The Philosophy of Hypothesis Testing, Questions and Answers

Question: So I'm hypothesis testing. What's the hypothesis I'm testing?

Answer: When you're testing a hypothesis on a regression coefficient, the hypothesis you're testing is that the true coefficient is equal to some value that you specify. For example, you often test the hypothesis that the true value of a coefficient is zero.

Question: Wait a minute. I just did a regression and found that the slope coefficient, the coefficient of my X variable, is 0.345678. Isn't that the true value?

Answer: Probably not, according to the classical philosophy of statistics. A classical statistician would say that the data you used in your regression were just one of many possible data sets that the true relationship might have produced. Even if the true value of the coefficient were 0.345678, you'd have to be extraordinarily lucky to find a data set that would give you that number when you ran your regression.

Question: I'd like to test the hypothesis that the true value of X's coefficient is zero. I find that the number in my regression output for the t-statistic is bigger than the number in the t table in the 0.05 column at the number of degrees of freedom my model has. Does this mean that I'm 95% sure that the true coefficient is not zero?

Some books' t tables label the column you use to do a two-tailed test at the 95% confidence level "0.025." Such a table is for a one-tailed t test. The critical value for a one-tailed test at the 97.5% level is the same as the critical value of a two-tailed test at the 95% level.

Answer: It depends on what you mean by "95% sure."

Question: Does it mean that the probability is 95% that the true coefficient is not zero?

Answer: Here's that fine point of statistical philosophy again.
There's nothing probabilistic or random about the true value. For example, suppose that we want to know the average age of students in the Public Health School. That's a specific non-random number. We don't know what it is, though. If we estimate based on a random sample of students, then it's our estimated value that's random, because we get a different number depending on who happens to be in our sample.

Question: That's OK for a random sample from a big population, but what if we have the whole population in our data? Suppose my dependent (Y) variable is the number of infant deaths in each county in South Carolina. Suppose my independent (X) variables are social and economic factors. Your theory has me treat my Y variable as if it were random. But, the number of infant deaths last year in Charleston County, for instance, is not a random number. DHEC counts it carefully, and it's based on the whole population of that county, not a sample.

Answer: You have a point there. It's strange to think of reality as random while maintaining that something that exists only in theory, the true value, is absolute. But, that's the classical statistical philosophy. Plato would like it.

Question: OK, classical statistical philosopher, what do I mean when I say that I'm 95% confident that some coefficient is not zero?

Answer: You mean this: If the true value of the coefficient were zero, the probability is at least 95% that a regression would have given an estimated value for that coefficient that was closer to zero than the estimated value that you got.

Question: That's quite a mouthful.

Answer: Yes, and it's stated that way because it's the estimated value that's considered random, not the true value.

Question: Still, it takes some work to follow the logic.

Answer: Suppose you could run lots of regressions on different data sets, each of which was independently generated by the same true relationship in which the true coefficient of X is 0. (This means that truly there is no relationship.) The more regressions you run, the more likely it is that 95% of the t-statistics you get will be closer to 0 than the critical value in the t table.

Question: By the way, what's so special about 95%? Why does everybody use it?

Answer: It's just the conventional confidence level. It's reasonably conservative for many purposes.

There are two types of errors of inference that you can make testing hypotheses. A Type I error (that's what it's actually called) is rejecting a hypothesis that is really true. A typical Type I error is declaring that a coefficient is not zero, but really it is. The 95% confidence level means that the probability of a Type I error is only 5%.

Question: Why not choose a 99% confidence level, or a 99.9999% confidence level, to make the probability of a Type I error very low?

Answer: The lower the probability of a Type I error, the higher the probability of a Type II error. A Type II error is refusing to reject a hypothesis that is really false. This happens if the true coefficient is not zero, but you are unwilling to say it's not zero because your estimated coefficient was not far enough from zero to satisfy you. There is an unavoidable tradeoff here. The 95% confidence level is the usual compromise. Sometimes you'll use the 99% confidence level, if you're a lot more worried about a Type I error than a Type II error.

This concern is not just academic. Consider the testing of the safety and efficacy of a new drug. Many lives and many dollars can ride on the selection of an appropriate confidence level.

This table summarizes what Type I and Type II errors are.

		We say	
		We do not reject the hypothesis.	We do reject the hypothesis.
Hypothesis is really	True	We are correct.	Type I error
	False	Type II error	We are correct.

The hypothesis can be true or false. We can refuse to reject the hypothesis, or we can reject the hypothesis. The table above shows the four combinations of those possibilities. If what we say matches the truth, then we are correct. Otherwise, we have made either a Type I or a Type II error.

Question: Suppose I want to test the hypothesis that a coefficient equals some number other than zero?

Answer: Subtract the hypothesized value from the estimated coefficient you got on your regression output. Divide that difference by the estimated standard error of that coefficient. Compare the quotient (ignoring the minus sign if you have one) with the critical value in the t table. If the quotient is bigger, reject the hypothesis.

Question: What's a "confidence interval"?

Answer: It's a short cut. It lets you avoid having to repeat the above calculation over and over. Once you compute the confidence interval, you know that you'd reject any hypothesized value that's outside the confidence interval, but not reject any hypothesized value inside the interval. Incidentally, you'll find that the 99% confidence interval is wider than the 95% confidence interval. The bigger confidence interval means a smaller chance of a Type I error, but a bigger chance of a Type II error.

Question: Is the probability 95% that the true value lies inside the 95% confidence interval?

Answer: Again, the answer is no, not according to classical statistics. The 95% confidence interval you calculate is considered random, because it depends on your data, which are presumed to be one of many possible realities. If, across lots of studies of different things, you make a practice of rejecting hypothesized values because they are outside the 95% confidence intervals, then the tendency is that you'll be right 95% of the time.

Question: Classical statistical philosophy is awfully convoluted.

Answer: Yes. There is an alternative approach, called Bayesian. This does allow you to apply the concept of probability to true values. However, the classical approach is by far the predominant one, so you should try to understand it so you can converse with other statisticians.

Question: I can grind through the mechanics of hypothesis testing, but whole thing seems mysterious. What am I really doing?

Answer: If you accept the idea that your data are the result of a random process, then any number ("statistic") that you calculate from your data is also the result of the random process. If you make certain assumptions ("hypotheses") about the random process that generated your data, then you can calculate the distribution of your statistic. You can calculate the probability that your statistic will take on any particular value or be inside any particular range of values. If your statistic is far from its expected value, such that being that far from the expected value is very unlikely, then you reject your hypothesis about how the data were generated.

Here's how that works with a coin toss experiment. Flip the coin six times. You can then calculate a statistic, the total number of times it comes up heads in six flips. (This is not a complex calculation. You just count the number of heads and that's your statistic.) Now, hypothesize that the probability that the coin will come up heads is 0.5, and that the flips are independent. Based on your hypothesis -- those assumptions -- you can calculate the distribution of your statistic. It is:

Possible values of the statistic, the number of heads in six coin tosses	Probability that the statistic is one of those values, assuming that the hypothesis that the coin is fair is true	Probability expressed as decimal
0	1/64	0.015625
1	6/64	0.09375
2	15/64	0.234375
3	20/64	0.3125
4	15/64	0.234375
5	6/64	0.09375
6	1/64	0.015625
All	1	1

This distribution has a "fat middle," in the sense that the most likely values are towards the middle (2, 3, and 4). The least likely values are towards the "tails" (0 and 6).

If our hypothesis is that the coin is fair, we want to reject that if our statistic is too far out towards either tail. That is, we want to be able to reject the hypothesis if we get either too many heads or too few. To do a "two-tailed" test, we combine the tails, like this:

Table for a two-tailed test that the coin is fair

Possible values of the statistic, the number of heads in six coin tosses	Probability that the statistic is one of those values, assuming that the hypothesis that the coin is fair is true	Probability that the statistic (the number of heads) is this far or further from its expected value (3).
0 or 6	2/64 = 0.03125	0.03125
1 or 5	12/64 = 0.1875	0.21875 = 0.03125+0.1875
2 or 4	30/64 = 0.46875	0.6875 = 0.03125+0.1875+ 0.46875
3 – the expected value	20/64 = 0.3125	0.3125
All	1	1

Compare the value of your statistic (number of heads in your actual six tosses) with the table. If your statistic is 0 or 6, you have an outcome which is quite unlikely to happen -- less than a 0.05 probability -- if the hypothesis that the coin is fair is true. You therefore reject the hypothesis and say the coin is not fair.

If your statistic is 1 or 5, you cannot reject the hypothesis, because your outcome is not unlikely enough. The probability of being that far from the expected value, 3 heads, is the probability of 1 or 5 plus the probability of 0 or 6, which is 0.1875+0.03125, for a total of 0.21875. This is greater than 0.05, your cutoff for rejecting an hypothesis. If you say that the coin is not fair, there is a 0.21875 probability that you are mistaken. That is too high a risk of a mistake for most researchers.

The t-test calculation is more complicated, but the idea is the same. You hypothesize that the true coefficient is 0 (or whatever number you want) and you assume that the errors have the normal distribution. The hypothesis, plus the very strong assumption about the errors being normal, imply that the statistic you calculate from the data and call t has a known distribution that you can look up in a table. If, according to the table, the t value you got is very unlikely, then you can conclude that hypothesis is wrong. You reject the hypothesis.

Question: We made two statements: (1) that the true coefficient is 0 (or whatever I wanted to test) and (2) that the errors have the normal distribution. When we reject a hypothesis, we always reject (1) and not (2). Why is that?

If you aren't sure about (2) then you shouldn't use the t statistic for your hypothesis test. There are other tests that can help you determine whether (2) is appropriate.

More on T Tests

You calculate a special number, called the t-statistic. You calculate it based on the assumption that a certain hypothesis is true. The hypothesis is usually that the true value of a certain one of your regression's coefficients is zero.

What does it mean if the coefficient of a variable in a regression is zero? It means that you can leave that variable out of your equation without losing any predictive power.

Consider a simple regression, of the form $Y = \alpha + \beta X + \text{error}$. If β is zero, the βX term disappears, and the reduces to $Y = \alpha + \text{error}$. Y does not depend on X. Knowing what X is does not help predict what Y is.

To calculate the t-statistic for testing whether a particular β is zero, we divide our estimate of β by its standard error. This measures how steep the regression line is, relative to how far the observed Y values are from the regression line and how spread out the X values are.

The t-statistic has no units of measurement. This means that the t-statistic comes out the same regardless of how X and Y are measured. For example, if X and Y are heights of people, as in Francis Galton's original study (to be mentioned in class), we get the same number for the t-statistic if we measure heights in inches or centimeters.

Better than that, the t-statistic has a known distribution *if the hypothesis that you are testing is true.* The printed t-table shows what that distribution is.

In any data set with two variables, X and Y, the data for the two variables will almost always correlate at least a little bit, even if there is no true underlying relationship. Suppose you have two rooms. In each room, you dump 1000 coins on the floor. Most likely, one of the rooms will have more "heads" than the other. A t-test can tell you if the difference is large enough to mean that there's really a difference between the rooms, or if the difference is just by luck.

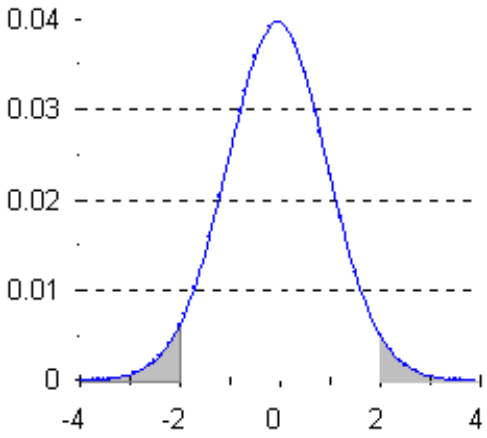
When we compare our calculated t-statistic with the critical value in the t-table, we are asking if the slope estimate we got might have happened just by luck. This could happen if the random errors correlate with X by luck. (Correlate means that there is a linear relationship.) When the errors correlate with X, then the Y values correlate with X, too. We can get fooled into thinking X and Y are related.

If there is really no true relationship between X and Y, then random errors will probably work out so that the t-statistic will be close to zero. T-statistics far from zero are unlikely, unless the assumption about there being no true relationship between X and Y is wrong.

If the data are such that higher X values are associated with higher Y values, then, when you do a regression with the equation $Y = \alpha + \beta X + \text{error}$, the t-statistic for the estimate of β will be positive. If lower X values are associated with higher Y values, then the t-statistic will be. The more that the X and Y data are correlated, and the steeper the line is that they seem to lie along, the further the t-statistic will be from zero, either positively or negatively.

The table of t-statistic critical values has lots of rows. Theoretically, it could have an infinite number of rows, but my table shows only what will fit on one page. The multitude of rows in the table means that

there is a family of t distributions, one t distribution for each number of degrees of freedom. (The degrees of freedom is the number of observations minus the number of parameters in the regression equation). Each of these t distributions has a bell shape. When the number of degrees of freedom is small, the bell is wide, so the critical value numbers near the top of the t table are higher. When the number of degrees of freedom is large, the bell is narrower, and the critical value numbers in the t table are smaller. If the number of degrees of freedom is very large, the t distribution becomes indistinguishable from the normal distribution.



The t distribution with 50 degrees of freedom. Shaded are the two "tails" that together have 5% of the total area under the curve. Reject the null hypothesis if the t-statistic is in either tail.

Here is a t-distribution bell curve for 50 degrees of freedom. Strictly speaking, this curve is the "density" function for the t-distribution.

T values are on the horizontal axis (here labeled -4 -2 0 2 4). The height of the bell curve over any t value is proportional to how likely it is that you would get something close to that t value purely by luck, with no true relationship between X and Y. The curve is highest above where the horizontal axis is 0, indicating that t-statistics near 0 are the most likely to happen by chance.

As you move away from zero, in either the + or the - direction, the curve gets lower down. If you move far enough, you get to where a t value that far from zero would happen by luck only 5 times out of 100. Those places are the critical values of the t distribution at the 5% level. They are shown on this graph as the inside edges of the shaded areas. Because the t-statistic's bell curve is symmetrical left and right, The two places have the same number, except that one is positive and one is negative. This is the number that you find in a statistics book's t table, or in our t table in the downloadable file of tables.

The portions of the t distribution that are to the right of the critical value in the positive direction, and to the left of the negative of the critical value in the negative direction, are called the "tails" of the distribution. These are shaded in the diagram above. The t distribution has two tails, one positive and one negative.

We use two-tailed t tests in this course. We reject the null hypothesis that β is 0 if our calculated t statistic is in either the positive tail or the negative tail. This is appropriate whenever you are not sure in advance whether X is positively or negatively related to Y, so you want to allow for both possibilities.

To put it another way, in a two-tailed test we reject the idea that the true $\beta = 0$ if our t statistic is greater than the critical t value *or* if it is less than the negative of the critical t value. That is the same as saying that we reject the hypothesis that β is 0 if the absolute value of our t-statistic is greater than the critical value.

The t-test can be used to test for other possible coefficient values besides 0. The general formula is

t-statistic = $(\hat{\beta} - \beta_{test})/(\text{standard error of } \hat{\beta})$, where β_{test} stands for the β value we want to test.

The null hypothesis is that the true β is β_{test}. If this t-statistic is larger (in absolute value) than the critical value from the t table, reject the hypothesis that the true β is β_{test}.