Lecture 11: CI Wrap-up, & Intro to Hypothesis Testing

STAT 324

Ralph Trane
University of Wisconsin–Madison

Spring 2020

Practicalities

- Midterm I on Tuesday 3/3 in class (4.00p-5.15p in Van Vleck B130)
    - Notes are allowed
    - Calculator recommended
    - No connected devices!
    - Practice problem set on Canvas
    - Q/A Review tomorrow and Thursday
    - No homework assigned this week

Misc.

- Recap of t-distribution/introduction of t-table
- Comment on homework questions re: sample size.

Recap of Confidence Intervals

1. Want to do better than just a single "best guess" (= point estimate)
2. Find a good estimator for parameter of interest
   - think $\bar{X}$ for μ, or P for π
3. Find distribution of some quantity involving estimator and parameter of interest
   - $\frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{\bar{X}-E(\bar{X})}{\widehat{\mathrm{SD}}(\bar{X})}$
4. Find critical values in the distribution
   - i.e. cut-off $\alpha/2$ on each side of distribution
5. Rearrange $P\left(x_1 < \text{ quantity from 3 } < x_2\right)$ until parameter of interest is in the middle
   - for example, $P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{s/\sqrt{n}} < z_{\alpha/2}\right)$ becomes $P(\bar{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\alpha}\frac{s}{\sqrt{n}})$.

Recap of Confidence Intervals

The hard part: how do we find the distribution of quantity?

Often, how do we find the distribution of $\frac{\bar{X}-E(\bar{X})}{\widehat{\mathrm{SD}}(\bar{X})}$?

A few tricks:

1. If $\bar{X} \sim N$, then $\frac{\bar{X}-E(\bar{X})}{\widehat{\mathrm{SD}}(\bar{X})} \sim t_{n-1}$.
2. If $\bar{X}$ not necessarily normal, then bootstrap.

Two things result in $\bar{X} \sim N$:

1. Data themselves are normal
   - i.e. $X_1, \ldots, X_n \sim N$
2. $n > 30$, because then CLT

Recap of Confidence Intervals

Once we get an interval, how do we interpret it?

- The standard phrase: "We are $(1-\alpha)\cdot 100\%$ confident the interval contains the true value"
- Confident = repeating the process many times will cover the truth $(1-\alpha)\cdot 100\%$ of the time
- The CI is a range of values we find plausible to be the truth given the data (at a certain level of confidence)

Introduction to Statistical Hypothesis Testing

Sometimes we have that one thing we are really interested in: "does it seem likely that the true mean is 5?"

First step in answering this question: write down your *statistical hypotheses*. These always come in pairs:

- the *null* hypothesis is the simplest (and often uninteresting) hypothesis
    - we take this to be "status quo" - we are trying to dismiss this
- the *alternative* hypothesis is what we will test against
    - this is what we will adopt, if we dismiss "status quo"

Example: "is the true mean 5?" would dictate the null hypothesis $H_0 : \mu = 5$. We now have a choice for the alternative:

- $H_A : \mu > 5$
- $H_A : \mu < 5$
- $H_A : \mu \neq 5$

For now, focus on the first one.

Introduction to Statistical Hypothesis Testing

Want to test $H_0 : \mu = 5$ vs. $H_A : \mu > 5$.

Gather data, estimate μ.

```
head(samp,n = 5)
```

```
## # A tibble: 5 x 1
##         x
##     <dbl>
## 1  3.36
## 2  5.05
## 3  5.83
## 4  7.23
## 5  5.17
```

We get $\bar{x}_{\text{obs}} = 6.468$.

Conclusion: since $\bar{x}_{\text{obs}} > 5$, surely $\mu > 5$?

Of course not! We never expect $\bar{X}$ to be **exactly** the true mean, simply "close to it".

Remember, $H_0 : \mu = 5$ is status quo. So the question is, is $\bar{x}_{\text{obs}}$ far enough from 5 that we no longer think $\mu = 5$?

Question: is $\bar{x}_{\text{obs}}$ far enough from 5 that we throw away H_0?

Introduction to Statistical Hypothesis Testing

Question: is $\bar{x}_{\text{obs}}$ far enough from 5 that we throw away H_0?

Question: is $\bar{x}_{\mathrm{obs}}$ far enough from 5 that we throw away H_0?

Question: is $\bar{x}_{\text{obs}}$ far enough from 5 that we throw away H_0?

Where do we draw the line?

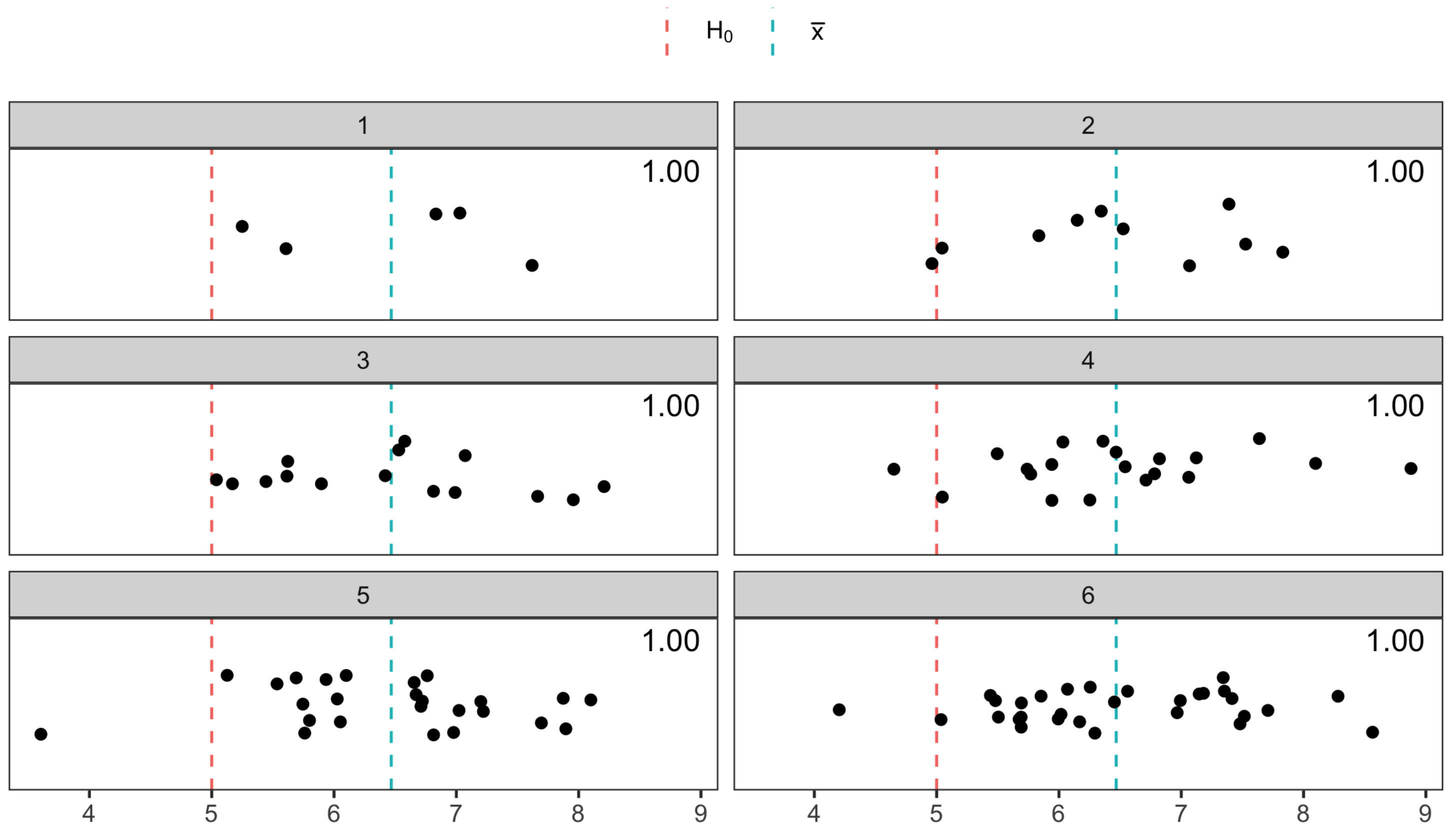
Introduction to Statistical Hypothesis Testing

Question: is $\bar{x}_{\text{obs}}$ far enough from 5 that we throw away H_0?

Where do we draw the line? Kind of comparing difference to variation.

Introduction to Statistical Hypothesis Testing

Comparing $\bar{X} - \mu$ to a measure of variation. Which of the following provide most evidence against the null $H_0 : \mu = 5$?



Would like our measure to have some dependency on n.

Introduction to Statistical Hypothesis Testing

What if we use $\frac{\bar{X}-\mu}{\widehat{\mathrm{SD}}(\bar{X})}$?

- measures deviation from hypothesized mean relative to measure of variation
- if difference $\bar{X} - \mu$ is large, this quantity is large
- since $\widehat{\mathrm{SD}}(\bar{X}) = s/\sqrt{n}$, larger n implies larger value

Intuitively, smart choice!

But when is this "large enough" that we would throw out $H_0 : \mu = 5$?

Introduction to Statistical Hypothesis Testing

Let's pretend that $\mu = 5$. I.e. pretend $E(X_i) = 5$. Then $E(\bar{X}) = 5$.

Let's pretend $\bar{X} \sim N$. (Does this seem outrageous? No, because CLT!)

So, $T = \frac{\bar{X}-5}{\widehat{\mathrm{SD}}(\bar{X})} \sim t_{n-1}$.

This is great! We can now say something about " is the observed $\bar{x}_{\text{obs}}$ far from 5?", **IF** $H_0 : \mu = 5$ is true!

If $\bar{x}_{\text{obs}}$ is close to 5, then T_{obs} is close to 0. (Here T_{obs} is the observed value of T.)

IF $H_0 : \mu = 5$, we can find $P\left(T > T_{\text{obs}}\right)$.

In other words, **IF** $H_0 : \mu = 5$, we can find *the probability of seeing something "more absurd"/"more extreme"/"further away"* than what we observe.

If this probability is small, our $\bar{x}_{\text{obs}}$ is deemed "far from" the hypothesized mean of 5, hence the sample seems to suggest that $\mu = 5$ is not the correct value.

Introduction to Statistical Hypothesis Testing

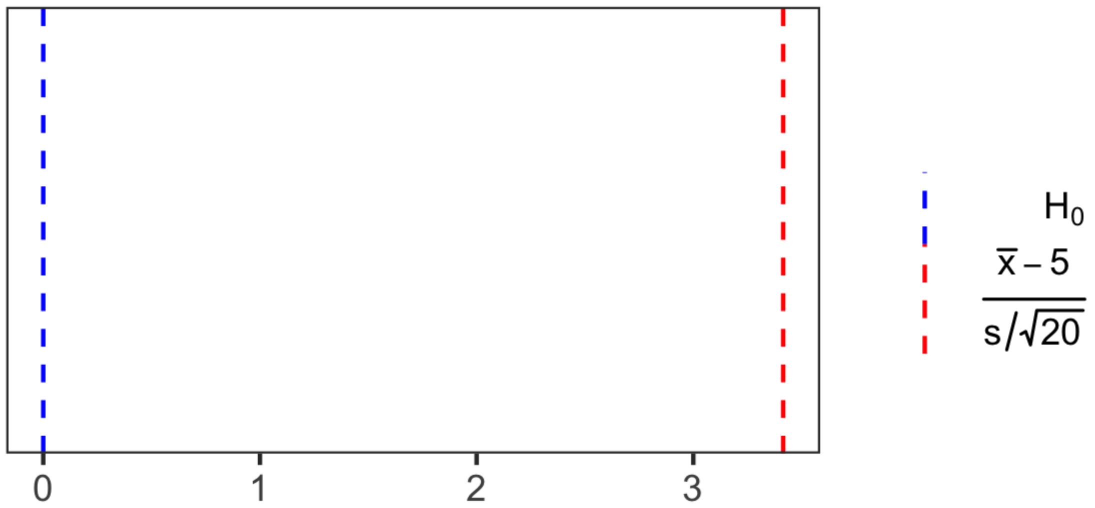
So, we have taken the question "is 6.468 far from 5?", which is highly subjective, and reformulated it as "is the probability of observing something "more extreme" large?", which is... still highly subjective, **BUT** now on a scale we all know and love!

This new question is unitless - it is a probability between 0 and 1.

You might think 0.1 is small, while I would use 0.01 as a cut-off for "small", but at least we are now operating on the same scale AND we can relate to the choice of others.

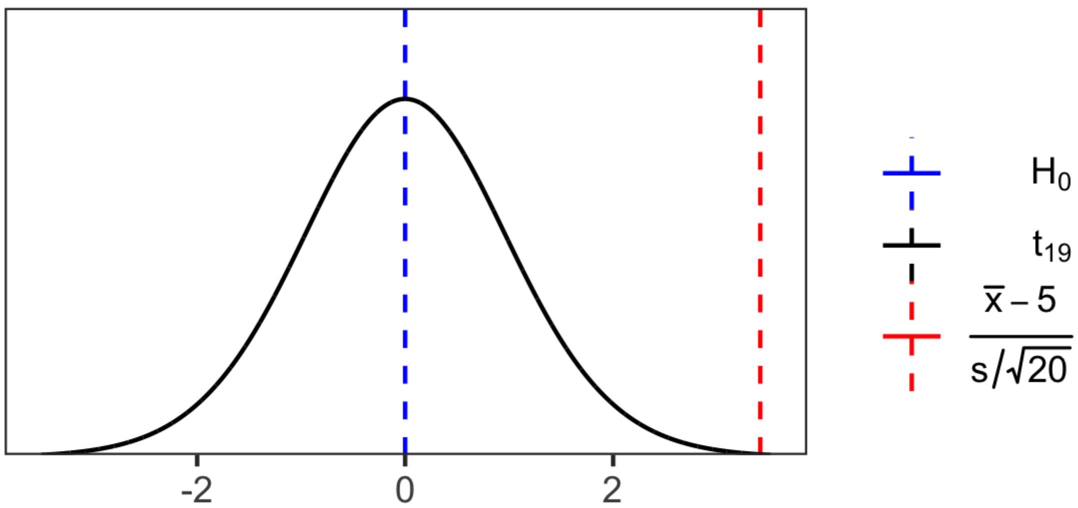
Introduction to Statistical Hypothesis Testing

So, is 6.468 far from 5? Find $T_{\text{obs}} = \frac{\bar{x}_{\text{obs}} - 5}{s/\sqrt{20}} = 3.415$ and see if it is far from 0:



Introduction to Statistical Hypothesis Testing

So, is 6.468 far from 5? Find $T_{\text{obs}} = \frac{\bar{x}_{\text{obs}} - 5}{s/\sqrt{20}} = 3.415$ and see if it is far from 0, when compared to the t_{n-1}-distribution:



Introduction to Statistical Hypothesis Testing

The most general strategy:

1. Set up null and alternative hypotheses:
   - $H_0 : \mu = \mu_0$ vs.
     - $H_A : \mu > \mu_0$ or $H_A : \mu < \mu_0$ or $H_A : \mu \neq \mu_0$
2. Pick cut-off for "small probability"
   - called *significance level* and is denoted by α
   - often 0.05, 0.01, or 0.001
3. Find good good *test statistic*
   - example: when wondering about true mean, $T = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$
4. Find distribution *assuming* H_0 *is true!*
   - if $H_0 : \mu = \mu_0$ is true, then $T \sim t_{n-1}$
5. Find the *p-value*: probability of being "more extreme"
   - if $H_A : \mu > \mu_0$, "more extreme" = even larger, so find $P(T > T_{\text{obs}})$
   - if $H_A : \mu < \mu_0$, "more extreme" = even smaller, so find $P(T < T_{\text{obs}})$
   - if $H_A : \mu \neq \mu_0$, "more extreme" = even further away from zero, so find $P(T > |T_{\text{obs}}|) + P(T < -|T_{\text{obs}}|)$

Introduction to Statistical Hypothesis Testing

Performing a hypothesis test results in one of two thing:

1. lots of evidence against the null (i.e. $\bar{x}_{\text{obs}}$ is far from μ above) leads us to *reject* H_0
2. less evidence against the null (i.e. $\bar{x}_{\text{obs}}$ is close to μ above) leads us to *not reject* H_0

Note:

- we **NEVER** accept the null, we **NEVER** accept the alternative.
- we **NEVER** find the truth, we simply reject suggestions.

Introduction to Statistical Hypothesis Testing

We will never know the truth, so mistakes happen:

1. If we reject the null, but the null is actually true, we make a **type I** error.
2. If we do not reject the null, but the null is actually false, we make a **type II** error.

Result\Truth	**H_0 true**	**H_0 false**
Reject	Type I	No mistake
Do not reject	No mistake	Type II

Since we never know the truth, we cannot check if a mistake occurred, but we can control the probability it occurs.

We use

$$P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ true}) = \alpha$$

and

$$P(\text{type II error}) = P(\text{do not reject } H_0 | H_0 \text{ false}) = \beta$$

Introduction to Statistical Hypothesis Testing

Generally, we want α and β to be small.

Closely related to β is the idea of *statistical power*: the probability that we reject H_0 when H_0 is indeed false! I.e.

$$\text{Power} = P(\text{reject } H_0 | H_0 \text{ false}) = 1 - P(\text{do not reject } H_0 | H_0 \text{ false}) = 1 - \beta.$$

So, we want power to be large.

Unfortunately, (for fixed sample size n) decreasing α increases β, and vice versa! So a trade-off to be made - do you want lower probability of Type I error, or lower probability of type II error.

Introduction to Statistical Hypothesis Testing

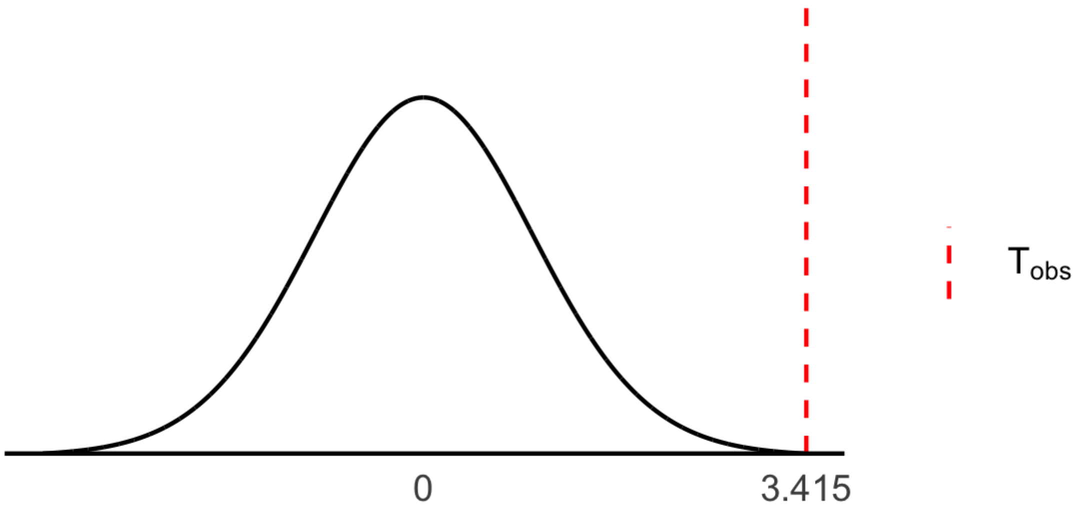
Three more concepts that are closely related: *significance level*, *p-value*, and *rejection region*.

- The *significance level* is your cut-off for what constitutes a small probability.
- The *p-value* is the probability of observing something more extreme **IF** the null hypothesis is true
    - p-value $= P(\text{more extreme}|H_0 \text{ true})$.
    - what it means to be "more extreme" is determined by H_A
- The *rejection region* (RR) is all the values that would result in a p-value smaller than the significance level
    - the opposite of the rejection region is called the *acceptance region*

Introduction to Statistical Hypothesis Testing

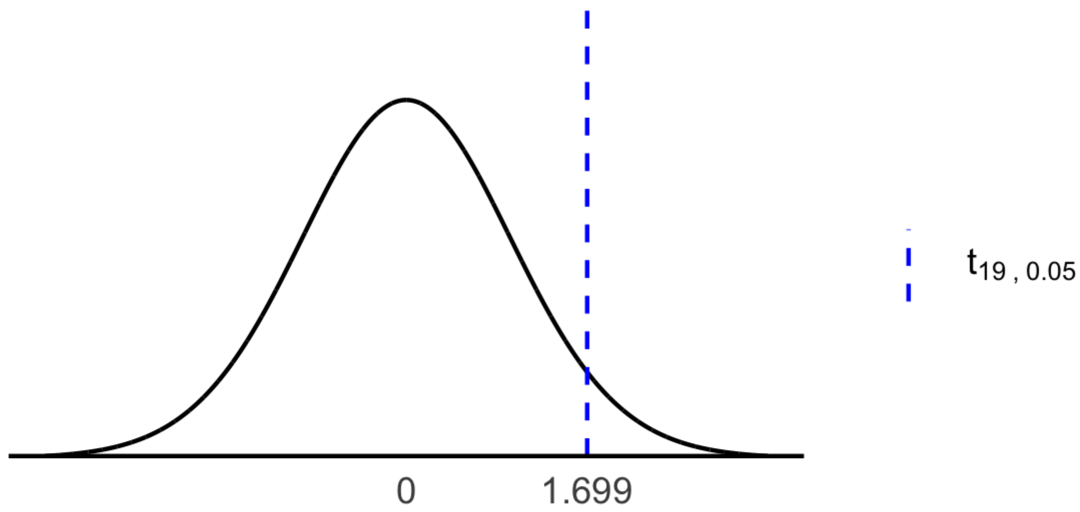
Consider $H_0 : \mu = 5$ vs. $H_A : \mu > 5$.

We might choose a significance level of 0.05. I.e. we would reject if area to the right of the observed value of our test statistic is greater than 0.05.



Consider $H_0 : \mu = 5$ vs. $H_A : \mu > 5$.

We might choose a significance level of 0.05. I.e. we would reject if area to the right of the observed value of our test statistic is greater than 0.05. In this case, it is (p-value = 9.5306084\times 10^{-4}). But we can also ask, where is the cut-off such that the area to the right is *exactly* 0.05:



I.e. we would reject if test statistic is greater than 1.699. So the rejection region on the test statistic scale is $[1.699, \infty)$.

Introduction to Statistical Hypothesis Testing

This is not super useful in the sense that we do not have a good understanding of what a test statistic of, say, 3 is.

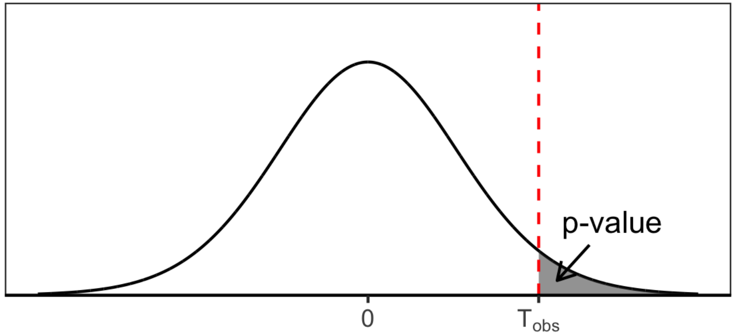
Fortunately, we can do even better! We can translate this back onto the $\bar{X}$ scale.

$T_{\text{obs}} = 1.699$, so $\frac{\bar{x}_{\text{obs}} - 5}{\widehat{\text{SD}}(\bar{X})} = 1.699$. I.e. $\bar{x}_{\text{obs}} = 1.699 \cdot \widehat{\text{SD}}(\bar{X}) + 5 = 5.607$.

We reject when the test statistic is greater than 1.699, which happens when $\bar{x}_{\text{obs}}$ is greater than 5.607. This is something we can actually relate to!

So,

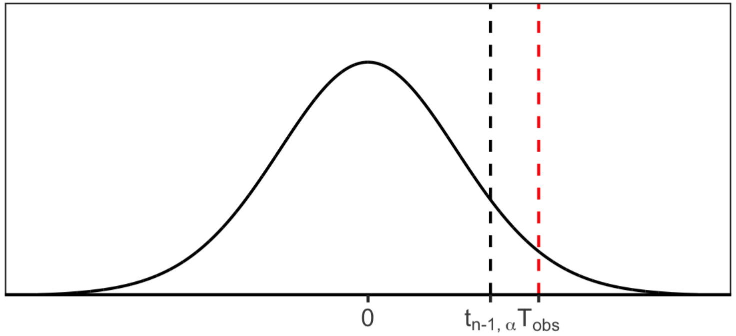
- reject when p-value $< \alpha$



Introduction to Statistical Hypothesis Testing

So,

- reject when p-value $< \alpha$
- happens when $\frac{\bar{x}_{\text{obs}} - 5}{\widehat{\text{SD}}(\bar{X})} = T_{\text{obs}} > t_{19,\alpha} = 1.699$
  - so RR for test statistic = $[1.699, \infty)$



Introduction to Statistical Hypothesis Testing

So,

- reject when p-value $< \alpha$
- happens when $\frac{\bar{x}_{\text{obs}}-5}{\widehat{\text{SD}}(\bar{X})} = T_{\text{obs}} > t_{n-1,\alpha} = 1.699$
    - so RR for test statistic = $[1.699, \infty)$
- happens when $\bar{x}_{\text{obs}} > 5.607$
    - so RR for $\bar{x}$ = $[5.607, \infty)$

Three different scales, but 1-to-1 path between them!