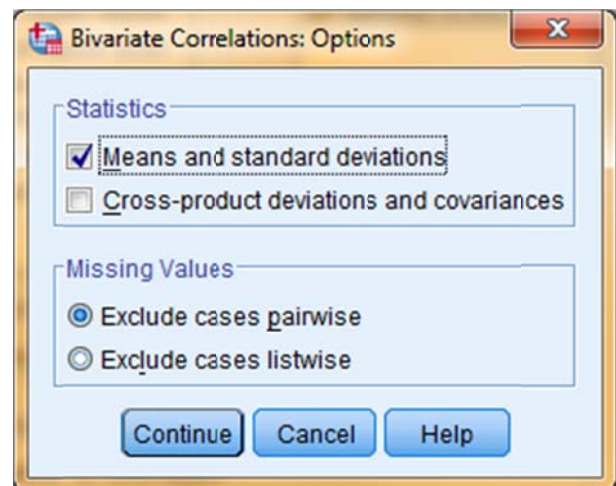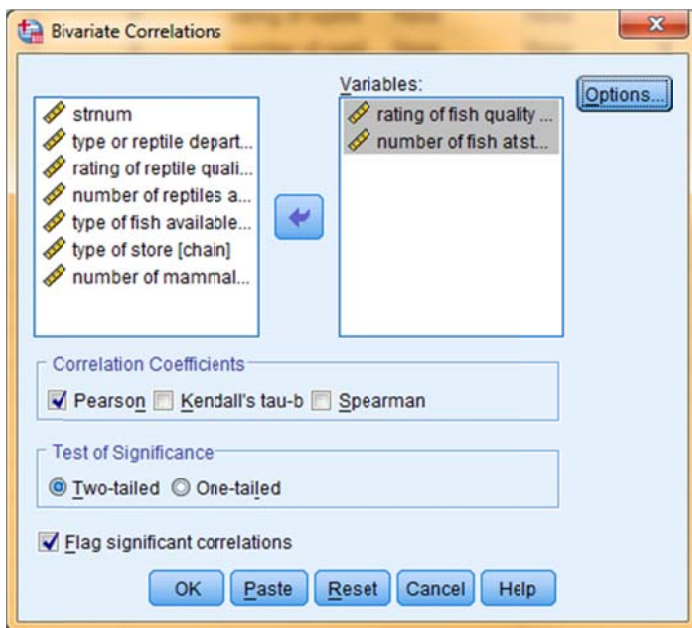
SPSS Syntax

```
CORRELATIONS
 /VARIABLES=fishgood fishnum        ← list the variables
 /PRINT=TWOTAIL                     ← get 2-tailed p-values (alternative is "ONETAIL")
 /STATISTICS DESCRIPTIVES           ← get means and standard deviations for variables
 /MISSING=LISTWISE.                 ← use listwise deletion (alternative is "PAIRWISE")
```

Descriptive Statistics

	Mean	Std. Deviation	N
rating of fish quality - 1-10 scale	6.67	2.146	12
number of fish at store	23.92	9.605	12

Correlations

		rating of fish quality - 1-10 scale	number of fish at store
rating of fish quality - 1-10 scale	Pearson Correlation	1	-.857
	Sig. (2-tailed)		.000
	N	12	12
number of fish at store	Pearson Correlation	-.857	1
	Sig. (2-tailed)	.000	
	N	12	12

p-value

SPSS presents a "correlation matrix"- notice:

- The correlation between a variable and itself is perfect (1.00)

- The correlation matrix is symmetrical (look at the upper right and lower left correlations)

The correlation of -.857 indicates that there is a significant **negative** linear relationship between these variables -- as the number of fish increases, the quality of those fish tends to decrease

SPSS reports the sample size (N), but you should report the degrees of freedom (df = N - 2)

Remember, even if the printout shows it, never report p = .000, because that would suggest there is no possibility of a Type 1 error. Instead report "p<.001"

Reporting the Results:

Table 1.
Univariate data summary (n=12)

Variable	Mean	Std
Number of fish at store	23.92	9.61
Rating of fish quality	6.67	2.51

Univariate statistics are presented in Table 1. Pearson's correlation between the number of fish displayed in these stores and the quality rating for the fish was r(10) = -.86, p<.05. This result supports the research hypothesis that those stores with fewer fish tended to have healthier fish, whereas those stores with more fish tended to have fish with lower health quality.

Reporting the Results:

Pearson's correlation between the number of fish displayed in these stores (M = 23.92, S = 9.61)and the quality rating for the fish (M = 6.67, S = 2.15) was r(10) = -.86, p<.05. This result supports the research hypothesis that those stores with fewer fish tended to have healthier fish, whereas those stores with more fish tended to have fish with lower health quality.

It is important to report the univariate statistics for the variables before presenting the correlation and the significant test. Often these are presented in a table.

When reporting correlations, try to describe the results so that a reader with limited statistical expertise and knowledge of the variables will understand how the variables are related.

Saying, "Number of fish and fish quality are significantly negatively correlated." or "Number of fish and fish quality have a significant inverse linear relationship." are statistically correct, but are less informative than phrasing like that in the example.

Also, as in the example, be sure to communicate:
- The research hypothesis (if there is one)
- The statistical results
- Whether or not those results support the research hypothesis

Sometimes, especially where there are only a few variables and correlations, the univariate statistics are presented in text, along with the correlation results.