Deciding between chi square and t test

I am always so confused whether to do a chi square test or a t test in the sums given by my biostats teacher. Does anyone have a simple rule to decide this?

homework biostatistics

asked Nov 14 '13 at 17:14

biogirl
3,791 4 19 58

3 Answers

This is a very subtle question and I encourage you to read the Wikipedia articles on these different subjects (t-test, chi-squared test, p-value, etc) because the authors worked hard to combat common misconceptions about these commonly used statistical tests. Here is a rather oversimplified rule-of-thumb for these different tests:

1. t-test: Used when you are looking at the *means* of different populations. For example, you might want to determine whether the difference in the mean gene expression level between treated and untreated cells is different, or if the gene expression level of cells in a certain environment differs from what you would expect in a null hypothesis. **Assumptions:** You are assuming that the populations you are looking at are normally distributed. The variance of the populations is not known (that would be a Z-test), but it is assumed that the variance of each population is the same. Finally, for the t-test to work, the samples of the data from the two populations are assumed to be independent.
2. χ^2 test: Several possibilities for this. The most common in biology is the Pearson χ^2 test, which is used when you are looking at *categorical data*, like the number of pea plants with white or purple flowers and round or wrinkled seeds, and trying to see whether the number of individuals in each category is consistent with some null hypothesis (such as the number in each category you'd expect if the genes for flower color and seed shape are not linked). **Assumptions:** The data points were randomly and independently gathered from the population, and you have a reasonably large number of samples.

I'd hate to have made a huge mistake, so please edit my answer and/or contribute your own if you think I am completely misrepresenting these topics!

answered Nov 14 '13 at 18:32

A. Kennard
1,409 5 18

8 This is correct, though I think the core of it is actually simple. Use chi-square if your predictor and your outcome are both categorical variables (e.g., purple vs. white). Use a t-test if your predictor is categorical and your outcome is continuous (e.g., height, weight, etc). Use correlation or regression if both the predictor and the outcome are continuous. – octern Nov 14 '13 at 22:57

Additional Info

T-test

As A.Kennard said t-test is applied when the random variable is normally distributed. How to know what is normally distributed is a relevant question. Regular measures which suffer some random error of measurement are normally distributed. The mean values estimated from different samples (the experiment that generates that sample may have **any** distribution) follow normal distribution. For e.g mean time interval of a radioactive decay- the interval itself is exponentially distributed but the mean of the mean decay interval will be normally distributed. You can reason that it is again an error of measurement that leads to variation in the mean value calculated in different samples. This is called the *central limit theorem*.

A normal distribution has two parameters- mean and variance i.e. you need to know these values beforehand to construct a normal distribution. A uniform distribution has no parameters- that doesn't mean that uniformly distributed samples have no mean or variance (in this case mean and variance are sample properties not distribution parameters). A t-test or z-test is done to see if

a sample is a representative of a given normal distribution. That again means that the calculated mean and variance are equivalent to the corresponding distribution parameters. In case of z-test you know the population variance (distribution parameter). You may ask how can anyone possibly know the population variance beforehand. An example is a case in which you already know the error rate of your measuring device (may be provided by the manufacturer or interpreted from its design).

χ^2 test

There are several variants of the χ^2 test. But what is common between them is that they refer to χ^2 **distribution**. Variances, which are always positive cannot be a normally distributed. These follow χ^2 distribution. The F-test for variances uses the ratio of the χ^2 statistic of the two random variables denoting variances. Even in the Pearson χ^2 test, the test statistic is a sum of squares which makes it always positive. In fact this χ^2 distribution is also used in t-test. As . Kennard said, one of the assumptions of t-test is that the population variance is unknown but assumed to be equal. Since population variance is unknown it has to be estimated from the sample. Like the case with all estimates you dont have a fixed value but a range of acceptable values falling in some confidence intervals. T-distribution is basically an average of several normal distributions with variance values falling in the allowed confidence interval of a χ^2 distribution.

It is not necessary that categorical data are to be tested by χ^2 test. Coin toss experiment gives rise to a categorical but it can be tested against a binomial distribution. So χ^2 test can be used for categorical data but it is not the only test.

Bottomline : a statistic tested by a χ^2 test has χ^2 **distribution** as its sampling distribution. That statistic should be a square/sum of squares- something that can never possibly have a negative value. Perhaps that is why it is called χ ***squared***.

edited Nov 16 '13 at 15:52

answered Nov 15 '13 at 6:42

WYSIWYG ♦
20.7k 3 28 97

Its true that T-test are used when your dependent variable is Numeric and Chi-Square test is used when you are analysis categrical variable. But how about this:

You have a categorical response (0,1) to a campaign. 1 who bought the product and 0 who did not. If you sum up the responses in your Test group and Control group and devide them by thier respective population size, lets say you get something like this - .23% response rate in Test Group and .01% response rate in Control Group.

Can't you use T-Test to see if these reponse rates are different? If yes, then let me remind that these variables were categorical (0,1) but we still used them as numeric.

All I want to say is that if we are comparing response rates or percentages, then T-Tests can be used irrespective if the dependent variable is character or numeric.

Sachin

answered Aug 19 '14 at 9:04

sachin01663
7 1

No, you can't compare just two numbers (.23 vs. .01) with a t-test. – kmm Aug 20 '14 at 10:58

@kmm He could've collected a series of results to test. Nevertheless, this would be a great mistake, not only due to excessive information (conduct a series of assays to run a t-test, whereas 1 assay was sufficient to run chisq), but due to compositional structure of the data that limits both variance and mean (hence you get an asymptotic - nonlinear expectation), hence breaking the assumption of linearity of expectation. – Eli Korvigo Dec 17 '15 at 12:54