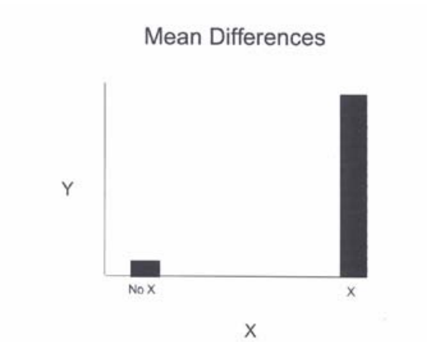
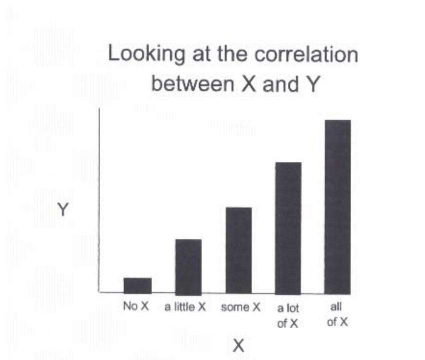
t-Tests, Chi-squares, Phi, Correlations: It's all the same stuff

In this handout, I provide illustrations of the connection between t-tests and correlations and between chi-square and correlation.

Correlation equivalents

Correlation is a statistic that describes the association between two variables. The correlation statistic can be used for continuous variables or binary variables or a combination of continuous and binary variables. In contrast, t-tests examine whether there are significant differences between two group means. With a t-test, we have binary independent variable (two groups, which could be coded 0 and 1) and a continuous dependent variable. If our study is an experiment, then a significant t-test comparing experimental group and control would suggest that our independent variable has a significant impact (and, therefore association with) the dependent variable. Significant group differences then imply a correlation between the independent and dependent variable. The graphs below illustrate that, even if the independent variable has few values, you can still observe a tendency for the dependent variable to increase in value as the independent variable increases in value.





t-Tests and Correlations

Below are the results from the t-test handout presented earlier in class, followed by a correlation analysis. The correlation gives the association between the independent (school type) and dependent variables (satisfaction). Notice that the p-values are identical in the two analyses.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
satisfaction rating of school	Equal variances assumed	3.200	.111	-3.000	8	.017	-3.00000	1.00000	-5.30600	-.69400
	Equal variances not assumed			-3.000	5.882	.025	-3.00000	1.00000	-5.45882	-.54118

Correlations

		satisfaction rating of school	type school type
satisfaction rating of school	Pearson Correlation	1	.728(*)
	Sig. (2-tailed)		.017
	N	10	10
type school type	Pearson Correlation	.728(*)	1
	Sig. (2-tailed)	.017	
	N	10	10

* Correlation is significant at the 0.05 level (2-tailed).

Notice that the p-value for both tests are identical. The significant difference between the means for charter and public schools is the same as testing whether the school type is associated with satisfaction. A special short cut formula called the point-biserial correlation used for the correlation between a binary and continuous variable is equivalent to the Pearson correlation coefficient.

Chi-square, Phi, and Pearson Correlation

Below are the chi-square results from the 2 X 2 contingency chi-square handout. With SPSS Crosstabs procedure, you can request Phi (for 2 x 2) or Cramer's V (for larger than 2 x 2) as a measure of association. Phi is identical to Pearson's correlation.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	81.940[b]	1	.000		
Continuity Correction[a]	80.322	1	.000		
Likelihood Ratio	84.606	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	81.776	1	.000		
N of Valid Cases	500				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 110. 40.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.405	.000
	Cramer's V	.405	.000
N of Valid Cases		500	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Correlations

		candidat candidate	party voters party
candidat candidate	Pearson Correlation	1	.405**
	Sig. (2-tailed)		.000
	N	500	500
party voters party	Pearson Correlation	.405**	1
	Sig. (2-tailed)	.000	
	N	500	500

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

The significance tests for chi-square and correlation will not be exactly the same, but will very often give the same statistical conclusion. Chi-square tests are based on the normal distribution (remember that $z^2 = \chi^2$), but the significance test for correlation uses the t-distribution. With large sample sizes (e.g., N = 120) will the t and the normal be the same (or, at least, very close).