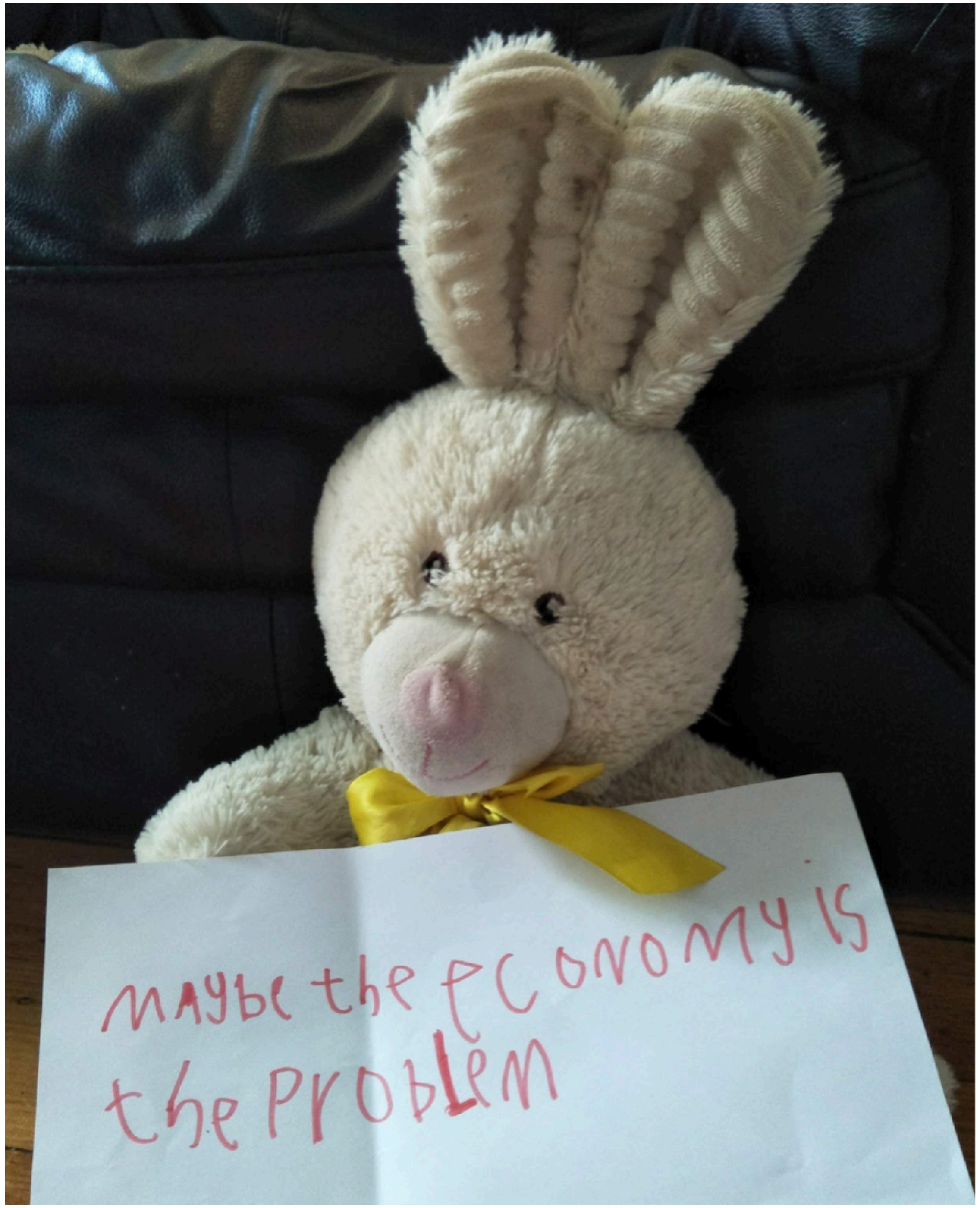
T-tests: Independent samples

Research Methods for Human Inquiry
Andrew Perfors

Today's story…

Flopsy, Shadow, and Bunny realise that they'll only be able to persuade people not to attack the Others if they have an alternate explanation and plan



Today's story…

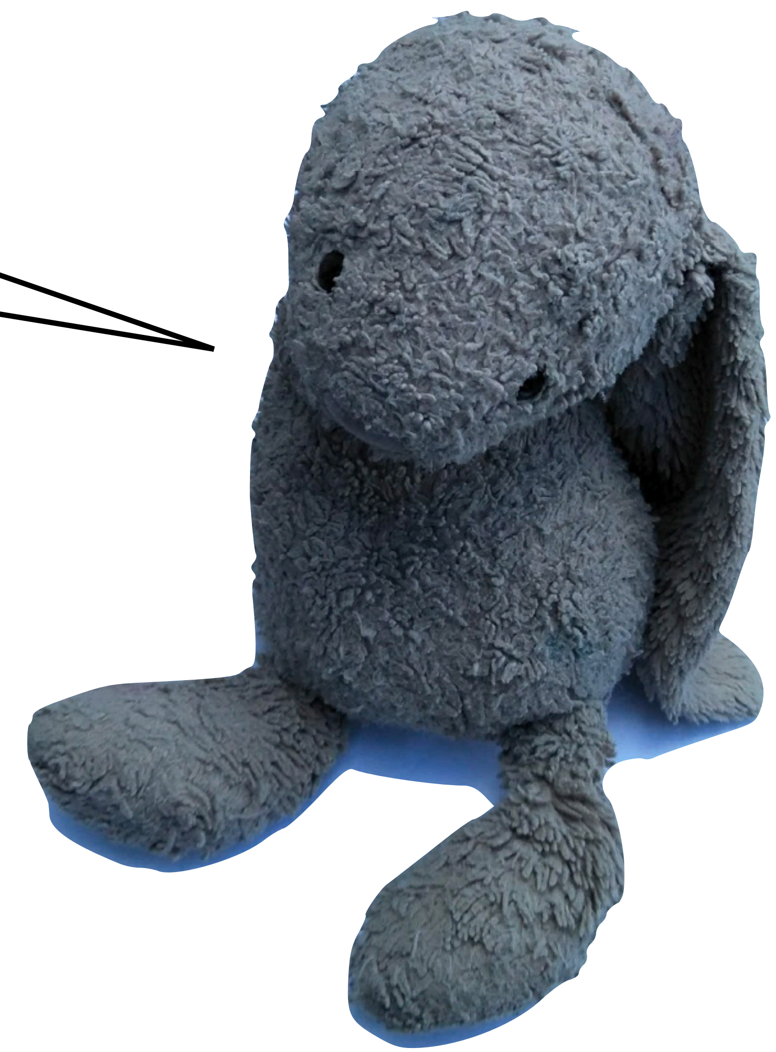
If it's mainly an economic problem, it predicts that poor people should be going hungry and rich people shouldn't

Today's story…

We can ask people about their hunger (relative to last year) and their income level, and see if the two groups are different!

Today’s story…

How about this as the question:



How would you rate your general hunger now relative to last year?

10 = far hungrier than last year
5 = somewhat hungrier than last year
0 = same as last year
-5 = somewhat less hungry than last year
-10 = far less hungry than last year

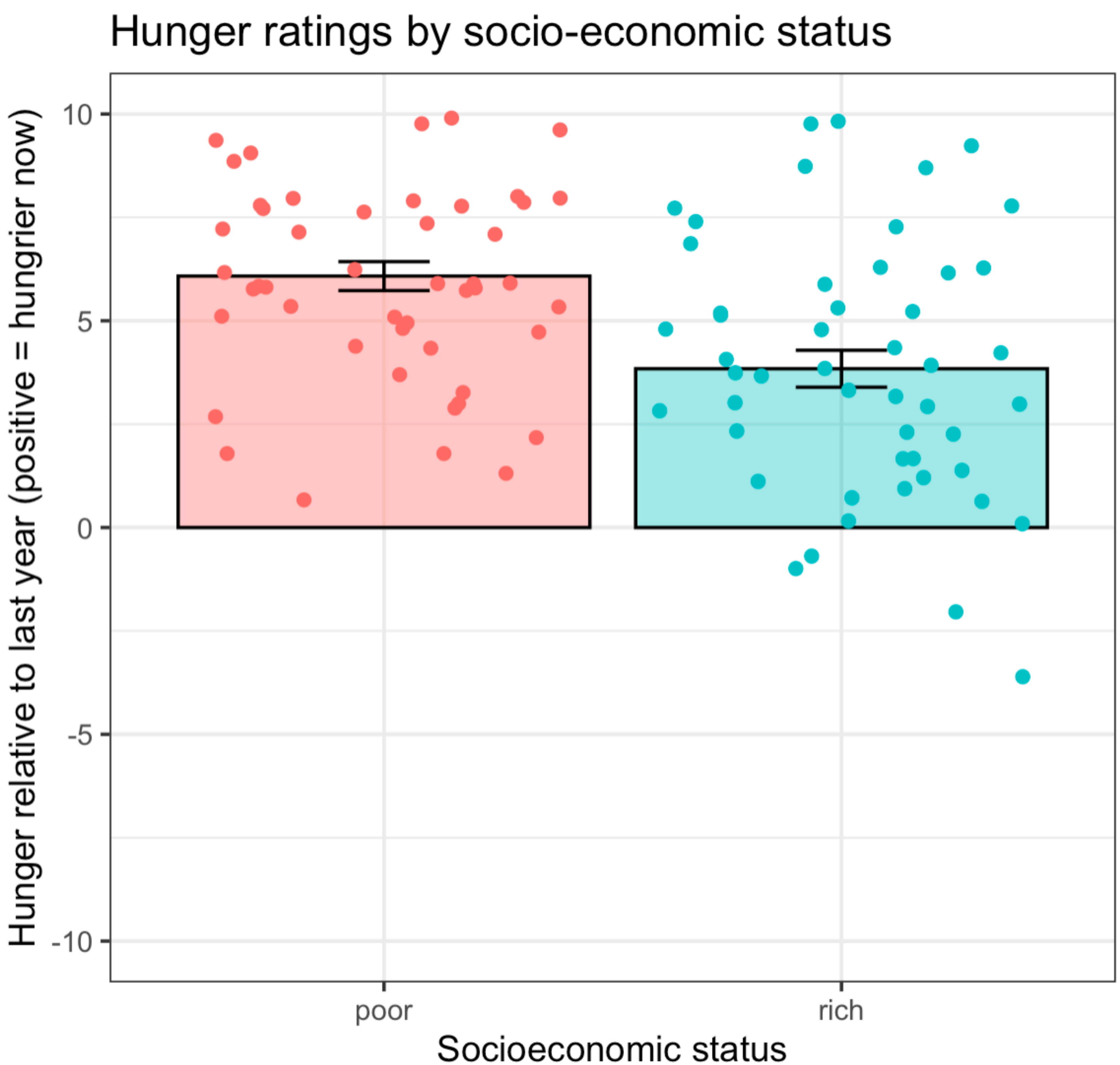
Our data

It has **two groups:** poor and rich. We'll want to compare them

Are poor people hungrier (relative to last year) than rich people?

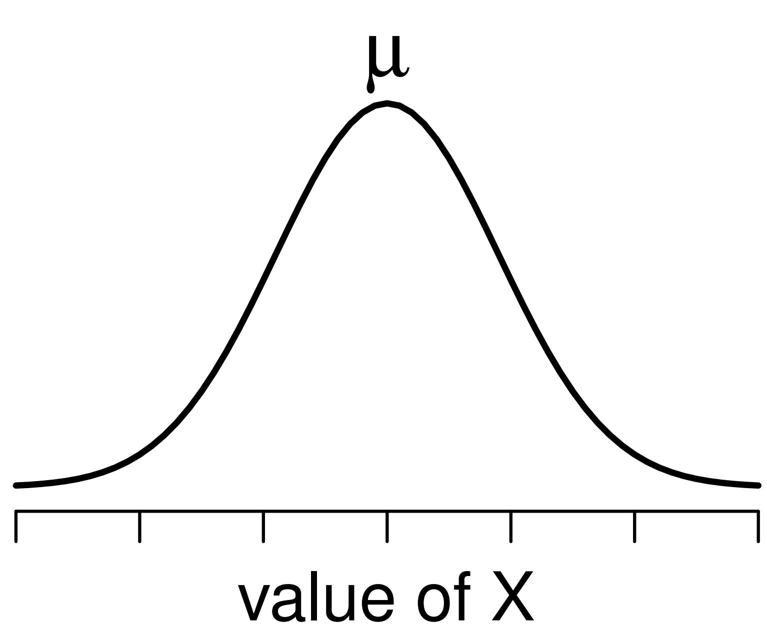
```
> de <- read_csv(file=here(“economy.csv"))
> head(de)

# A tibble: 6 x 3
  person   ses   rating
  <chr>    <fct>  <dbl>
1 person 1 rich       9
2 person 2 rich       2
3 person 3 rich       0
4 person 4 poor       2
5 person 5 rich       3
6 person 6 poor      10
```

Always look at the data first!



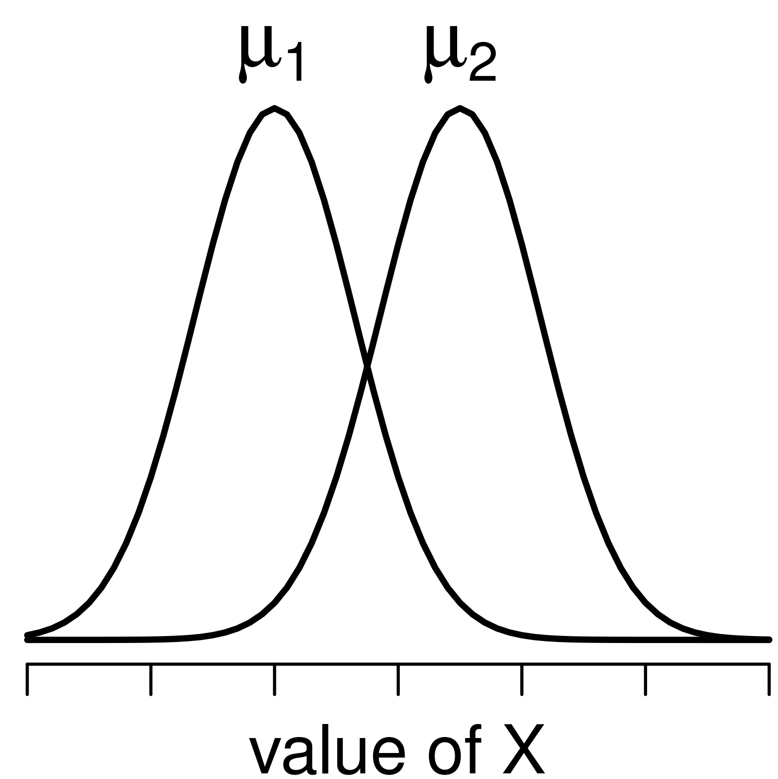
Null hypothesis, H0

population means are
equal for both
SES "groups"



Alternative hypothesis, H1

population means are
different for the two groups
(i.e. rich and poor)

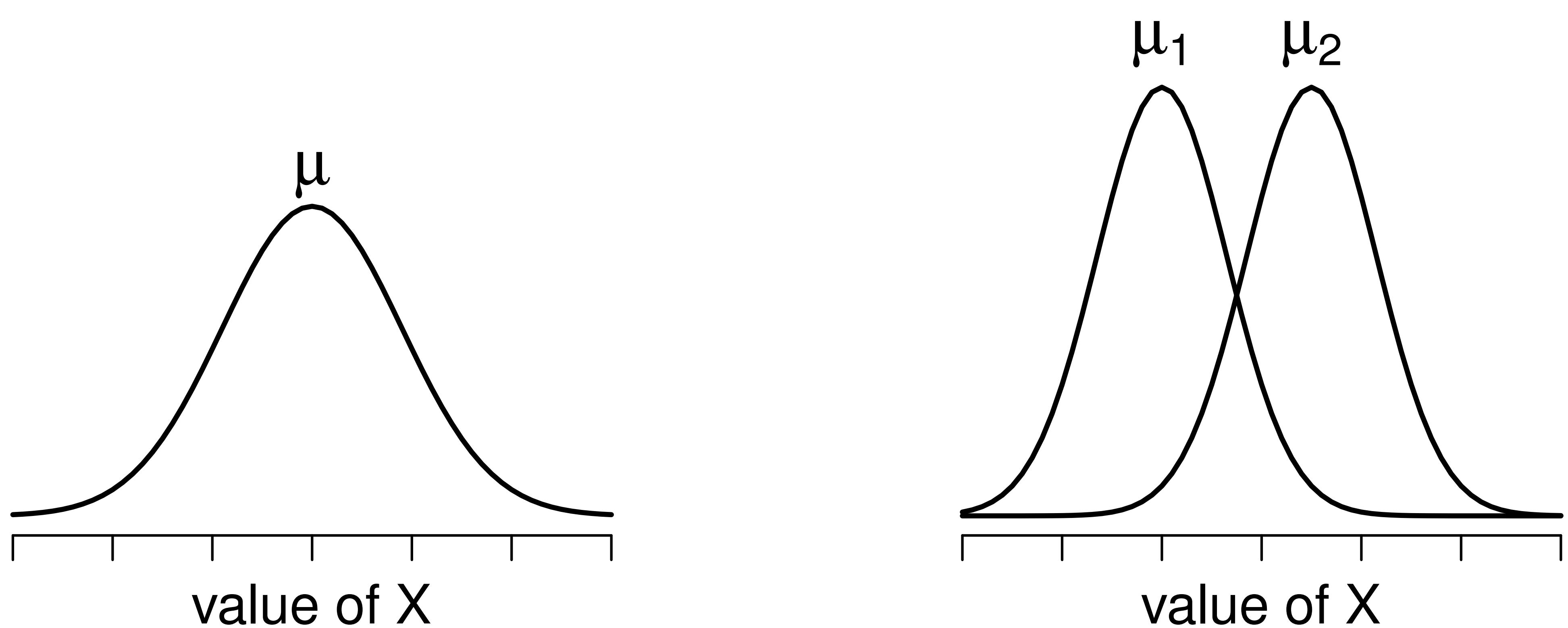


Background assumptions

population distributions are normal

observations are independently sampled

groups have the same standard deviation
(a.k.a. “homogeneity of variance”, “homoscedasticity”)



Let's construct a test

1) A diagnostic test statistic, T
2) Sampling distribution of T if the null is true
3) The observed T in your data
4) A rule that maps every value of T onto a decision (accept or reject H0)

Diagnostic test statistic?

As before, our t-statistic is based on the difference between group means

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

We “standardise” the difference by dividing it by the **standard error** of the (difference between) sample means

This is our estimate of the the population standard deviation (which we <u>assume</u> is the same for both groups)

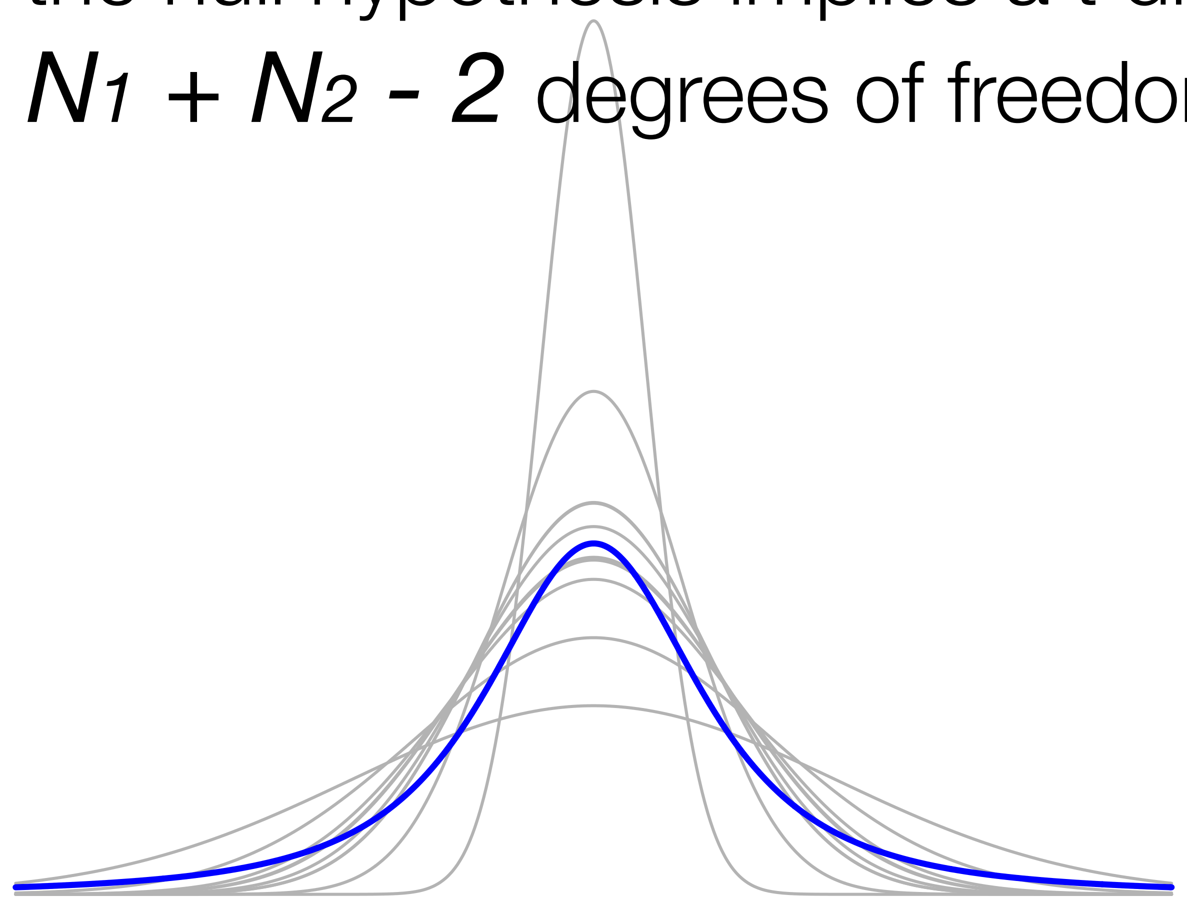
This part tells us how much the sampling distribution “narrows” with sample size (it depends on N for both groups!)

Sampling distribution of t

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

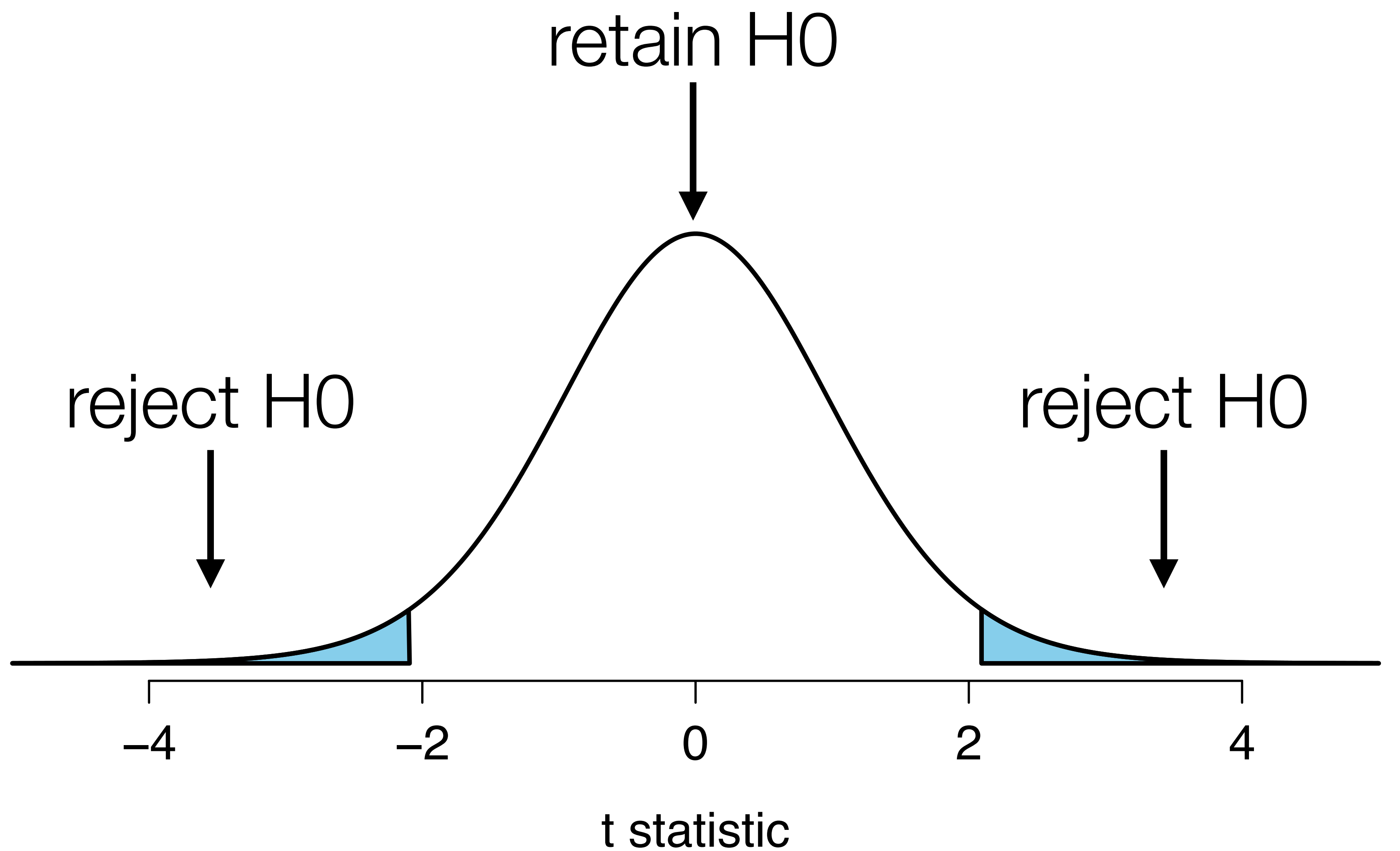
As usual… if the null hypothesis is true we expect t to be close to zero. If the alternative is true, it's more likely to differ substantially from zero

Specifically, the null hypothesis implies a t distribution with $N_1 + N_2 - 2$ degrees of freedom

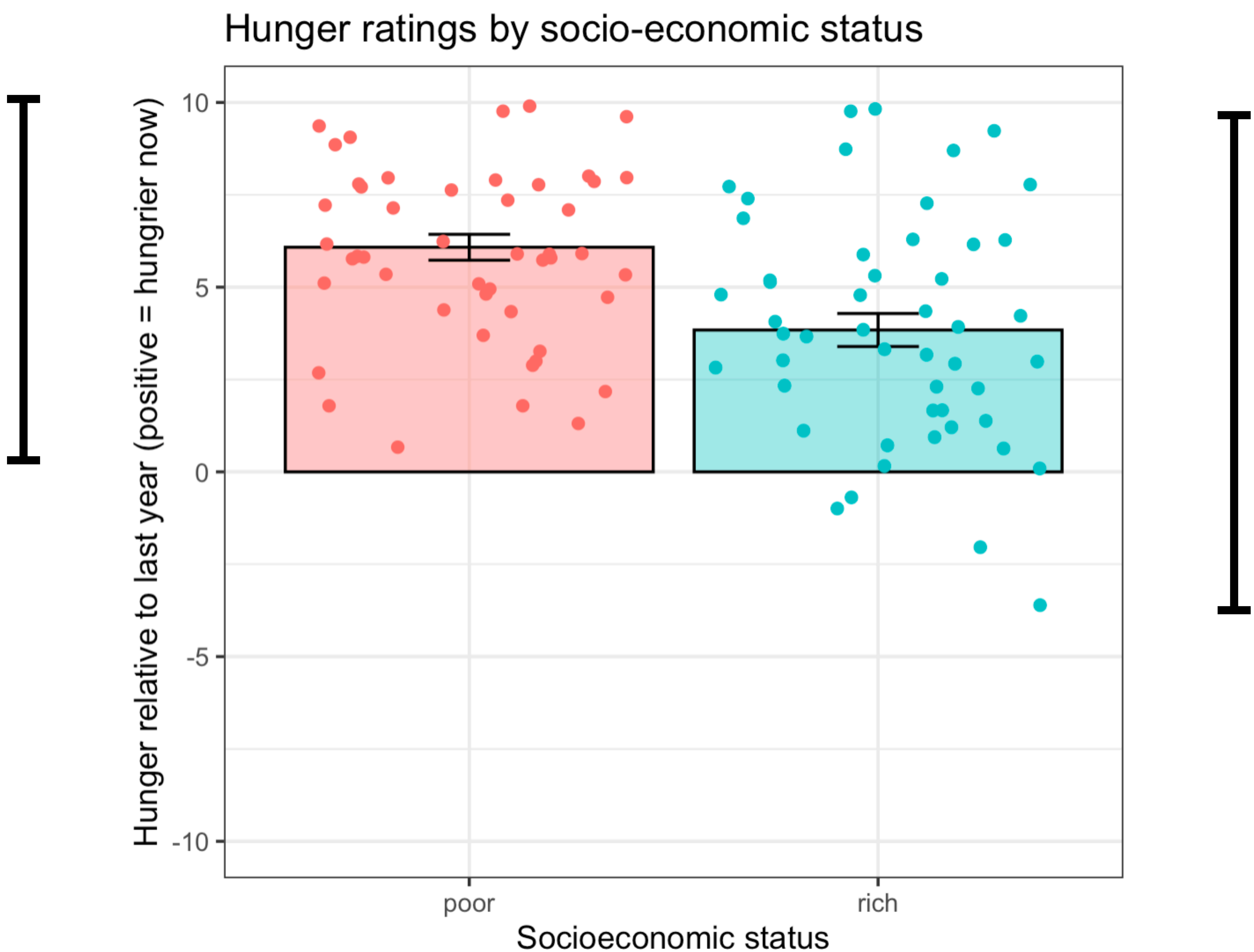


Rule for making a decision

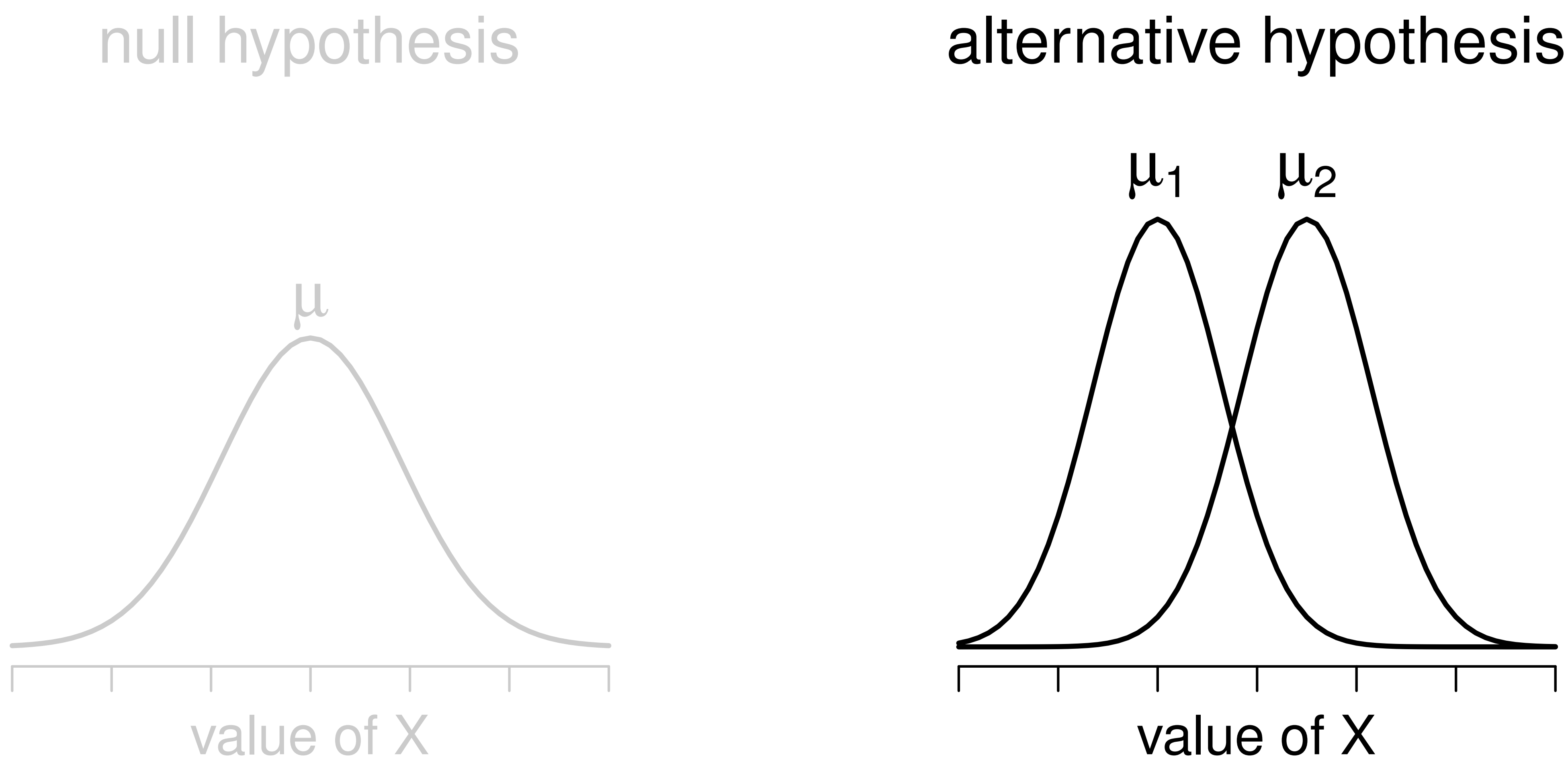
Same as we've seen before. For a two-sided test:



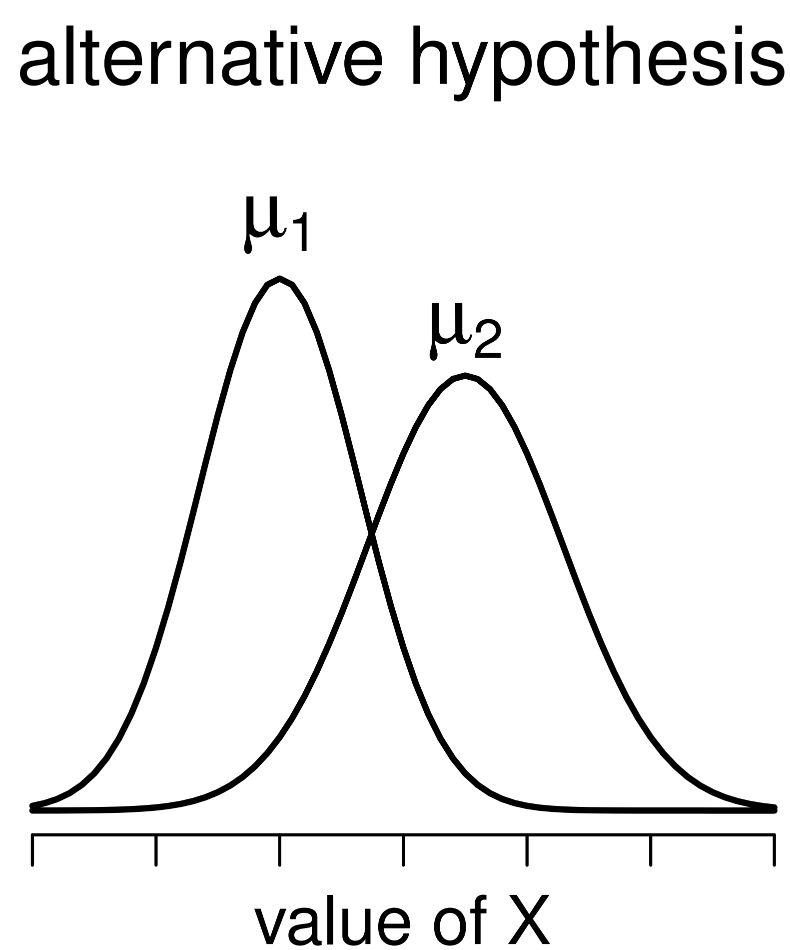
But… it looks like the rich are a lot more variable



The Student t-test (which we just saw) assumes both groups have equal variance…



Fortunately, the <u>Welch</u> t-test does not…



Big picture: what's the difference?

- Both tests assume
  - Population is normal
  - Observations are independently sampled

- Student:
  - Populations have equal variances

- Welch:
  - Populations can have different variances
  - This is a <u>much</u> safer thing to do

Under the hood, what does that mean?

- Student:
  - Degrees of freedom equals $N_1 + N_2 - 2$
  - Estimates a single “pooled” standard deviation

- Welch:
  - “Effective” degrees of freedom is adjusted to be a bit smaller, depending on just how unequal the sample standard deviations are (not necessarily a whole number)
  - Uses both groups’ standard deviations separately

Which one should I use?

- Welch test is usually better. It is the default in R
- But Student's test is somewhat better known. It’s the default in SPSS
    - Many people don’t know Welch
    - Many times when people say they ran an independent samples t-test they mean a Student test.

- **Use Welch unless told otherwise** for this class!
- It’s actually easier for you!

Welch t-test in R

If your data is in long form:

Predictor	Outcome
GroupA	8.3
GroupB	2.9
GroupA	7.5
GroupB	4.0
GroupA	6.2
GroupB	5.1

```
t.test(outcome ~ predictor, data=d)
```

Our data is in long form:

ses	rating
rich	9
rich	2
rich	0
poor	2
rich	3
poor	10

```
t.test(rating ~ ses, data=de)
```

Welch t-test in R

```
> t.test (formula = rating ~ ses, data = de)
```

This is the dataset

This is the formula that describes our test: we want to see if rating (the outcome variable) is related to ses (the predictor/grouping variable)

In general in R you can use formulas to specify all kinds of relationships:

```
outcome ~ predictor
```

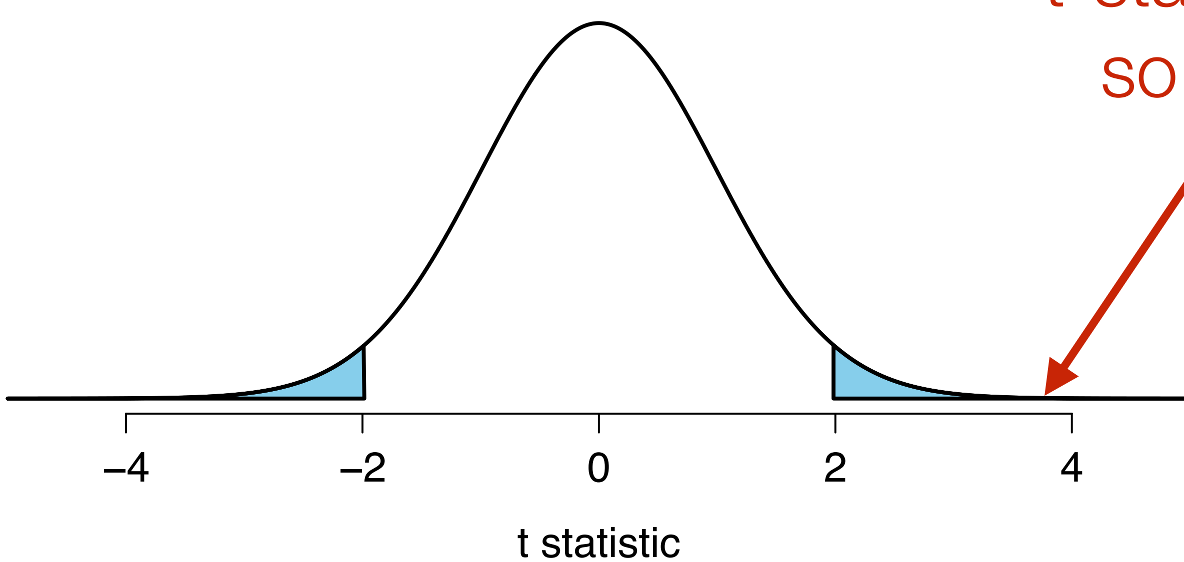
Welch t-test in R

```
> t.test (formula = rating ~ ses, data = de)

    Welch Two Sample t-test

data:  rating by ses
t = 3.9487, df = 92.672, p-value = 0.000153
```

Default is Welch

Test statistic, degrees of freedom, p-value

t-statistic falls here, so we reject H0



Welch t-test in R

```
> t.test (formula = rating ~ ses, data = de)

    Welch Two Sample t-test

data:  rating by ses
t = 3.9487, df = 92.672, p-value = 0.000153
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.11346 3.36654
sample estimates:
mean in group poor mean in group rich
              6.08               3.84
```

Default is Welch

Test statistic, degrees of freedom, p-value

H1 is about the difference in means

confidence interval for the *difference* between means

means of each group

Student t-test in R

You can modify the `t.test()` command to get a Student t-test

All you need to do is tell R to *assume* that the variances are equal

```
> t.test (formula = rating ~ ses, data = de, var.equal=TRUE)

	Two Sample t-test

data:  rating by ses
t = 3.9487, df = 98, p-value = 0.000148
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.11427 3.36573
sample estimates:
mean in group poor mean in group rich
              6.08               3.84
```

Output is similar except for the name of the test and the exact numbers

Can do all the same things in wide form

If your data is in wide form:

GroupA	GroupB
8.3	NA
NA	2.9
7.5	NA
NA	4.0
6.2	NA
NA	5.1

```
t.test(d$groupA, d$groupB)
```

We can put our data in wide form

rich	poor
9	NA
2	NA
0	NA
NA	2
3	NA
NA	10

```
t.test(de_wide$rich, de_wide$poor)
```

Putting our data in wide form

```
> de_wide <- de %>%
      pivot_wider(names_from=ses,values_from=rating)

> head(de_wide)

# A tibble: 6 x 3
  person    rich  poor
  <chr>    <dbl> <dbl>
1 person 1     9    NA
2 person 2     2    NA
3 person 3     0    NA
4 person 4    NA     2
5 person 5     3    NA
6 person 6    NA    10
```

Welch t-test

```
> t.test( x=de_wide$rich, y=de_wide$poor)

	Welch Two Sample t-test

data:  de_wide$rich and de_wide$poor
t = -3.9487, df = 92.672, p-value = 0.000153
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.36654 -1.11346
sample estimates:
mean of x mean of y 
     3.84      6.08 
```

Student t-test

```
> t.test( x=de_wide$rich, y=de_wide$poor, var.equal=TRUE)

   Two Sample t-test

data:  de_wide$rich and de_wide$poor
t = -3.9487, df = 98, p-value = 0.000148
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.36573 -1.11427
sample estimates:
mean of x mean of y
     3.84      6.08
```

Exercises are in w7day2exercises.Rmd