StatiBot:

Tutorial

Aim of this tutorial

This tutorial explains how StatiBot works using a typical example.


Data and problem

It is well known that hedgehogs eat snails. You conduct a study to determine the occurrence of hedgehogs and snails in ten districts of your town. Since the districts differ in surface area you convert the occurrences: hedgehogs per square kilometer and snails per square meter:

District	Hedgehogs per square kilometer	Snails per square meter
Down Town	15.0	4.9
Upper Gate	16.5	2.9
Lower Gate	38.5	5.2
West Ring	37.0	6.1
South Ring	32.0	8.3
East Ring	26.5	7.0
Sunny Hills	54.0	7.2
Heaven's Valley	57.5	15.3
Duck's Wood	71.0	13.8
Abbey Forest	83.0	12.0

There are basically two possible situations:

Situation 1: High snail densities occur together with high hedgehog densities, and low snail densities with low hedgehog densities. Possible reason: a rich snail supply attracts hedgehogs.
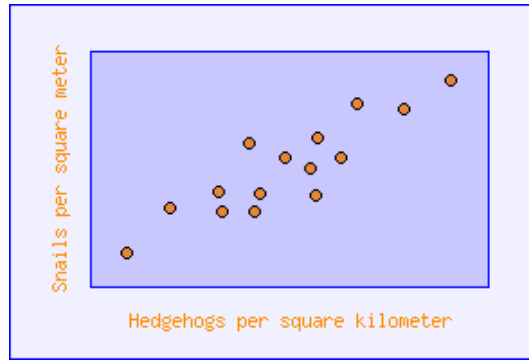
Situation 2: High snail densities occur together with low hedgehog densities and vice versa. Possible reason: hedgehogs decimate the original snail population:
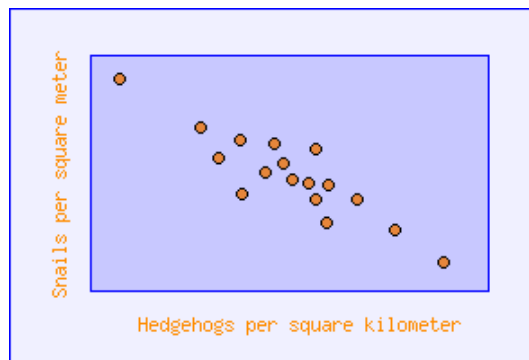
Hypothesis

The table seems to support situation 1: the more hedgehogs there are, the more snails there are. However, the link is not very close. For example, the highest snail density (15.3) is observed at Heaven's Valley, and not at Abbey Forest, where there is the highest number of hedgehogs. The same is true for low densities (15.0 and 2.9). So the link hypothesized may be rather accidental. On the other hand, the link may indeed be a true causal link. It is your aim to discriminate between the two.

Data entry

StatiBot is launched. The values are typed into the table provided by StatiBot, as in the table above. The program identifies a table with 11 times 3 cells. It prompts you to confirm that the first row contains column headings. At the end of this tutorial, other ways of entering the same information in StatiBot will be shown.

Determination of the problem

After data entry, the range of possible methods of analysis is reduced step by step. Each question is illustrated with examples to help you find the correct answer:
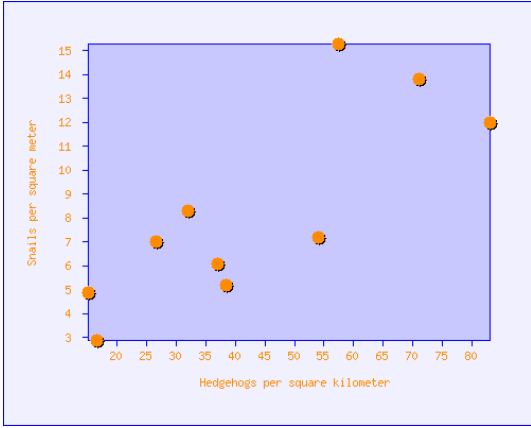
| Question	Answer
Are the units of the two columns "Hedgehogs per square kilometer" and "Snails per square meter" identical?	No.

| Question	Answer
The values of one of the two columns were obtained during the investigation. How did you obtain the values of the other column?	Also during the investigation.
Are you sure that the values of one column are responsible for the values in the other column, at least to some extent? In other words: are you sure that one quantity represents the cause and the other one the effect?	No, the causality is open.
Are you interested in whether the values of the two quantities change together?	Yes.
How did you obtain the values?	By measuring or by counting.
Are the values distributed normally?	Unknown.
StatiBot has executed a D'Agostini test. According to this, the data in the "Hedgehogs per square kilometer" column are distributed normally. Do you want to confirm this decision?	Yes.
StatiBot has executed a D'Agostini test. According to this, the data in the "Snails per square meter" column are distributed normally, too. Do you want to confirm this decision?	Yes.

Results

The unique feature of StatiBot is its ability to help you find the right test. In addition, a number of references are provided (similar tests, synonyms, literature) that allow you to obtain further information on the analysis. And finally, StatiBot furnishes all the key values of the test that any other statistical program might provide.

Here are the actual results of the hedgehog-snail example:

Name of the appropriate test	Pearson correlation analysis
Main result of the test, graph and words	The larger the values in the "Hedgehogs per square kilometer" column, the larger the values in column "Snails per square meter". Almost certainly, this result represents a real connection. The probability that the result is accidental is extremely low.
Conditions for the test	StatiBot has executed a D'Agostini test. According to this, the data in column "Hedgehogs per square kilometer" are distributed normally. If you want to cite the test give the following information: $n = 10$, $T = 608$, $D = 0.2819$, $p > 0.2$.
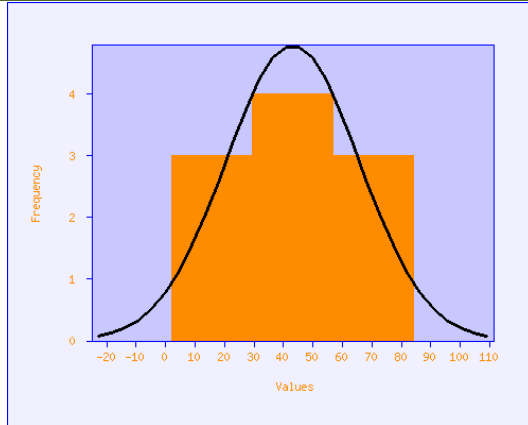
StatiBot has executed a D'Agostini test. According to this, the data in column "Snails per square meter" are distributed normally as well. If you want to cite the test give the following information: n = 10, T = 107, D = 0.2763, p > 0.2.
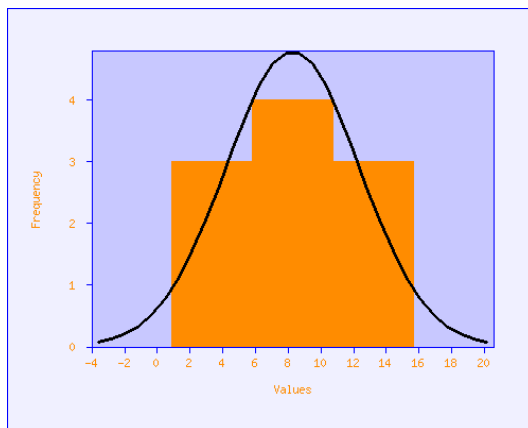
Information on the P value	The P value is 0.00455. A low P value means that the result is meaningful (significant) from a statistical point of view. A high P value means that the differences found are accidental. In many situations it is practical to use a 0.05 threshold as a decision criterion (i.e. an error probability of 5 percent). However, this threshold is actually arbitrary.
Scientific quotation	Give the following information to cite the test: n = 10, r = 0.809, t = 3.90, P(two-tailed) = 0.00455.
Explanation of the key values	The r value is called the correlation coefficient. Its range is from -1 to +1. Correlation coefficients above zero mean that the values of the two columns move in the same direction. Correlation coefficients below zero mean that small values of one column are associated with large values of the other column (and vice versa).

	Correlation coefficients close to zero point to a weak link. Those close to -1 or +1 point to a strong link.
Interpretation aids	You cannot base a causal direction on the correlation found. There are three possibilities: 1. "Hedgehogs per square kilometer" influences "Snails per square meter". 2. "Snails per square meter" influences "Hedgehogs per square kilometer". 3. A third unknown factor influences both "Hedgehogs per square kilometer" and "Snails per square meter". A study in which either "Hedgehogs per square kilometer" or "Snails per square meter" are determined experimentally defines the direction of causality (if such a study is possible at all).
Synonyms	Synonyms of this analysis are: "simple linear correlation analysis", "product moment correlation" and "parametric correlation analysis".
References	J. Zar (1984) *Biostatistical Analysis*, Prentice Hall, Englewood Cliffs, p. 306ff. R. Sokal & F. Rohlf (1987). *Introduction to Biostatistics*, W. H Freeman and Company, New York, p. 267ff.

Open data entry structures

StatiBot features open data entry structures. Data can simply be copied from spread-sheets or text editors by common copy and paste procedures.

Of course, StatiBot is open to different types of input tables for one single study design. So, in the hedgehog-snail example for instance, the district names could be dropped:

Hedgehogs per square kilometer	Snails per square meter
15.0	4.9
16.5	2.9
38.5	5.2
37.0	6.1
32.0	8.3
26.5	7.0
54.0	7.2
57.5	15.3
71.0	13.8
83.0	12.0

Or the data can be input vertically without headings, and with commas rather than decimal points:

15	16,5	38,5	37	32	26,5	54	57,5	71	83
4,9	2,9	5,2	6,1	8,3	7	7,2	15,3	13,8	12