

Friday, November 21, 2014

Relation between two variables:
association, correlation, and regression

Recapitulating

Uncertainty about reality

Probability as a numeric measurement of uncertainty about
an event

Interval of confidence of a measurement based on a sample
containing the true value with a given probability
(random sampling in R of normal distribution, t-distribution, real mean
bounded between confidence interval using mean and SE of sample, that
can be defined at any level one chooses i.e. at any level of uncertainty)

Propagate the uncertainty

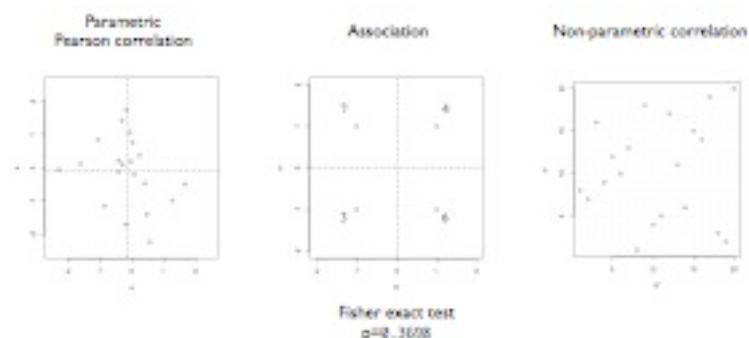
Statistical testing by defining a statistics (e.g. t) with known or
derivable distribution such that we can bound it at some
significance level α

Hypothesis Testing

Inference procedure

1. State the null hypothesis (H_0) and its alternative (H_1). Decide what data to collect and under what conditions.
2. Choose a test, the model of which most closely approximates the conditions of the research in terms of the assumption on which the test is based.
3. Find the sampling distribution of the statistical test under the assumption that H_0 is true.
4. Specify a significance level (α) and a sample size (n).
5. On the basis of 2, 3, and 4 above, define the region of rejection for the statistical test.
6. Collect the data. Using the data compute the value of the test statistic. If the value is in the region of rejection, the decision is to reject H_0 ; otherwise, the decision is that H_0 cannot be rejected at the chosen level of significance.

Relation between (random) variables



	Below	Above
Below	a	b
Above	c	d

	Below	Above
Below	$(1-p)(1-q)$	$p(1-q)$
Above	$(1-p)q$	pq

Chi-squared

$$\chi^2 = \sum_{\text{column} \in \{\text{columns}\}} \sum_{\text{row} \in \{\text{rows}\}} \frac{[\text{Observed}(\text{column}, \text{row}) - \text{Expected}(\text{column}, \text{row})]^2}{\text{Expected}(\text{column}, \text{row})}$$

$$df = (\text{columns} - 1)(\text{rows} - 1)$$

Odds ratio

	Below	Above
Below	$(1-p)(1-q)$	$p(1-q)$
Above	$(1-p)q$	pq

Odds ratio

	Below	Above
Below	a	b
Above	c	d

Covariance

$$\begin{aligned} Cov_{XY} &= E\left((X - E(X))(Y - E(Y))\right) \\ &= \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \end{aligned}$$

$$\begin{aligned} Cov_{XY} &= E(XY) - E(X)E(Y) \\ &= \frac{1}{N} \sum_{i=1}^N (X_i Y_i)(Y_i - \bar{Y}) - \bar{X} \bar{Y} \end{aligned}$$

Correlation

Pearson-correlation coefficient

$$r_{XY} = \frac{Cov_{XY}}{s_X s_Y}$$

Correlation

Spearman-correlation coefficient

$$r_s = 1 - \frac{6 \sum_{i=1}^N d(X_i, Y_i)^2}{N^3 - N}$$