Section 14: Loose ends

STOR 155.02, Spring '21

updated 2021-05-04

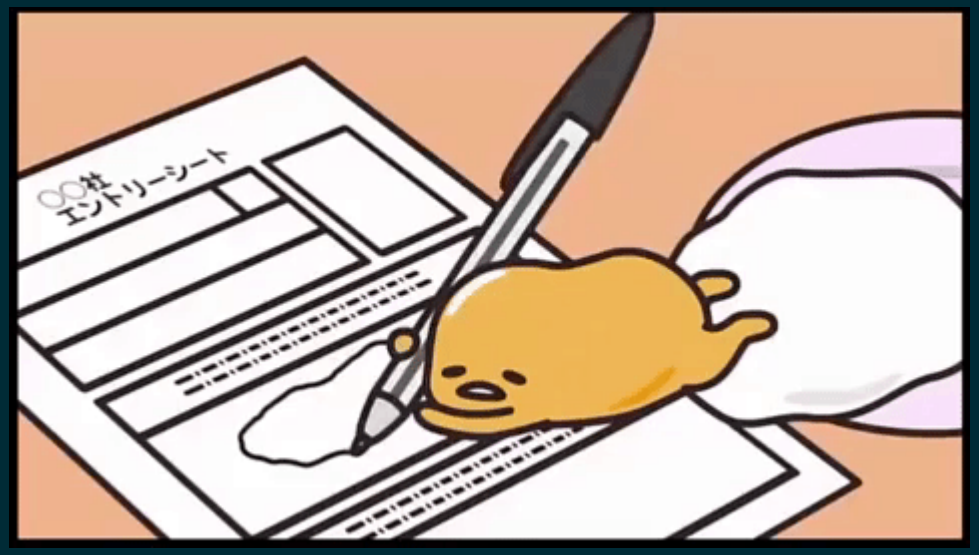
What you will learn

- t-tests
- why t-tests aren't worth much for this course

Resources

- Textbook ch 6.3, 6.4 and 7.1

it's the end of the semester, finally



please do the course evaluations

Final project

due May 11, 3p.m. sharp

instructions posted right after class

Special office hours

1:45-2:45 pm Thursday

for final project Qs only

Small-sample hypothesis tests

So the Central Limit Theorem is pretty good

but what if your sample is too small for the CLT to hold?

Some rules of thumb from textbook

For n random samples from Bernoulli(p)

$$np \geq 10, \quad \text{and} \quad n(1-p) \geq 10$$

meaning you see enough successes and enough failures. Textbook ch 7.1.1.

but rules of thumb are 💩

better to evaluate CLT assumptions (might be hard)

better still to get more random samples (often less hard)

Easy fix: Swap normal distribution for t-distribution

For example in a **difference of means test**

$$p = 2P\left(T_d < -\frac{|\hat{m}_{n_1} - \hat{m}_{n_0}|}{\hat{s}_{n_1,n_0}}\right)$$

where T_d is a t-distributed random variable with d 'degrees of freedom'.
$d = n_1 + n_2 - 2$ in the two-sample difference of means test.

A **similar formula holds for the one-sample test**, replacing the standard normal Z with T_d where $d = n - 1$.

with a computer

Excel: `T.DIST`

Python: `scipy.stats.t`

and there are even baked-in functions for doing t-tests or z-tests. But **make sure you use them correctly** as there are several versions with various options.

I  CLT

but really, this isn't worth much talk

the t-distribution is almost identical to the normal for any reasonable sample size

You should just get more data rather than worry about the t distribution

worry more about

independence assumption of sampling

more complex requirements for more complex hypothesis test methods

which can be hard in practice to evaluate

in rare cases, small sample sizes are unavoidable

rare events

more generally: when data collection is very very expensive


Five more minutes