DSC 152: Applied Statistical Data Analysis and Inference

Lecture #2
The t-Test and Type I Error Estimation

Thursday, April 2
Spring Quarter 2026
Peter Chi

Last time

Let X be the number of heads when we flip a coin 6 times. We have:

H_0: the coin is fair; $p = 0.5$

H_A: the coin is unfair; $p \neq 0.5$

where p is the true probability of heads with this coin.

With a rejection region of $\{X = 0, X = 1, X = 5, X = 6\}$, the probability of making a Type I Error was:

```r
dbinom(x=0, size=6, prob=0.5) + dbinom(x=1, size=6, prob=0.5) +
  dbinom(x=5, size=6, prob=0.5) + dbinom(x=6, size=6, prob=0.5)
```

```
## [1] 0.21875
```

Last time

What if we used a smaller rejection region, say of $\{X = 0, X = 6\}$?

Then the probability of making a Type I error becomes:

```
dbinom(x=0, size=6, prob=0.5) + dbinom(x=6, size=6, prob=0.5)
```

```
## [1] 0.03125
```

Side question: if shrinking the rejection region leads to a lower probability of making a Type I Error, then why don't we always just use the smallest possible rejection region?

Type I Error with a t-Test

Suppose we want to know if there is evidence that the average amount of sleep that a UCSD student gets is <u>different from</u> 6 hours per night. We obtain the following data on a random sample of five students and ask them how many hours of sleep they got on the previous night, with the data below:

$$3, 7, 1, 2, 2$$

Type I Error with a t-Test

Suppose we want to know if there is evidence that the average amount of sleep that a UCSD student gets is <u>different from</u> 6 hours per night. We obtain the following data on a random sample of five students and ask them how many hours of sleep they got on the previous night, with the data below:

$$3, 7, 1, 2, 2$$

Our hypotheses are:

$$H_0 : \mu = 6$$
$$H_A : \mu \neq 6$$

Now how do we make a decision?

Type I Error with a t-Test

The t-Test:

We calculate:

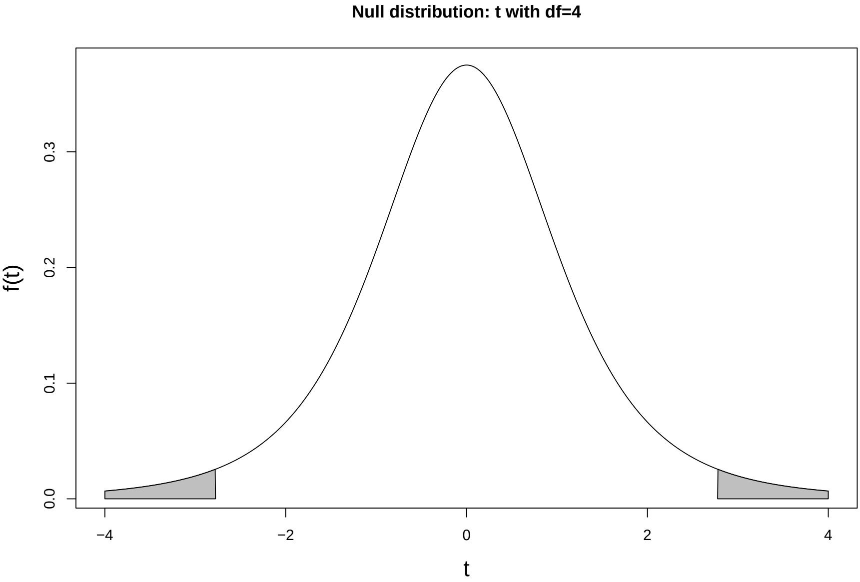
$$t_s = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \ldots$$

```
sleep <- c(3, 7, 1, 2, 2)
t_s <- (mean(sleep) - 6) / (sd(sleep) / sqrt(5))
t_s
```

```
## [1] -2.860388
```

We then compare this to its null distribution. . .

Type I Error with a t-Test



Type I Error with a t-Test

At $\alpha = 0.05$, the critical values that delineate the rejection region are:

```
qt(0.025, df=4)
```

```
## [1] -2.776445
```

```
qt(0.025, df=4, lower.tail=FALSE)
```

```
## [1] 2.776445
```

Type I Error with a t-Test

At $\alpha = 0.05$, the critical values that delineate the rejection region are:

```
qt(0.025, df=4)
```

```
## [1] -2.776445
```

```
qt(0.025, df=4, lower.tail=FALSE)
```

```
## [1] 2.776445
```

So our test statistic of -2.86 is in the rejection region, meaning that we would reject H_0. Specifically, the p-value is:

```
pt(t_s, df=4) * 2
```

```
## [1] 0.04591151
```

Type I Error with a t-Test

Also note that there is a built-in R function that does the t-Test for us and will give the same exact answer as the manual calculations from the previous slides:

```
t.test(sleep, mu=6)
```

```
##
##  One Sample t-test
##
## data:  sleep
## t = -2.8604, df = 4, p-value = 0.04591
## alternative hypothesis: true mean is not equal to 6
## 95 percent confidence interval:
##  0.08803981 5.91196019
## sample estimates:
## mean of x
##         3
```

Type I Error with a t-Test

And if we just want to extract the p-value:

```
model1 <- t.test(sleep, mu=6)
names(model1)
```

```
##  [1] "statistic"   "parameter"   "p.value"     "conf.int"
##  [6] "null.value"  "stderr"      "alternative" "method"
```

```
model1$p.value
```

```
## [1] 0.04591151
```

So either way works, and again, we reject H_0 at $\alpha = 0.05$. But...

Type I Error with a t-Test

Might we have made a Type I Error?

We can't know for sure in any given case, but we CAN know the probability of making a Type I Error.

If the conditions of the test are met, the probability of making a Type I error should just be the α level of the test (here, 0.05). Why?

https://pollev.com/chi

Type I Error with a t-Test

But wait, what conditions?

https://pollev.com/chi

Type I Error with a t-Test

And what are the consequences of performing a test when its conditions are not met?

What happens if the conditions of your test were not met?

Primarily, your probability of making a Type I error may be inflated.

In other words, if you thought you were doing a 0.05-level test, but the conditions required for your test were not met and you still did the test anyways, then it might actually not be a 0.05-level test!

(P.S. this is really bad!!)

Type I Error with a t-Test

Sidenote: why is it called a t-Test (or also the "Student's t-Test")?

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in

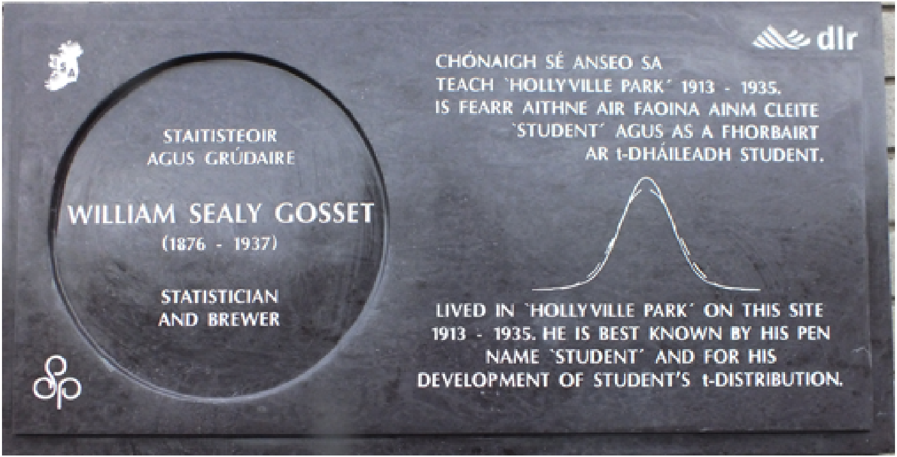
Type I Error with a t-Test

Sidenote: why is it called a t-Test (or also the "Student's t-Test")?

- If σ is known and the conditions just mentioned are met, $\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$ follows a normal distribution. But usually, σ is NOT known! Recall:
  - σ is the population standard deviation
  - s is the sample standard deviation
- Prior to this work, people just treated $\frac{\overline{x}-\mu_0}{s/\sqrt{n}}$ like it also follows a normal distribution, knowing that it didn't with small samples, but not knowing how to fix it.
  - Since s is calculated from the data, s itself has variability which is what messes things up.
- What "Student" did was derive the probability distribution of $\frac{\overline{x}-\mu_0}{s/\sqrt{n}}$ for any sample size (e.g. the one on Slide 6 when $n = 5$).

So who was "Student" and why did they call themselves that?

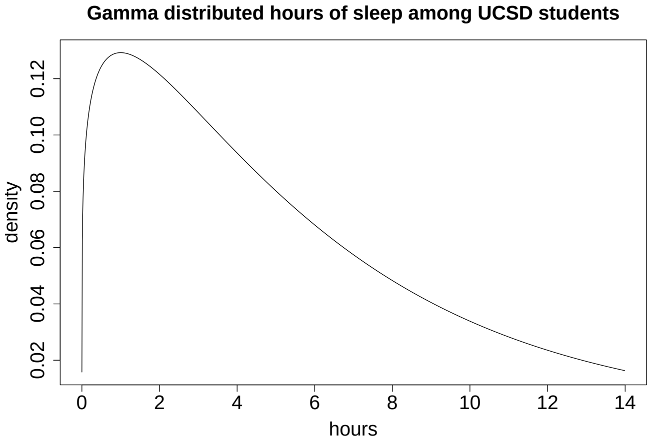
Type I Error with a t-Test




CHÓNAIGH SÉ ANSEO SA
TEACH 'HOLLYVILLE PARK' 1913 - 1935.
STAITISTEOIR
AGUS GRÚDAIRE
WILLIAM SEALY GOSSET
(1876 - 1937)
STATISTICIAN
AND BREWER
1913 - 1935.
Ah, beer.
The cause of and the solution
to all of life's problems.

Determining Type I error rates

Example: let's suppose that in the population, the distribution of hours of sleep follows a Gamma(1.2, 5) distribution (don't worry too much about the specifics; just know that this is a skewed distribution):



Determining Type I Error rates

The Gamma distribution shown in the previous slide does have a mean of $\mu = 6$. Based on this, can we figure out what the Type I Error rate of a t-Test would be?

Wait what's the Type I Error rate again?

It's the probability of rejecting H_0 if H_0 is actually true…

So, here, it would be the probability that we get a t_s value that is in the rejection region of the t-Distribution.

Determining Type I Error rates

The Gamma distribution shown in the previous slide does have a mean of $\mu = 6$. Based on this, can we figure out what the Type I Error rate of a t-Test would be?

Wait what's the Type I Error rate again?

It's the probability of rejecting H_0 if H_0 is actually true...

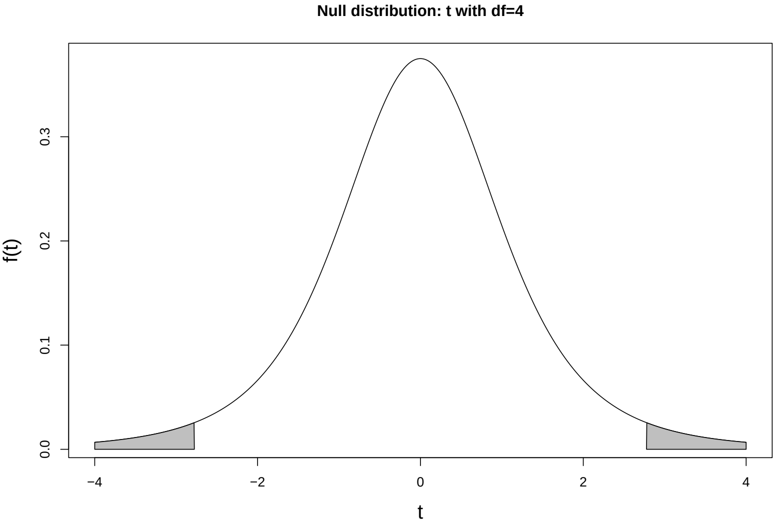
So, here, it would be the probability that we get a t_s value that is in the rejection region of the t-Distribution.

Recall the test statistic t_s:

$$t_s = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

When H_0 is true, this thing follows the t-Distribution on Slide 6, but ONLY IF the conditions are met (which here we know they are not).

The t-Distribution again



The problem: since X is highly skewed, the distribution of t_s under H_0 is going to look quite different from this!

Determining Type I Error rates

Therefore, to calculate the Type I Error rate in this situation theoretically, we would have to know what the distribution of

$$t_s = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

is when X follows a Gamma distribution, and then use that to determine the probability of t_s falling in the rejection region of the original t-Distribution.

Unfortunately...

There is actually no closed-form solution to the distribution of t_s if X follows a Gamma distribution. :(

But, we can go back to our DSC roots and simulate!

Determining Type I Error rates

What do we simulate?

- Simulate a sample of size 5 under the Gamma distribution
- Run the t-Test on that sample, get the p-value
- Check if the p-value is less than 0.05 (in which case, that sample gave a Type I Error)
- Do this repeatedly, count each time that p-value was less than 0.05 and thus gives a Type I Error
- The proportion of times that a Type I Error occurs in the simulations is an estimate of the Type I Error rate.

Determining Type I Error rates

```
count <- 0

for(i in 1:10000){
  gam_data <- rgamma(n=5, shape=1.2, scale=(6/1.2))
  p.val <- t.test(gam_data, mu=6)$p.value
  if(p.val < 0.05){
    count <- count + 1
  }
}

TypeI <- count / 10000
TypeI
```

```
## [1] 0.1071
```

That's more than twice as big as 0.05!

Your Turn

Let's do the same simulation, but under two other scenarios:

- Let X actually follow a normal distribution, with $\mu = 6$ and $\sigma = 1$.
  - The `rnorm` function will be useful.

- Let X follow a Uniform distribution from 2 to 10.
  - The `runif` function will be useful.

In an R Markdown file, find a simulated estimate of the Type I Error rate in each case, again with samples of size $n = 5$, and briefly comment on what you observe.

Recap

- Any parametric statistical test has conditions that must be met for validity.
  - By parametric, we essentially mean one that has distributional conditions on the data in order to have a closed-form null distribution (such as the t-Test)
  - By validity, we primarily mean that a statistical test is valid if its Type I Error rate will in fact be the stated α-level.

- We can evaluate Type I Error rates via simulation. We will extend this to more complicated statistical analyses as we proceed in this course.

Recap

Things still to come:

- Non-parametric tests (e.g., but not limited to, what you've learned in DSC 10 and 80 where you get a p-value via simulating the null distribution) are less dependent on conditions to be valid.
  - By non-parametric, essentially we mean tests that do not rely on distributional conditions on the data in order to have a closed-form null distribution
  - Tests performed via simulation are one example of this, but there are others as well.
- What about statistical power?

Daily Check for today

Upload your .pdf output file from R Markdown to Gradescope, consisting of:

1. The answer to the question on Slide 10.
2. The answer to the question on Slide 11.
3. Your two simulations from the “Your Turn” on Slide 23, with brief commentary on what you observed in each case.

If you have not yet been able to get R Markdown to output to a pdf file, we will be lenient on that for this assignment; please just find some way to upload a single pdf with your answers, code and output.