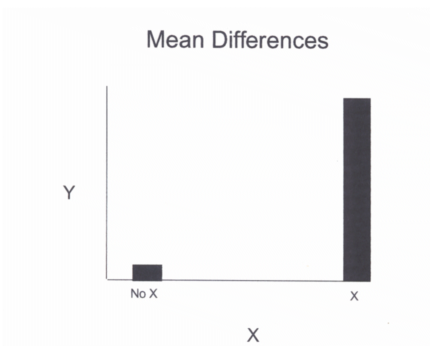
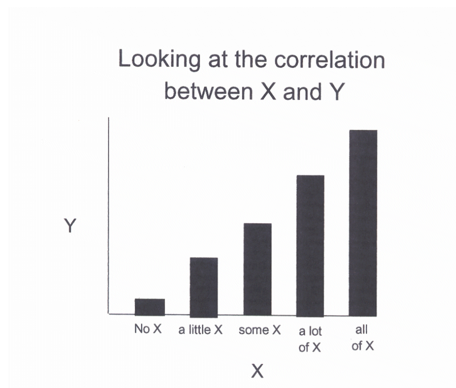
t-Tests, Chi-squares, Phi, Correlations: It's all the same stuff

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).

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*-values for the two 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 shortcut 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 × 2 contingency chi-square handout. With SPSS Crosstabs procedure, you can request *Phi* (for 2 × 2) or Cramer's V (for larger than 2 × 2) as a measure of association. Phi is identical to Pearson's correlation.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.122[a]	1	.290		
Continuity Correction[b]	.977	1	.323		
Likelihood Ratio	1.123	1	.289		
Fisher's Exact Test				.322	.161
Linear-by-Linear Association	1.121	1	.290		
N of Valid Cases	982				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 132.49.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.034	.290
	Cramer's V	.034	.290
N of Valid Cases		982	

Correlations

		ind vs affil	intended vote
ind vs affil	Pearson Correlation	1	.034
	Sig. (2-tailed)		.290
	N	982	982
intended vote	Pearson Correlation	.034	1
	Sig. (2-tailed)	.290	
	N	982	988

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., $N \geq 120$) the t and the normal z-distributions will be the same (or, at least, extremely close).[1]

Correlation and Spearman's Rank-Order Correlation (Rho)

A third type of correlation equivalent is one that is named for a correlation between two sets of ranks. For example, if I had 20 products lined up and I asked you to rank them from 1 to 20. I could then examine how these ranks were related to another set of ranks, say for a second type of product. The association between the two sets of rank scores is described by the Spearman's rho. Notice the correlation coefficient and the p-value are exactly the same for the two analyses (Pearson's correlation on the left and Spearman's rho on the right).

Correlations

		product1	product2
product1	Pearson Correlation	1	.241
	Sig. (2-tailed)		.307
	N	20	20
product2	Pearson Correlation	.241	1
	Sig. (2-tailed)	.307	
	N	20	20

Correlations

			product1	product2
Spearman's rho	product1	Correlation Coefficient	1.000	.241
		Sig. (2-tailed)	.	.307
		N	20	20
	product2	Correlation Coefficient	.241	1.000
		Sig. (2-tailed)	.307	.
		N	20	20

[1] This "approximate significance" is reported by most researchers regardless of the sample size. Perhaps this is because, for smaller sample sizes, there are not any simple recommendations. There does not seem to be much consensus in favor of the Yates continuity correction, and the Fisher Exact test, which is more powerful, depends on assumption that the marginal frequencies are fixed (determined values as in an experiment).