

Test Statistics

John Kubale

2024-09-06

What is a test statistic?

A measure that allows us to evaluate how consistent your data are with the the null hypothesis. It boils everything down to a single number which quantifies the amount which your data diverges from the null hypothesis. You can use it to calculate a p-value and/or compare it to a critical value to determine whether the data are different enough to warrant rejecting the null hypothesis.

Where do test statistics come from?

There are a variety of test statistics that are used in different situations. Here we will focus on the t statistic (i.e., what is used in t tests), specifically that used in a one-sample t-test.

```
# Let's say we have a class of 40 students with a mean height of 60 inches  
# and a standard deviation of 1. We want to compare this to the national average  
# for this age group of 55 to see if class1 is actually taller than the national  
# average.
```

```
# create dataset for class 1  
set.seed(894894)  
class1 <- rnorm(40, 60, 1)
```

```
# conduct 1 sample t-test  
t.test(class1, mu=55)
```

```
##  
## One Sample t-test  
##  
## data: class1  
## t = 39.696, df = 39, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 55  
## 95 percent confidence interval:  
## 59.88474 60.40928  
## sample estimates:  
## mean of x  
## 60.14701
```

The output above returns a t statistic of 39.7 with 39 degrees of freedom which leads to a p-value of $< 2.2 \times 10^{-16}$. This is much smaller than our typical alpha value of 0.05 so we would reject the null hypothesis

and conclude that class 1 is taller than the national average. But where did we get the t statistic, degrees of freedom, and the p-value?

Let's start with calculating our t statistic. The formula for a one-sample t-test is as follows $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$.

```
# define class mean and national mean
class_mn <- mean(class1)
nat_mn <- 55

# define class standard error
class_se <- sqrt(var(class1)/length(class1))

t <- (class_mn - nat_mn)/class_se
t
```

```
## [1] 39.6955
```

When we plug the numbers into the formula we see that we get the same result. But what does this number mean? Let's look at a situation where there is no difference in the numbers being compared to see what our t statistic looks like then.

```
set.seed(894894)
class2 <- rnorm(40, 55, 1)

class2_mn <- mean(class2)
nat_mn <- 55

# define class standard error
class_se <- sqrt(var(class2)/length(class2))

t.test(class2, mu=55)

##
## One Sample t-test
##
## data: class2
## t = 1.1338, df = 39, p-value = 0.2638
## alternative hypothesis: true mean is not equal to 55
## 95 percent confidence interval:
## 54.88474 55.40928
## sample estimates:
## mean of x
## 55.14701

t <- (class2_mn - nat_mn)/class_se
t
```

```
## [1] 1.1338
```

We can see that for class2 we set the mean of the normal distribution that we're drawing from to 55 (i.e., national mean for this age group). We can see there is essentially no difference in our sample mean of class 2,

55.2, and the national average 55. The mean of class2 is not exactly 55, but this is due to sampling variability – we’re drawing from a distribution with a mean of 55 forty times. Looking at our t statistic, 0.87, we can see this is much smaller than what we saw in the first example with class 1. We can imagine a more extreme case where the means we are comparing are exactly the same which would result in the numerator of the formula for the t statistic, and therefore the t statistic itself being equal to 0. Thinking about this from another perspective, the further away from 0 our t statistic is, the less consistent our data is with the null hypothesis. But at what point do we consider it inconsistent enough to reject the null hypothesis? To do this we need to decide on a threshold **before** conducting the test.

As we’ve discussed in class we set this threshold by defining α and most commonly in the social sciences $\alpha = 0.05$. But how does this relate to Student’s t distribution? To understand this we need to talk about critical value(s). These correspond to the points in a distribution where if our test statistic falls beyond we would reject the null hypothesis. When $\alpha = 0.05$ we are rejecting the null hypothesis anytime the test statistic falls in the 5% of the distribution furthest from the mean. In the case of a two-sided test this 5% is divided equally to the right and left of the mean (2.5% on each side) as reflected in the red-lined regions below. In the case of a one-sided test, the entire 5% would be to the right *or* left (depending on the hypothesis being tested).

These critical values will vary depending on what we set α to and our sample size (via the degrees of freedom). So how do we calculate them? Previously students had to look them up in tables that were painstakingly created by pioneering statisticians. Fortunately R has provided a number of helpful functions to make this simple and easy.

```
# Since we're running a t test we will use the qt() function. There are a number
# of options in this function, but we're going to focus on just 3: p, df,
# and lower.tail.
```

```
# The first argument p corresponds to our alpha value or the probability we're
# setting as our threshold. The df argument tells R the degrees of freedom to
# use, calculated as n-1 where n is our sample size. And the lower.tail argument
# tells R a bit more about the probability we're interested in more, or more
# roughly which side of the mean (left or right) we're looking at.
```

```
# Let's calculate the critical value for our first example class 1.
length(class1) # sample size is 40
```

```
## [1] 40
```

```
df_class1 <- length(class1)-1 # calculate df and save as df_class1
```

```
# set alpha to 0.05 and save as alph as alpha is already used by an
# R function in ggplot
alph <- 0.05
```

```
# Now we can use the qt() function to calculate our critical value. We'll
# conduct a two-sided test so we are dividing alpha by 2. This way half is on
# the right and the other half is on the left. We should also include
# "lower.tail=T" as we're interested in the probability beyond our critical
# value (see what happens when we don't include this below).
qt(0.05/2, df_class1, lower.tail = F)
```

```
## [1] 2.022691
```

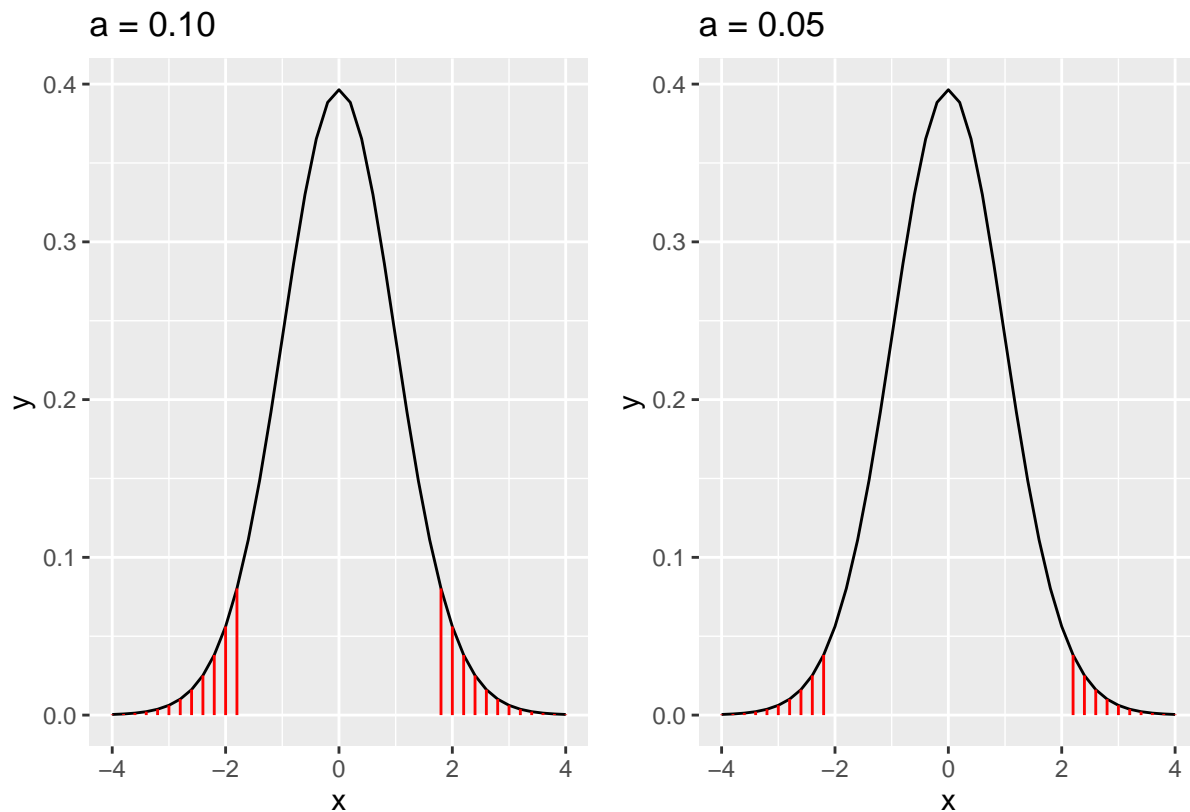
Now what would happen if we choose a different α ? Let’s look to see what this means for both our critical value and the resulting rejection region.

First, let's set $\alpha = 0.1$ and see what this does to our critical value for the class 1 example.

```
qt(0.1/2, df_class1, lower.tail = F)
```

```
## [1] 1.684875
```

When we increase to $\alpha = 0.1$ we get a critical value of 1.7, smaller than the 2.0 we got when $\alpha = 0.05$. Do you think this makes us more or less likely to reject the null hypothesis? Let's look at the rejection region graphically to see what this tells us. Remember if a value falls in the red-lined region we would reject the null hypothesis.



What about if we decrease to $\alpha = 0.01$?

```
qt(0.01/2, df_class1, lower.tail = F)
```

```
## [1] 2.707913
```

This time we get a larger critical value of 2.7. Let's look at what this does to our rejection regions.

