z-test, *t*-test

Kenneth A. Ribet

Math 10A
November 28, 2017

This is the last week of classes.

RRR Week December 4–8:

This class will meet in this room on December 5, 7 for structured reviews by T. Zhu and crew.

Final exam
Thursday evening, December 14, 7–10PM.

If your GSI is Teddy Zhu or Freddie Huang, you are in 120 Latimer.

Otherwise you are in 1 Pimentel.

More on the exam

No devices or books are allowed. Each student can bring one two-sided standard-sized sheet of paper to the exam room, but otherwise the exam is closed-book.

The final exam is intended to be comprehensive, but it's easy to imagine that there will be more questions on the final third of the course (meaning the material discussed beginning October 24) more than on either of the first two thirds of the course.

Don't even think about submitting a solution with no words of explanation. Explanations pay a crucial role in possible partial credit for solutions that are correct in principle but flawed in some way. A "correct answer" with no supporting words and sentences is very unlikely to earn full credit.

Little-known fact: my first Faculty Club student breakfast was held on February 1, 2013. It was for Math 55. I intend to organize a breakfast on February 1, 2018.

To be informed of spring semester Faculty Club events, please send me email asking that your email address be added to my “spring dining” list.

Breakfast Friday, December 1 at 8AM. One or two slots may still be available—sign up!

Pop-in lunch on Friday, December 1 at High Noon (Faculty Club).

The Faculty Club gets dressed up for the holidays.

The last topic of last Tuesday

A random sample of 1562 students was asked to respond on a scale from 1 (strongly disagree) to 7 (strongly agree) to the proposition: "Advertising helps raise our standard of living." The sample mean response was 4.27 and the sample standard deviation was 1.32. Decide whether or not to reject this (null) hypothesis: the mean for the full population of 1562 undergrads is 4.

We imagine that undergraduates respond to the proposition by rolling a 7-sided biased die, with each side having an associated probability that we don't know. The probability space is then

$$\Omega = \{ 1, 2, 3, 4, 5, 6, 7 \}$$

and there's an implicit random variable $X : \Omega \to \{ \text{numbers} \}$ that takes an outcome i to the number i. We admit as a *null hypothesis* that X has mean $\mu = 4$, and we estimate the standard deviation σ of X to be 1.32 because we have no better idea of what σ might be.

As on November 14, we consider

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n},$$

where $n = 1562$ and the variables $X_1, \ldots, X_n$ correspond to the 1562 rolls of the 7-sided die, one roll for each student. The function $\overline{X}$ is a random variable defined on the space of all possible outcomes of 1562 rolls of a 7-sided die; the space has 7^{1562} elements.

A review: Dividing by n in the formula for $\overline{X}$ amounts to taking the average or mean value of the responses of the students who were sampled. The mean of $\overline{X}$ is still 4, but the standard deviation of $\overline{X}$ is $\frac{\sigma}{\sqrt{n}}$.

The particular value of $\overline{X}$ that we observe from our sample is the *sample mean response*, namely $\overline{x} = 4.27$.

The Central Limit theorem suggests that the random variable $Z = (\overline{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma}$ is distributed like the standard normal variable.

If the particular value of Z is very far from 0 (the mean of the standard normal), then we reject the hypothesis that the mean is μ.

This is called the Z-test, or z-test.

The variable Z is just a synonym for $(\overline{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma}$.

The value of Z in this case is

$$z = (4.27 - 4) \cdot \frac{\sqrt{1562}}{1.32} \approx 8.08,$$

which is enormous. (According to Table 7.3 on page 566, the probability of being more than 3 standard deviations from the mean is 0.001.)

We were right to reject the null hypothesis last Tuesday!

Issues

There are two problems with this failing analysis:

- We conflated the sample standard deviation with the actual standard deviation, which we don't know.
- We invoked the Central Limit Theorem, which says that Z is distributed like a standard normal variable when n is "infinitely big," even though n was 1562, which is finite.

Sad!

Not on the exam

The number 1562 is “big enough” that we actually don’t need to worry.

Unfortunately, a precise statement in this direction is beyond the scope of the course.

Next question

What if there were 156 students, or 15, or two, or even 1?

> *A random sample of ten undergraduates enrolled in marketing courses was asked to respond on a scale from 1 (strongly disagree) to 7 (strongly agree) to the proposition: "Advertising helps raise our standard of living." The sample mean response was 4.27 and the sample standard deviation was 1.32. Can we reject the hypothesis: the mean* μ *for the full population of 1562 undergrads is 4?*

We will explore this question from the perspective of the `Prob_Stat` notes that are online in `bCourses`.

Imagine that we have n responses $x_1, \dots, x_n$. The *sample mean* response is their average $\overline{x} = \dfrac{x_1 + \cdots x_n}{n}$. Their *sample standard deviation* is the non-negative number s whose square is

$$s^2 = \frac{1}{n-1}\left((x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2\right).$$

Dividing by $n-1$ instead of n matters numerically when n is small, although we don't care much about the difference when n is larger.

As in the previous slides, there are random variables $X_1, \ldots, X_n$ in the picture, and the extremely mean random variable $\overline{X} = \dfrac{X_1 + \cdots + X_n}{n}$. The numbers $x_1, \ldots, x_n$ and $\overline{x}$ are the numerical values of $X_1, \ldots, X_n$ and $\overline{X}$ for a particular sequence of rolls of the die (i.e., a sequence of specific student responses).

Let S be the square root of

$$S^2 = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right).$$

Then the expected value of S^2 is $\mathrm{Var}[X]$ and the expected value of S is the standard deviation of X. (We saw this in HW for $n = 2$, $n = 3$.)

That's why we divide by $n - 1$ instead of by n.

Not on the exam

Here's a discussion of the case $n = 2$ when we no longer think we're flipping coins. The random variables X_1 and X_2 are independent "copies" of X; their expected values are both $E[X]$. Since $\overline{X} = \dfrac{X_1 + X_2}{2}$,

$$(X_1 - \overline{X})^2 = (X_2 - \overline{X})^2 = \frac{(X_1 - X_2)^2}{4}, \quad S^2 = \frac{(X_1 - X_2)^2}{2}.$$

In view of the independence of X_1 and X_2,

$$\begin{aligned} E[S^2] &= \frac{1}{2} E[X_1^2 + X_2^2 - 2X_1 X_2] \\ &= \frac{1}{2}(E[X^2] + E[X^2] - 2E[X]^2) \\ &= \mathrm{Var}[X]. \end{aligned}$$

Recall that we introduced

$$Z = (\overline{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma},$$

where μ was the hypothetical mean (4 in this case) and σ was the standard deviation. We really didn't know σ but just used the sample standard deviation s as a proxy for it. We said that Z was (for all practical purposes) normally distributed because n was big (1562).

We now introduce instead

$$T = (\overline{X} - \mu) \cdot \frac{\sqrt{n}}{S},$$

where S is the random variable that represents the sample standard deviation. Abandoning the fantasy that n is as good as infinite, we ask how T is distributed. The answer will depend on n.

Numerically, nothing has really changed because we were approximating σ by s when we calculated

$$(4.27 - 4) \cdot \frac{\sqrt{1562}}{1.32} \approx 8.08.$$

If there are n students ($n = 15, 10, 2$, whatever), we are still calculating

$$(4.27 - 4) \cdot \frac{\sqrt{n}}{1.32}.$$

For example, for $n = 10$, this number is 0.65.

The number that we get is called t, and we say that the random variable T is distributed according to Student's t-distribution with $n - 1$ degrees of freedom.

Not on the exam

We say it because it's a theorem. See Wikipedia for more info and background.

Not on the exam, but it's fun

Student was a dude, William Sealy Gosset, 1876–1937. He worked for a brewery and "Student" was his statistical pen name.

Not on the exam, but it's not a big deal

The number of degrees of freedom in the distribution is usually written ν (the Greek lower-case letter "nu"). Student's t-distribution with ν degrees of freedom has PDF

$$\frac{1}{C} \cdot (1 + \frac{x^2}{\nu})^{-(\nu+1)/2},$$

where $C = \int_{-\infty}^{\infty} (1 + \frac{x^2}{\nu})^{-(\nu+1)/2} \, dx$. The area C can be calculated and the formulas are in Wikipedia.

That's possibly all you want to know—but you should definitely know what's on this slide.

Not on the exam

Here are the formulas for C: When ν is even,

$$C = \frac{2 \cdot 4 \cdots (\nu - 4)(\nu - 2) 2\sqrt{\nu}}{3 \cdot 5 \cdot 7 \cdots (\nu - 3)(\nu - 1)}.$$

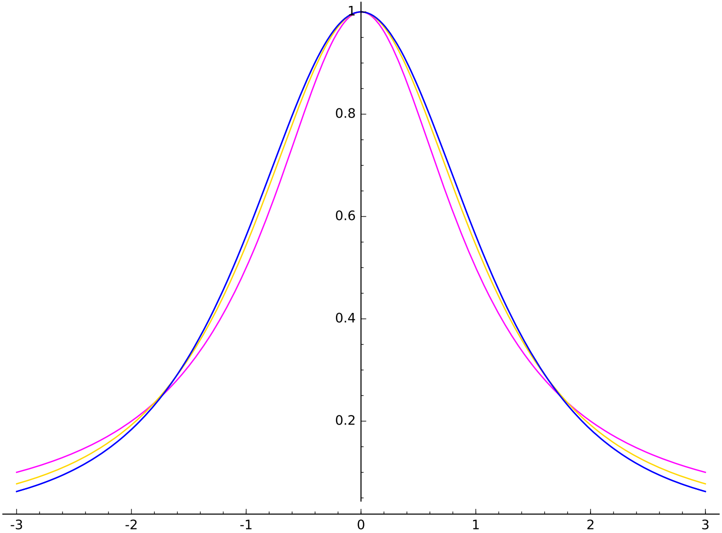
For example, when $\nu = 2$, $C = 2\sqrt{2}$; also, $C = \dfrac{8}{3}$ when $\nu = 4$.

When ν is odd,

$$C = \frac{3 \cdot 5 \cdots (\nu - 4)(\nu - 2)\pi\sqrt{\nu}}{2 \cdot 4 \cdots (\nu - 3)(\nu - 1)}.$$

I hope that this is all right—I just copied the formulas from Wikipedia and might have made a mistake in copying.

The graphs of $(1 + \frac{x^2}{\nu})^{-(\nu+1)/2}$:



Color code: $\nu = 1$, magenta; $\nu = 2$, gold; $\nu = 3$, blue.

The graphs are symmetrical about the y-axis because of the "x^2" in the formula. Although they look qualitatively like the PDF of the standard normal, their tails are much fatter than the tails of the normal PDF when ν is small.

To get the PDFs from these functions, you need to divide by the numbers C.

Exercise: Compute $\lim_{\nu\to\infty}(1+\frac{x^2}{\nu})^{-(\nu+1)/2}$.

The analogue of table 7.3 on page 566

After we calculate a t-statistic, we want to know the probability of getting that number or something even further from the mean (which is 0).

The probability depends on ν.

In the old days, people consulted a book of tables. Now we can look online (or calculate locally, using "technology"). I will use the stat department online t-statistic calculator.

Recall that we have to calculate

$$(4.27 - 4) \cdot \frac{\sqrt{n}}{1.32},$$

where n is the number of students. We unleash technology on this number, using $n - 1$ as the number of degrees of freedom.

Take $n = 10$, $\nu = 9$, $t = 0.65$. Then

$$P(-0.65 \leq T \leq 0.65) \approx 46.81\%,$$

so we can't reject the null hypothesis. (We reject it if the probability is more than 95%.)

For 15 students, $t \approx 0.79$, and the probability is 55.75%.

For 36 students, $t \approx 1.82$, prob $\approx 92.27\%$.

For 49 students, $t \approx 2.12$, prob $\approx 96.08\%$ and we can reject the hypothesis that the mean response is 4.

This is not surprising because the distribution is pretty normal-looking when $\nu = 48$, and the z-value for rejection in the normal case is $z = 2$ (as explained last Tuesday).

For 15 students, $t \approx 0.79$, and the probability is 55.75%.

For 36 students, $t \approx 1.82$, prob $\approx 92.27\%$.

For 49 students, $t \approx 2.12$, prob $\approx 96.08\%$ and we can reject the hypothesis that the mean response is 4.

This is not surprising because the distribution is pretty normal-looking when $\nu = 48$, and the z-value for rejection in the normal case is $z = 2$ (as explained last Tuesday).

For 15 students, $t \approx 0.79$, and the probability is 55.75%.

For 36 students, $t \approx 1.82$, prob $\approx 92.27\%$.

For 49 students, $t \approx 2.12$, prob $\approx 96.08\%$ and we can reject the hypothesis that the mean response is 4.

This is not surprising because the distribution is pretty normal-looking when $\nu = 48$, and the z-value for rejection in the normal case is $z = 2$ (as explained last Tuesday).

A takeaway

Suppose that we had only four responses and that the sample mean and standard deviation were 5.32 and 1.32. Then the number

$$(5.32 - 4) \cdot \frac{\sqrt{4}}{1.32}$$

would be 2, which is just large enough to reject the null hypothesis under the assumption of normality.

In the Student world, for 3 degrees of freedom, the probability that t is between -2 and $+2$ is vastly less than 95%: it's only 86.07%. Thus there is no way to reject the null hypothesis given the four responses, even though their average (5.32) is strikingly higher than 4. In fact, the probability that t is between -3 and 3 is still less than 95%—it's 94.25%.

What t-value is needed?

For the applet that I consulted, one enters a degree of freedom and t-values; the applet returns the probability.

If you start with a number of degrees of freedom and a probability threshold, you might ask what t-value is needed. You can consult this applet, which I found with a quick google search. There must be many others.