A Short Review of the "Normal vs. Student" Issue

Math&146 - 211

1 Basic Idea

Let's assume, at least at first, that we are making observations of a genuine normally distributed random variable. According the the usual mathematical model then, we come up with numbers $x_1, x_2, \ldots, x_n$, that we consider observations of n independent identically distributed variables, all with a normal distribution, and let's call the mean and the variance of this distribution μ, and σ^2. Traditionally, we denote this variables by $X_1, X_2, \ldots, X_n$.

The first observation (its relevance is due to the fact that for decades people had to rely on tables to compute probabilities of the normal distribution, and, even today with our powerful computers, it may be faster to use a table for a quick and dirty estimation) is that, if we call $\bar{X}$ the random variable $\frac{1}{n}\sum_{k=1}^{n} X_k$, this has again a normal distribution, with mean μ, and variance $\frac{\sigma^2}{n}$, so that

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \tag{1}$$

has a *standard* normal distribution (mean 0 and variance 1). This means that, if we know the value of σ, we can, for example build confidence intervals for μ, or check if the observed value for the mean ($\bar{x} = \sum_{k=1}^{n} x_k$) happens to be so far from the mean μ that we assume is true, that it suggests that we were wrong in assuming that to be the mean (in other words, we have a tool to reject hypotheses on the mean).

Thus, if we are looking for, say, a 95% confidence interval for the mean, we can write that (with that level of confidence)

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \tag{2}$$

This is all fine, but, in reality, it is highly improbable that we know the value of the variance. We can create some situations where this may be the case, but they are few, and not always ironclad. If we don't know the variance, we cannot use the methods sketched above for estimating or testing for the mean. This is where a statistician working for the Guinness brewery in Dublin, Mr. William Gosset, enters the scene. He could not publish his results under his own name, because Guinness was paranoid about publication of "trade secrets", but, after developing a rigorous methods to deal with this issue, he improved the brewery's quality control enormously, and published (under the auspices of Roland Fisher) his results under the pseudonym "Student". Too bad we keep calling this "Student's" this and that (method, distribution, and so on), instead of Gosset...

"Student's" trick was, essentially, to sit down and do a lot of tedious calculations, resulting in the computation of the distribution of

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \sqrt{n}\frac{\bar{X} - \mu}{s} \tag{3}$$

i.e., something that looks very much like (1), but involves the observed *sample*[1] standard deviation, rather than the unknown square root of the "true" variance. While the distribution of (1) is the same for all n, this is not the case for (3), and, for quirky historical reasons, the distribution of (3) with a given value of n is called the "Student distribution with $n - 1$ degrees of freedom".

Thus, if we have our observations and we cannot pretend to now the value of the true variance, we will write a confidence interval very much like (2), except that, instead of "1.96" we will plug in a number picked form the table for the Student distribution with $n-1$ degrees of freedom, corresponding to the appropriate confidence level.

If you look at the graph of the density of Student's distributions, you will see that they too are symmetrical and bell-shaped, but they are squatter, wider, than the normal distribution (they result in larger confidence intervals, as is reasonable, since we have less information). The higher n the closer the corresponding Student distribution is to a standard normal, but there are small differences even for $n = 100$ or higher.

In practice, if n is large, it really makes very little difference: most observations are not so precise that a change in the third decimal from 1.96 will make any difference at all, but it all depends on your data. If you are into high precision measurements, it may be worth the trouble.

2 Cautionary Notes

Assuming an underlying normal distribution is, in most applications, not quite as solid an assumption as we may like. However, as long as we are dealing with the "known variance" case, the Central Limit Theorem assures us that if n is reasonably large, even if the individual $X_1, X_2, \ldots$ are not quite normal, $\bar{X}$ will essentially be. That makes the use of (1) fairly universal. However, as noted, its use requires us to know the true variance, and that is pretty unusual.

So, what about the more realistic case? Strictly speaking, Gosset's calculations are strongly dependent on the assumption that the $X_1, X_2, \ldots$ are indeed normal so the reliability of its method is, for very strict theoreticians, pretty

[1] Gosset could, just as well, have used the *population* standard deviation, resulting in slightly different numbers, but he did not, so we are stuck with his choice

limited. Practitioners couldn't care less, and use Student distributions all the time, regardless. While this can be, at times, a bit reckless (and is sometimes really unnecessary, since one could use different distributions, more appropriate to the experiment, and obtain better grounded results), there are theoretical results that show that results obtained this way are reasonably accurate, provided the true distribution, even if not quite normal, is at least symmetric and not too spread out. What you or I can make out of this is very subjective: after all, statistics is a bridge connecting a fairly sophisticated theory with everyday concerns and needs, and what compromise is acceptable depends on the circumstances, and, also, who you ask.

Let's spell out the main point of this issue: while the Central Limit Theorem guarantees that, for n sufficiently large (and "sufficient" is not very large, if the true distribution is not terribly asymmetric), $\bar{X}$ has, for all practical purposes, a normal distribution. However, to state that (3) has a Student distribution, we also need s^2 to have a fairly approximate χ^2 distribution. While one can argue that "for sufficiently large" n this will happen, the "sufficient" size is much higher than the one needed for the normality of $\bar{X}$. Thus, the studies mentioned above, about symmetry and not excessive dispersion ensuring a useful outcome of an analysis using Student's distribution are highly non trivial.

3 In Practice: Tests

3.1 The formulas

Suppose we are implementing a *two-tailed test* (one-tailed tests are even simpler), for the mean of a normal random variable X, say $mean = \mu_0$, and the average of our n observations was $\bar{x}$. First of all, let's set up a test, with significance level α.

3.1.1 Case 1: We Know The Variance σ^2

We use the notation z_γ (a widely used notation) to indicate a number such that $P\left[Z > z_\gamma\right] = \gamma$, where Z is a standard normal variable. Under our assumptions,

$$\sqrt{n}\frac{\bar{x} - \mu_0}{\sigma}$$

is a standard normal, hence it falls between $-z_{\alpha/2}$, and $z_{\alpha/2}$ with probability α. Consequently, we will not reject the hypothesis $mean = \mu_0$, if $\bar{x}$ is such that

$$l = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = u$$

The interval $[l, u]$ could be called the "acceptance region", and it is the range of observed mean that make the test non significant.

Let's now consider an alternate hypothesis for the mean, say $mean = \mu_1$. Even if this was true, we would still not reject μ_0, as long as the observed mean was in $[l, u]$. This "mistake" can happen with a probability that we can compute, under the assumption that the true mean is μ_1: in fact, if Y is a normal random variable with mean μ_1, and variance σ^2, then

$$P[l \leq Y \leq u] = P\left[\sqrt{n}\frac{l-\mu_1}{\sigma} \leq \sqrt{n}\frac{Y-\mu_1}{\sigma} \leq \sqrt{n}\frac{u-\mu_1}{\sigma}\right]$$

and the quantity in the middle is a standard normal variable. This probability is the probability of an "Error of Type II". The power of the test, at μ_1 is given by $1-$this probability.

Repeating for other values of the mean, we can compute how the power changes as we change the mean.

3.1.2 The Variance Is Unknown

The convenience of the Student distribution is that we need not change anything in the discussion above, safe making a few simple substitutions:

- use the sample standard deviation s, where we were using the true standard deviation σ
- use the t_{n-1} distribution, where we were using the standard normal distribution

The formulas translate directly, with $t_{n-1,\gamma}$ (the number such that $P[T > t_{n-1,\gamma}] = \gamma$, if T is distributed according to t_{n-1}), in place of z_γ. The advantage of using software instead of tables is huge now: tables for the Student distributions only list values $t_{m,\gamma}$ for a very limited choice of γ. This is not a big deal in interval estimation, or in significance testing, since people are accustomed to consider only these special values as levels. When it comes to compute powers, however, we end up looking for probabilities that may not be listed in such tables, and would have to employ a lot of guesswork. A computer has no such limitation, of course.

3.2 A Numerical Example

3.2.1 Known Variance

Suppose $n = 25, \mu_0 = 10$, $\alpha = 0.05$. Suppose, at first, that we know that $\sigma^2 = 9$. Then, our acceptance interval is an interval such that a normal random variable with mean 10 and standard deviation $\frac{3}{5}$ (that's $\frac{\sigma}{\sqrt{n}}$, the standard deviation of the average observation) falls there with probability 0.95. We can use tables, or we can use a computer, and end up with the interval ($z_{0.025} = 1.96$, approximately)

$$\left[10 - 1.96 \cdot \frac{3}{5}, 10 + 1.96 \cdot \frac{3}{5}\right]$$

or

$$[8.824, 11.176]$$

Now, assume $\mu_1 = 11$. From the previous discussion, we need to find the probability that a standard normal variable falls within

$$\left[5 \cdot \frac{8.824 - 11}{3}, 5 \cdot \frac{11.176 - 11}{3}\right]$$

$$[-3.627, 0.293] \tag{4}$$

Using tables or software, that's about 0.615. The power of our test, for $\mu = 11$, is equal to 0.385

> The calculation in a spreadsheet is very quick: `normsdist(0.293) - normsdist(-3.627)`

Consider now as an alternate hypothesis that the mean is 15. In this case, we repeat the calculation for the interval

$$\left[5 \cdot \frac{8.824 - 15}{3}, 5 \cdot \frac{11.176 - 15}{3}\right]$$

$$[-10.30, -6.37]$$

which has a really small probability: about 9.25×10^{-11}, for a power of practically 1.

3.2.2 Unknown Variance

To check how the choice of a Student distribution inevitably creates larger acceptance regions (and hence, makes it harder to have significant results), let us now assume that $s^2 = 9$, that is let's work with a sample standard deviation equal to the assumed "true" variance in 3.2.1.

The calculations are similar, with the only change being the use of the distribution t_{24}, instead of the standard normal. The acceptance interval is now ($t_{24,0.025} \approx 2.064$)

$$\left[10 - 2.064 \cdot \frac{3}{5}, 10 + 2.064 \cdot \frac{3}{5}\right]$$

$$[8.7616, 11.2384]$$

(obviously larger than (4)).

The probability of ending up there if the true mean is 11, is given by the probability that a variable, with t_{24} distribution falls in

$$\left[5 \cdot \frac{8.7616 - 11}{3}, 5 \cdot \frac{11.2384 - 11}{3}\right]$$

$$[-3.7307, 0.3973]$$

From software, this probability turns out to be 0.6521, and the power is, approximately, 0.348. Note that if we are restricted to tables, we would have been in quite a bind, since 0.3973 is so far from the tails that almost no table will be of help.

The alternate hypothesis of a mean of 15 similarly leads to a power of practically 1: the interval is now

$$\left[5 \cdot \frac{8.7616 - 15}{3}, 5 \cdot \frac{11.2384 - 15}{3}\right]$$

$$[-10.397, -6.269]$$

and the probability of ending there is again very small: about 8.82×10^{-7} (still quite larger than what we got for a normal distribution).

3.2.3 Reading The Results

Of course, we would repeat the calculations for many more values of the mean. By the symmetry of the distributions involved, our result for 11 is the same as what we would get for 9, the result for 12, would be the same as the result for 8, and so on. Now, that we have these numbers we would like to do something with them.

First, note how, as expected, we have a more powerful test if we can refer to the normal distribution. This is very reasonable, since in the second case, we have less information (we don't know the true variance, and so we have higher incertitude.

Second, what does, for example, the result that the power of the test is 0.348 for a mean of 11 mean? That means that, if the true mean is not 10, but 11, our test will recognize that 10 is not the correct value with probability of only 35%. Clearly, most of the time we will not reject the hypothesis that the mean is 10. In other words, if the mean is 11, we won't be really able to tell the difference with any likelihood.

We will be much more confident about our ability to spot the mistake in assuming that the mean is 10 for alternate values such that the power of the test is much higher, for example 90% or better. There is no question that the likelihood of not rejecting the hypothesis that the mean is 10 would be negligible if the true value was 15.

This kind of information allows us to be much more precise about what we mean when we say that a test is not significant. We could not reject the Null Hypothesis, and while that does not imply that the Null Hypothesis is actually true, it gives us some confidence about the fact that the true value, while not necessarily 10, is not terribly different either.

The final observation concerns the difference in outcomes between calculations using the normal and the Student distributions. The difference becomes smaller as the sample size increases (it can be shown that as n grows the Student distribution becomes indiscernible form the standard normal distribution). If you look up a table for the t-distribution, you will notice this tendency. Since it is mildly easier to deal with the normal tools, we may wonder when will it be OK to use them, even if the variance is unknown. As expected, the answer has to be "it depends". As you can check, there are still differences when $n = 1000$, but they are very small. Whether this difference is worth worrying about depends on what precision your data actually has. If you are working at the High Energy Hadron Collider in Geneva, you are probably working with extremely high precision data, and small differences can have a big impact. If you are dealing with rough data, where two significant digits, at most, are reliable (that is, for example, 2.31 or 2.34 are to be considered meaningful only as 2.3, the extra digit being essentially a random guess), you won't need a huge sample before you may feel fine with a simple normal model. Statistics, since it straddles between sophisticated abstract math and practical, down to Earth applications, rarely provides a clean cut prescription: common sense will be your most valuable asset in this field.

4 A More Detailed Example

Check out the accompanying spreadsheet file (in *.gnumeric, *.ods – Open/LibreOffice – and *.xls formats), for a one-tailed test setup, both in the normal and in the Student case. In particular, the different coding of normal and Student distribution functions requires different formulas for the two cases.