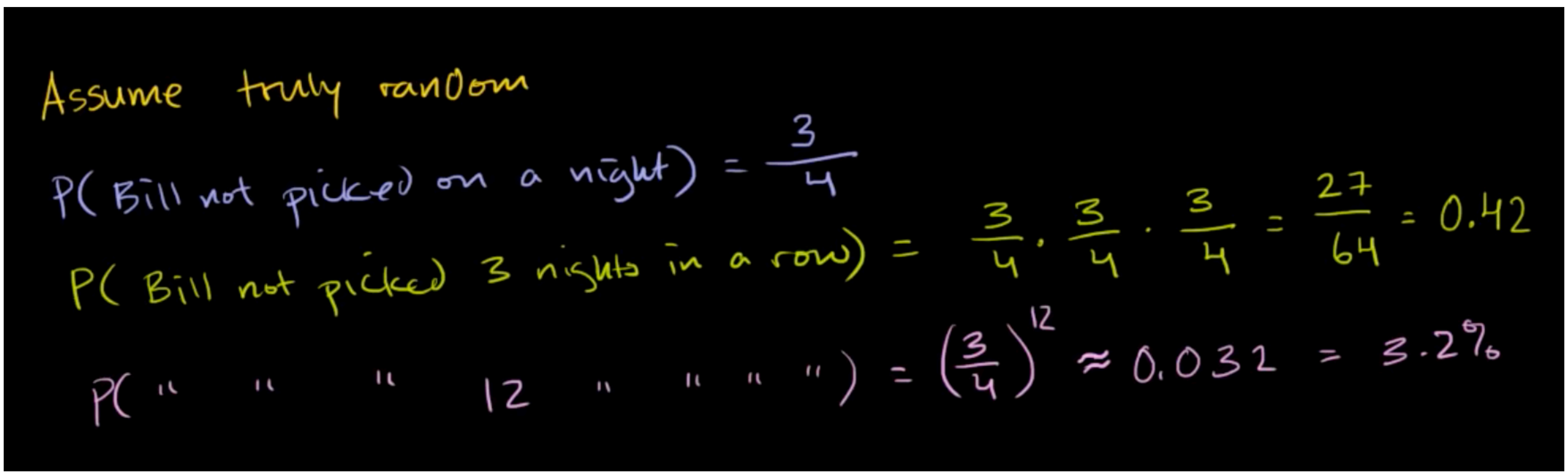
SIMPLE TESTS

Tutorial #4

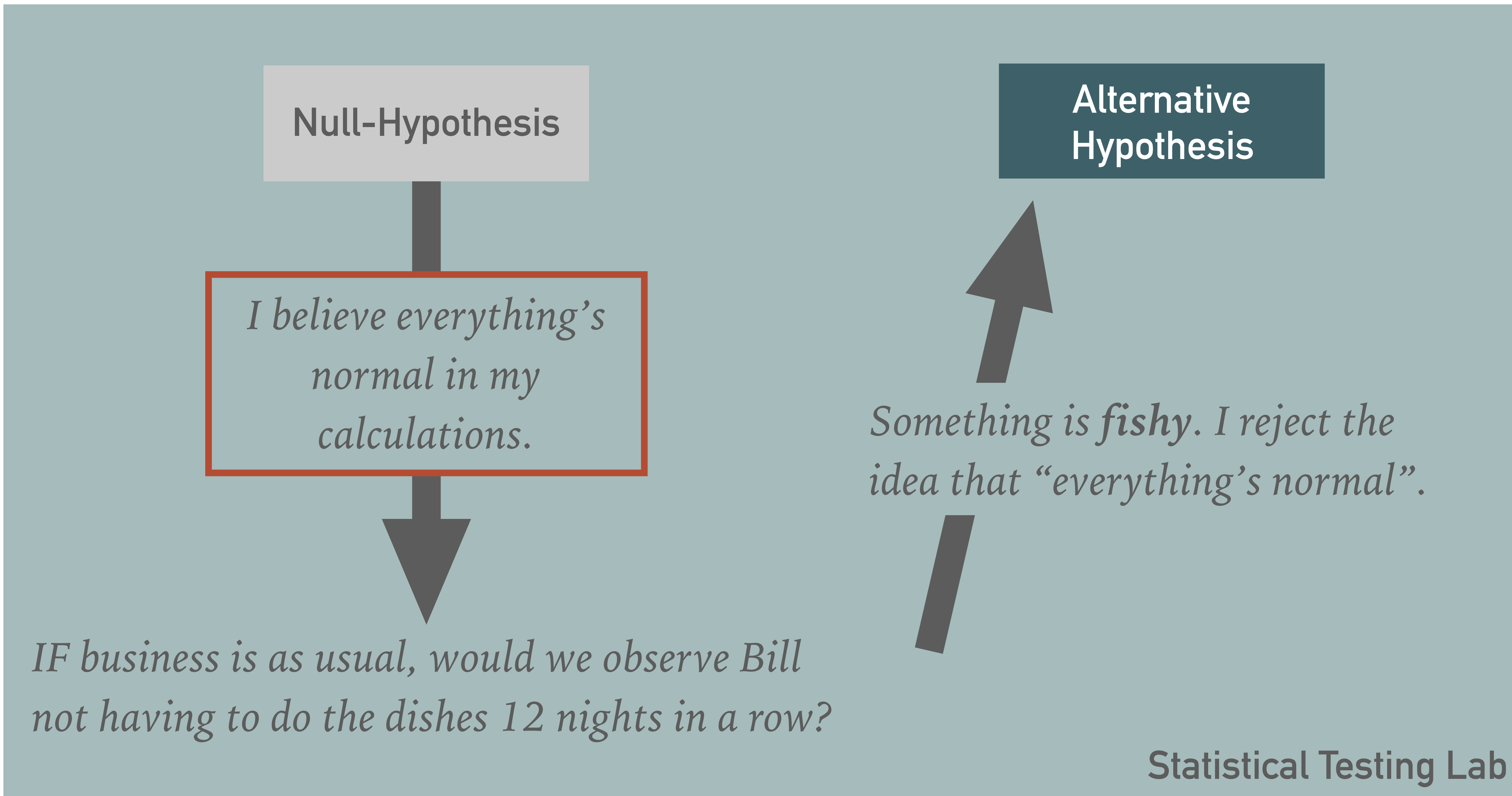
INTRODUCTION

- Video by Sal Khan on Significance

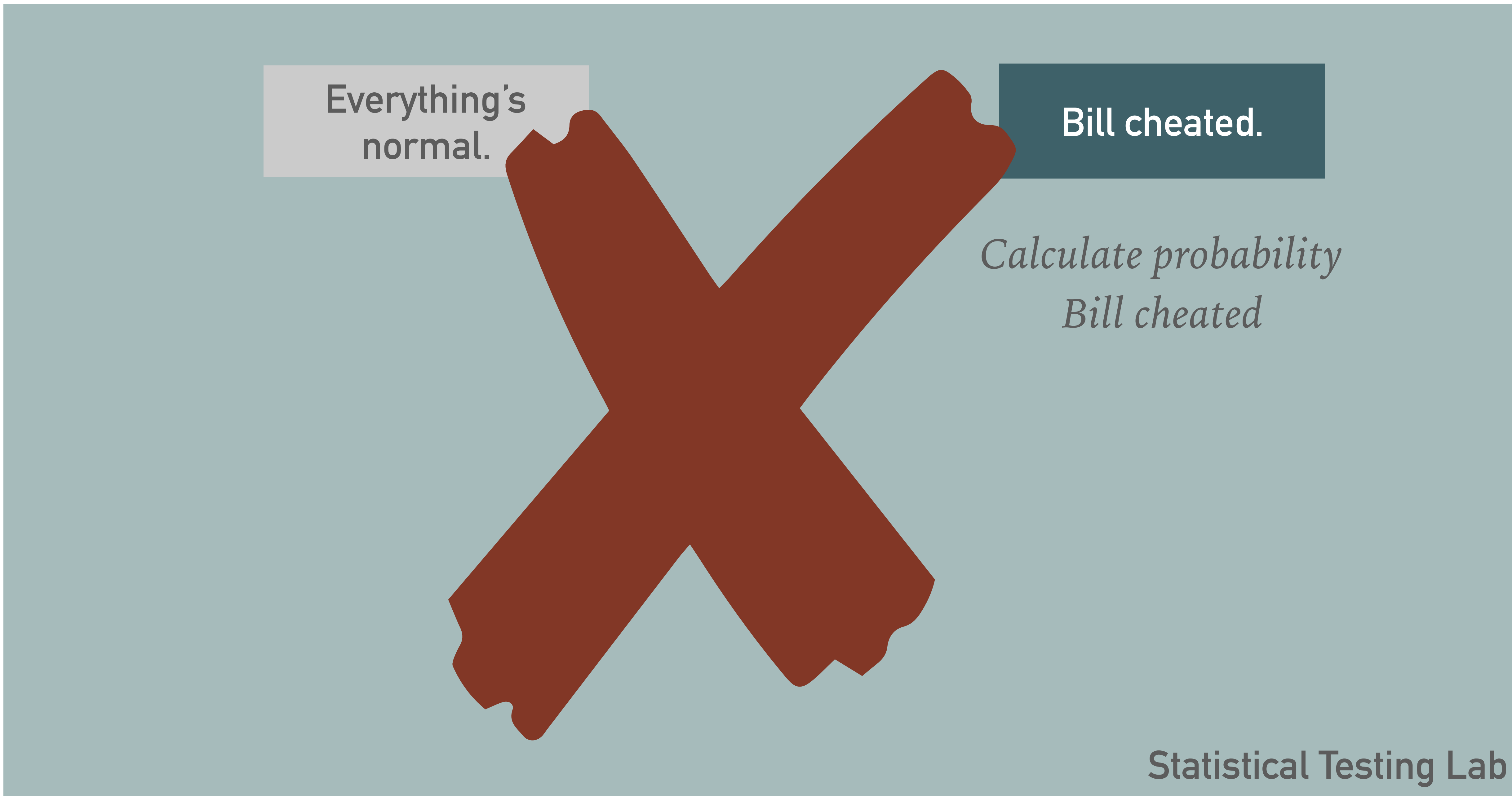


STATISTICAL TESTING: THE FALSIFICATION PRINCIPLE

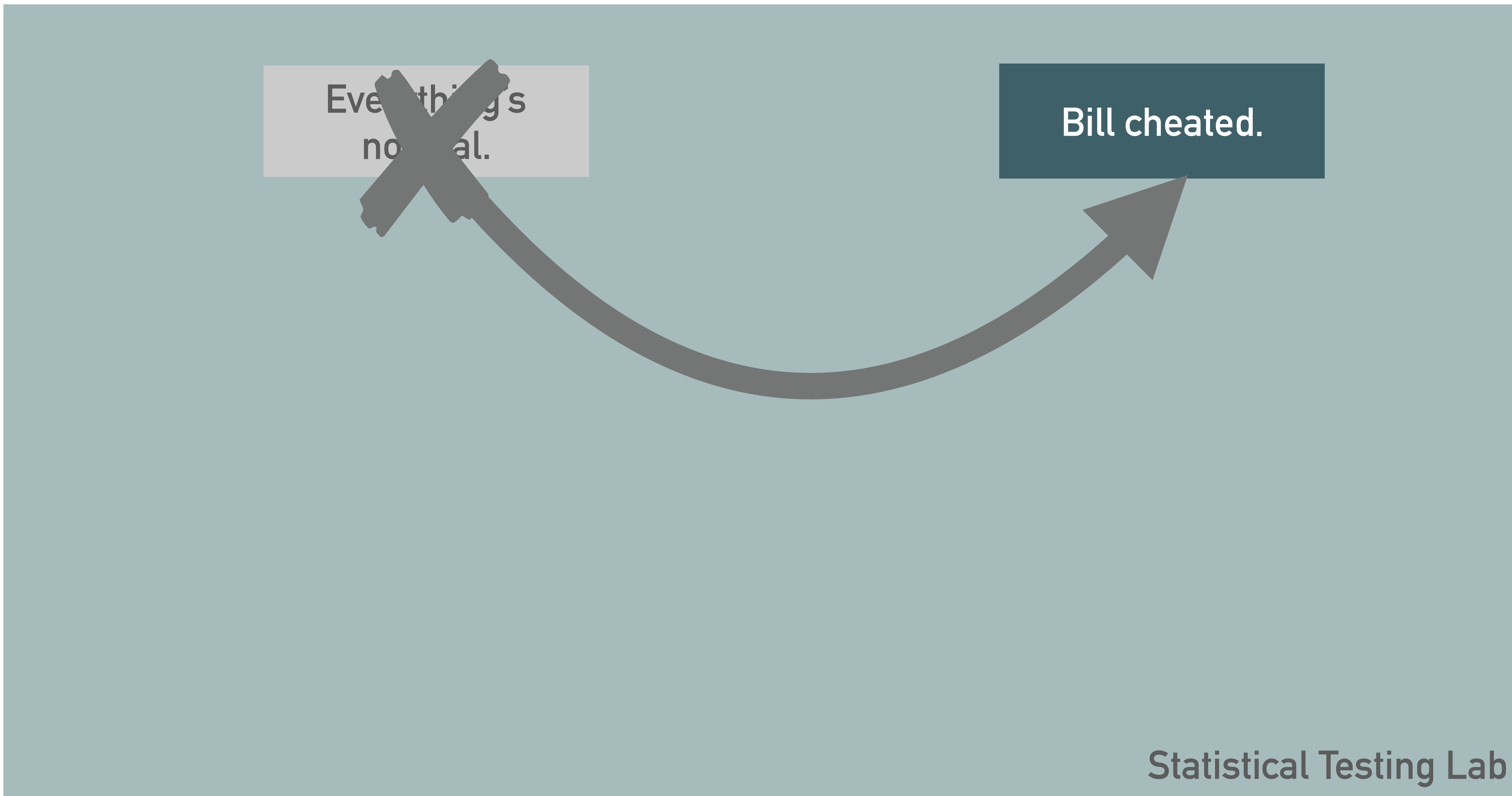
Falsification principle*: I cannot "prove" things, I can only falsify them.*



WHAT WE DON'T DO



WHAT WE DO



STATISTICAL TESTING: MORE GENERAL

p-value

- “trust” in our Null-Hypothesis
- p(the data|Null-Hypothesis) < 5%
- the probability, calculated under null hypothesis, of observing a result as extreme as, or more extreme than, the one that was obtained/observed

Statistical Testing Lab

STATISTICAL TESTING: MATHEMATICALLY

probability *of* *observing data like we did* *given* *everything is normal*

p(the data | Null-Hypothesis) < 5%

➙ I found evidence against the Null-Hypothesis if this probability is below 5%.

We call this the **p-value**.

EXERCISE

- What's the *Null* and the *Alternative Hypothesis* in the following cases?

    1. You know the coffee drinkers' grades in this study program and you'd like to know if that's normally distributed.

    2. You have coffee drinkers and non-coffee drinkers and you'd like to know if their grades have a similar variance, ergo if it's even comparable.

    3. You have coffee drinkers and non-coffee drinkers and you'd like to know if the coffee drinking is related to being late or early. (late/early = categorical)

    4. You have coffee drinkers and non-coffee drinkers and you'd like to know if the coffee drinkers usually come to a different time than the non-coffee drinkers. (late/early in minutes)

    5. You have coffee-drinkers and you'd like to know if on days where they drink coffee, they have better grades.

You have 3 minutes.

OVERVIEW OF TESTS

Test	Data type	Purpose	Null Hypothesis	Alternative Hypothesis
Shapiro-Wilk test	(Ordinal)/ Interval/Ratio	**Normally** distributed?	Data is normally distributed.	Data is not normally distributed.
F-Test				
Chi-Square Test				
t-test				
Wilcoxon Test				

SHAPIRO WILK TEST

Is my data normally distributed?

LET'S TAKE THE COFFEE DRINKERS

Shapiro Wilk Test

Null-Hypothesis	➤ Data is normally distributed.
Alternative Hypothesis	➤ Data is not normally distributed.

Scenario: 1. You know the coffee drinkers' grades in this study program and you'd like to know if that's normally distributed.

Null-Hypothesis	➤ The coffee drinkers' grades in this study program are normally distributed (= I can proceed with other tests).
Alternative Hypothesis	➤ The coffee drinkers' grades in this study program are *not* normally distributed (= I have to find the right distribution).

THAT'S WHAT OUR DATA LOOKS LIKE

Scenario: 1. You know the coffee drinkers' grades in this study program and you'd like to know if that's normally distributed.

coffee drinkers
1.3
1.7
3.7
2.7
3.7
3.7
2.3
2.0
1.3
1.7
…

THAT'S WHAT R GIVES ME

Shapiro Wilk Test

```
> shapiro.test(sample1)

	Shapiro-Wilk normality test

data:  sample1
W = 0.98824, p-value = 0.7587
```

p(the data | H0) > 0.05

!!

H0: coffee drinkers' grades normally distributed (= I can proceed with other tests).

H1: coffee drinkers' grades not normally distributed (= find distribution).

Note: This is actually what we want.

OVERVIEW OF TESTS

Test	Data type	Purpose	Null Hypothesis	Alternative Hypothesis
Shapiro-Wilk test	(Ordinal)/ Interval/Ratio	**Normally** distributed?	Data is normally distributed.	Data is not normally distributed.
F-Test	(Ordinal)/ Interval/Ratio	2 **variances** the same?	No difference in variances.	There is a difference in variances.
Chi-Square Test				
t-test				
Wilcoxon Test				

F-TEST

Are the variances of two samples the same?

LET'S TAKE THE COFFEE DRINKERS

F-Test

Null-Hypothesis

- ➤ No difference in variances.

Alternative Hypothesis

- ➤ Difference in variances.

Scenario: 2. You have coffee drinkers and non-coffee drinkers and you'd like to know if their grades have a similar variance, ergo if it's even comparable.

Null-Hypothesis

- ➤ Coffee drinkers and non-coffee drinkers have a similar variance (= comparable).

Alternative Hypothesis

- ➤ Coffee drinkers and non-coffee drinkers have different variances (= *not* comparable).

THAT'S WHAT OUR DATA LOOKS LIKE

Scenario: 2. You have coffee drinkers and non-coffee drinkers and you'd like to know if their grades have a similar variance, ergo if it's even comparable.

coffee drinkers	non-coffee drinkers
1.3	2.3
1.7	2.7
3.7	2.0
2.7	2.0
3.7	1.7
3.7	1.3
2.3	3.0
2.0	3.0
1.3	1.3
1.7	3.7
…	…

THAT'S WHAT R GIVES ME

F-Test

```
> var.test(sample1, sample2)

	F test to compare two variances

data:  sample1 and sample2
F = 1.2172, num df = 49, denom df = 29, p-value = 0.5785
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6115387 2.2900198
sample estimates:
ratio of variances
          1.217179
```

p(the data | H0) > 0.05

!!

H0: Similar variances (= comparable)

H1: Different variances (= not comparable)

Note: This is actually what we want.

OVERVIEW OF TESTS

Test	Data type	Purpose	Null Hypothesis	Alternative Hypothesis
Shapiro-Wilk test	(Ordinal)/ Interval/Ratio	**Normally** distributed?	Data is normally distributed.	Data is not normally distributed.
F-Test	(Ordinal)/ Interval/Ratio	2 **variances** the same?	No difference in variances.	There is a difference in variances.
Chi-Square Test	Categorical	# Categories **relationship**?	The variables are all similarly distributed, no relationship.	The variables depend upon each other, there is a relationship.
t-test				
Wilcoxon Test				

CHI-SQUARE TEST

Do my categories depend on each other?

LET'S TAKE THE COFFEE DRINKERS

Chi-Square Test

Null-Hypothesis

- The variables are all similarly distributed, no relationship.

Alternative Hypothesis

- The variables depend upon each other, there is a relationship.

Scenario: 3. You have coffee drinkers and non-coffee drinkers and you'd like to know if the coffee drinking is related to being late or early.

Null-Hypothesis

- Coffee and non-coffee drinkers both come early or late, no relationship.

Alternative Hypothesis

- Coffee drinkers tend to come to a different time than non-coffee drinkers.

THAT'S WHAT OUR DATA LOOKS LIKE

Scenario: 3. You have coffee drinkers and non-coffee drinkers and you'd like to know if the coffee drinking is related to being late or early.

	come early	come late
coffee drinkers	40	60
non-coffee drinkers	30	20

THAT'S WHAT R GIVES ME

Chi-Square Test

our data

chi-square test

```
> coffeeEarly
            time
coffee       early late
  coffee        40   60
  non-coffee    30   20
> chisq.test(coffeeEarly)

	Pearson's Chi-squared test with Yates' continuity correction

data:  coffeeEarly
X-squared = 4.5837, df = 1, p-value = 0.03228
```

$\boxed{p(\text{the data} \mid H0) < 0.05}$

H0: No relationship between coffee and coming early/late

H1: I believe coffee has something to do with coming early/ late.

OVERVIEW OF TESTS

Test	Data type	Purpose	Null Hypothesis	Alternative Hypothesis
Shapiro-Wilk test	(Ordinal)/ Interval/Ratio	**Normally** distributed?	Data is normally distributed.	Data is not normally distributed.
F-Test	(Ordinal)/ Interval/Ratio	2 **variances** the same?	No difference in variances.	There is a difference in variances.
Chi-Square Test	Categorical	# Categories **relationship**?	The variables are all similarly distributed, no relationship.	The variables depend upon each other, there is a relationship.
t-test	Interval/Ratio	2 samples significantly **different**?	The two samples means are not different.	The two samples means are different.
Wilcoxon Test				

T-TEST

Do the two samples have different means?

LET'S TAKE THE COFFEE DRINKERS

T-Test

Null-Hypothesis	➤ The two samples' means are not different.
Alternative Hypothesis	➤ The two samples means are different.

Scenario: 4. You have coffee drinkers and non-coffee drinkers and you'd like to know if the coffee drinkers usually come to a different time than the non-coffee drinkers.

Null-Hypothesis	➤ Coffee and non-coffee drinkers don't come at different times.
Alternative Hypothesis	➤ Coffee and non-coffee drinkers come at different times.

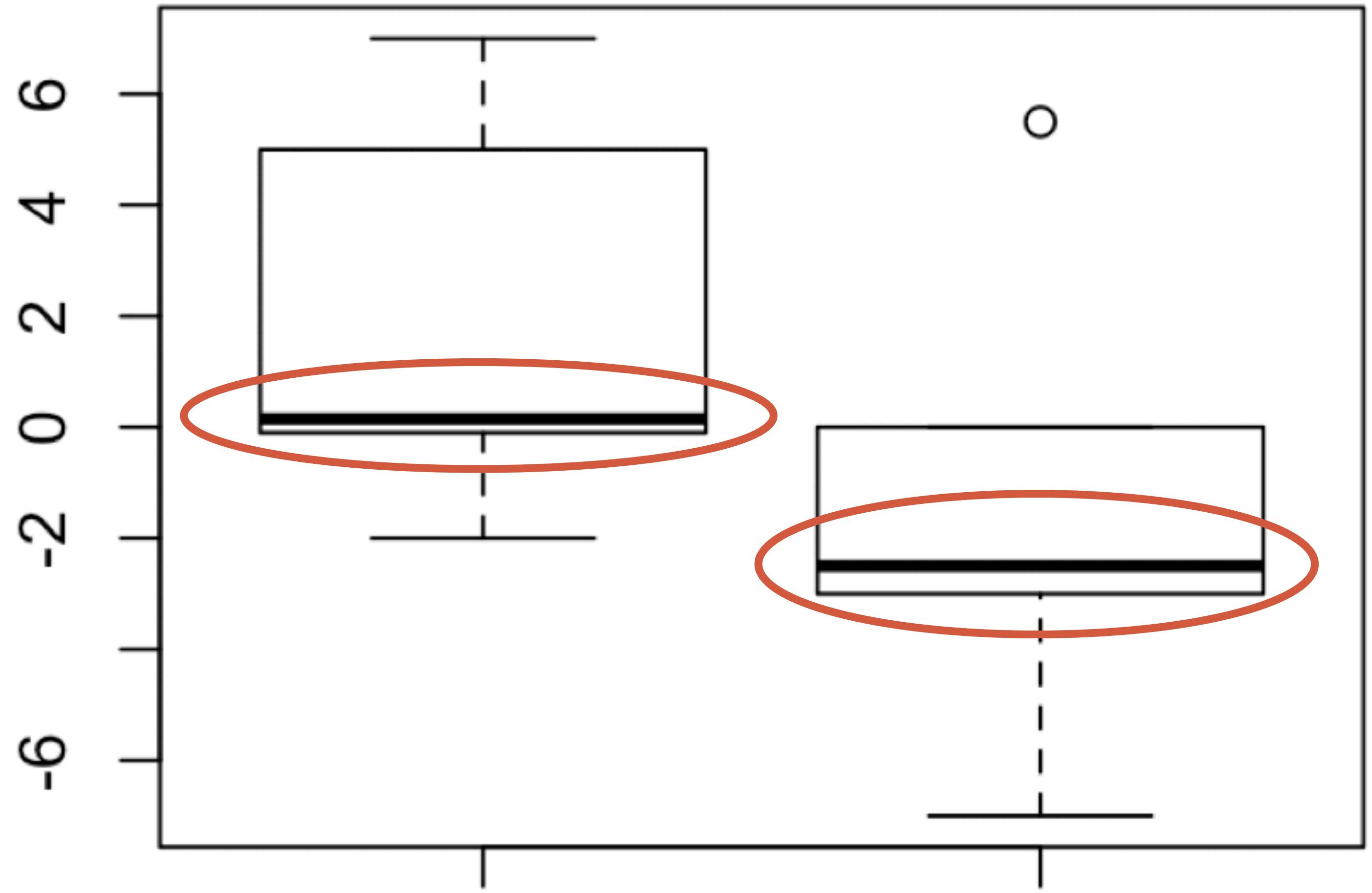
THAT'S WHAT OUR DATA LOOKS LIKE

Scenario: 4. You have coffee drinkers and non-coffee drinkers and you'd like to know if the coffee drinkers usually come to a different time than the non-coffee drinkers (in min).

coffee drinkers' arrival	non-coffee drinkers' arrival
0.3	-3.0
5.0	-2.5
6.8	0.0
7.0	5.5
-0.1	-0.1
-0.5	-2.5
0.5	-3.0
0.0	-7.0
0.0	-4.5
-2.0	0.0
…	…

THE T-TEST

Arrival compared to the lecture starting time (in min)



coffee drinkers

(“sample”)

non-coffee drinkers

(“normal population”)

*We compare the **sample** to our **population**.*

THAT'S WHAT R GIVES ME

T-Test

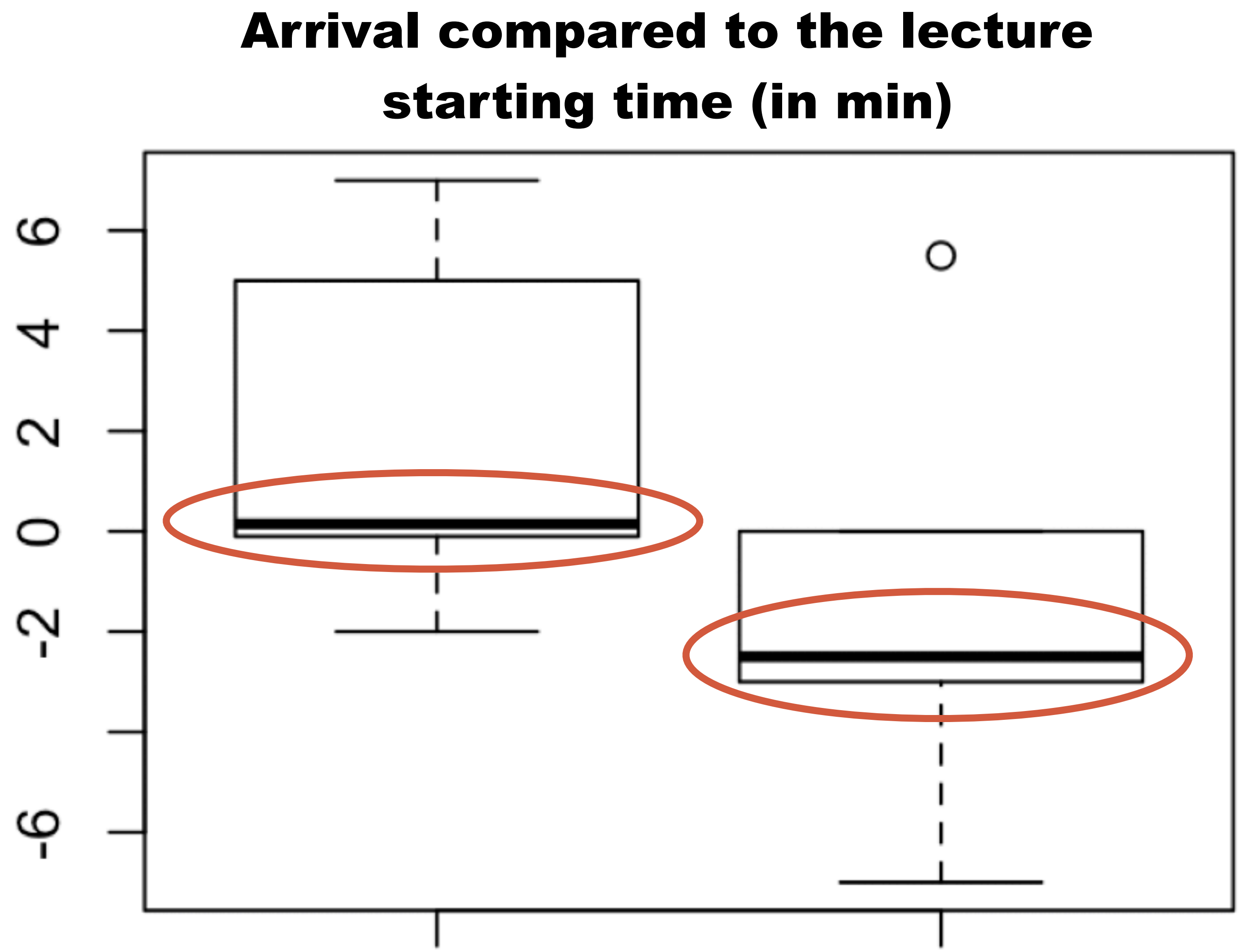
```
> t.test(sample1, sample2)

	Welch Two Sample t-test

data:  sample1 and sample2
t = 2.3066, df = 17.989, p-value = 0.03319
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3038705 6.5161295
sample estimates:
mean of x mean of y 
     1.70     -1.71 
```

$p(\text{the data} \mid H0) < 0.05$

H0: Coffee and non-coffee drinkers don't come at different times.

H1: Coffee and non-coffee drinkers come at different times.

THE T-TEST



coffee drinkers

("sample")

non-coffee drinkers

("normal population")

*We compare the **sample** to our **population**.*

THE T-TEST FORMULA

- $\bar{x}$ = **sample** mean
- μ_0 = **population** mean
- s = standard deviation of the **sample**
- n = sample size

$$t = \frac{\text{sample - population}}{\text{standardized by something}}$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

*We compare the **sample** to our **population**.*

LET'S DO THIS BY HAND…

- … to appreciate our beloved R once more ;)
- Calculate the variance and the t-value.

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

You have 10 minutes.

coffee drinkers' arrival	non-coffee drinkers' arrival
0.3	-3.0
5.0	-2.5
6.8	0.0
7.0	5.5
-0.1	-0.1
-0.5	-2.5
0.5	-3.0
0.0	-7.0
0.0	-4.5
-2.0	0.0
…	…

A NOTE ON THE FORMULA

	General case	Our case "two samples" case A = sample 1, B = sample 2
standard deviation	$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$	$s = \sqrt{\frac{\sum (x - \bar{x}_A)^2 + \boxed{\sum (x - \bar{x}_B)^2}}{n_A + \boxed{n_B} - 2}}$
t-value	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \frac{\bar{x}_A - \bar{x}_B}{\frac{s}{\sqrt{n_A}} + \boxed{\frac{s}{\sqrt{n_B}}}}$

c	nc
0.3	-3.0
5.0	-2.5
6.8	0.0
7.0	5.5
-0.1	-0.1
-0.5	-2.5
0.5	-3.0
0.0	-7.0
0.0	-4.5
-2.0	0.0
…	…

BLACKBOARD: HOW TO WRITE THIS DOWN

Data types: ratio / ratio

Test: t-test

~~Null~~ Hypo: → Alt Hypo:

oder Null Hypo — Alt Hypo:

Varianz: Formel — Berechnung

t-value: Formel — Berechnung

p-value:

} you may use R

⇒ evidence for rej. H_0

FROM T-VALUES TO P-VALUES

T-Test

Two Tails T Distribution Table

the p-value you'll get

DF	A = 0.2	0.10	0.05	0.02	0.01	0.002	0.001
∞	t_a = 1.282	1.645	1.960	2.326	2.576	3.091	3.291
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101 *x*	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883

df ≈ n-2

$t = 2.3006$

$p \approx 0.03$

https://www.statisticshowto.datasciencecentral.com/tables/t-distribution-table/

THAT'S WHAT R GIVES ME


```
> t.test(sample1, sample2)

        Welch Two Sample t-test

data:  sample1 and sample2
t = 2.3066, df = 17.989, p-value = 0.03319
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3038705 6.5161295
sample estimates:
mean of x mean of y
     1.70     -1.71
```

$p(\text{the data} \mid H0) < 0.05$

H0: Coffee and non-coffee drinkers don't come at different times.

H1: Coffee and non-coffee drinkers come at different times.

TASK: ONE MORE TIME!

- Data which show the effect of two soporific drugs (increase in hours of sleep compared to control) on 10 patients.

- Help for typing in R:
- `data(sleep)`
- `t.test(extra ~ group, data = sleep, paired = TRUE)`

OVERVIEW OF TESTS

Test	Data type	Purpose	Null Hypothesis	Alternative Hypothesis
Shapiro-Wilk test	(Ordinal)/ Interval/Ratio	**Normally** distributed?	Data is normally distributed.	Data is not normally distributed.
F-Test	(Ordinal)/ Interval/Ratio	2 **variances** the same?	No difference in variances.	There is a difference in variances.
Chi-Square Test	Categorical	# Categories **relationship**?	The variables are all similarly distributed, no relationship.	The variables depend upon each other, there is a relationship.
t-test	Interval/Ratio	2 samples significantly **different**?	The two samples means are not different.	The two samples means are different.
Wilcoxon Test	Ordinal	2 samples significantly **different**?	The two samples means are not different.	The two samples means are different.

WILCOXON TEST

Ordinal alternative to the dependent samples t-test.

LET'S TAKE THE COFFEE DRINKERS

Wilcoxon Test

Null-Hypothesis

- The samples' means are not different.

Alternative Hypothesis

- The samples means are different.
- *Important: they need to be **dependent** samples.*

Scenario: 5. You have coffee-drinkers and you'd like to know if on days where they drink coffee, they have better grades.

"dependent sample": same people with different conditions

Null-Hypothesis

- Coffee drinkers don't have different grades depending on their coffee input.

Alternative Hypothesis

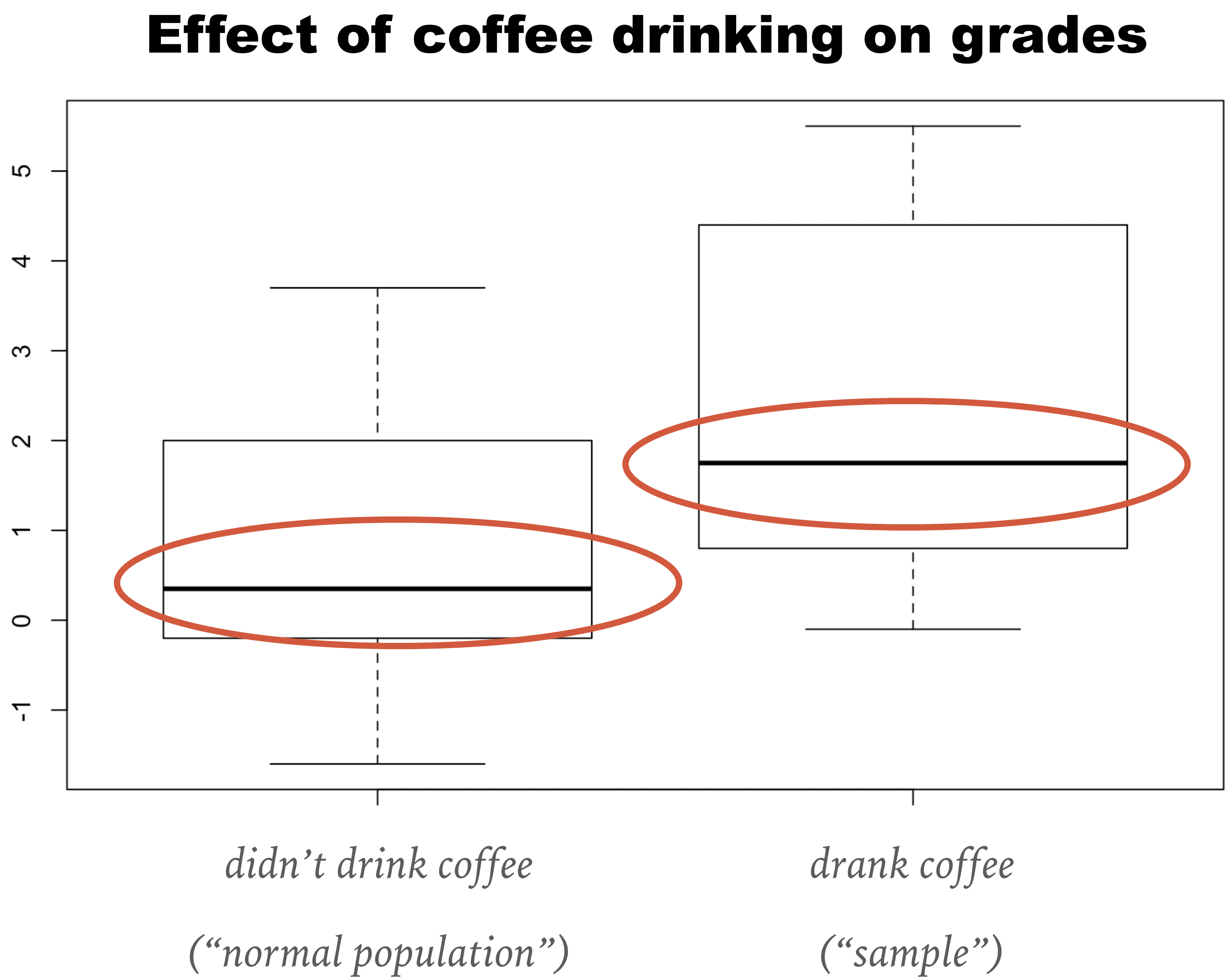
- Coffee drinkers have different grades depending on their coffee input.

THAT'S WHAT OUR DATA LOOKS LIKE

Scenario: 5. You have coffee-drinkers and you'd like to know if on days where they drink coffee, they have better grades.

drank coffee	didn't drink coffee
1.3	2.3
1.7	2.7
3.7	2.0
2.7	2.0
3.7	1.7
3.7	1.3
2.3	3.0
2.0	3.0
1.3	1.3
1.7	3.7
…	…

THE WILCOXON TEST



THAT'S WHAT R GIVES ME


```
> wilcox.test(sample1, sample2)

	Wilcoxon rank sum test with continuity correction

data:  sample1 and sample2
W = 1327, p-value = 2.905e-08
alternative hypothesis: true location shift is not equal to 0
```

p(the data | H0) < 0.05

H0: Coffee drinkers don't have different grades depending on their coffee input.

H1: Coffee drinkers have different grades depending on their coffee input.