Data Analysis in Sociology

Lecture 3. t-test.

February 2023

NATIONAL RESEARCH UNIVERSITY
SAINT PETERSBURG

A couple of organizational announcements

Data clinic session: today (Feb, 6), 19:50, online.

My office hours: Thursday (Feb, 9), 15:00, online.

Deadline for Project 1 submission: Feb 13, 23:59

Previously on this course

Research hypothesis vs. statistical hypothesis

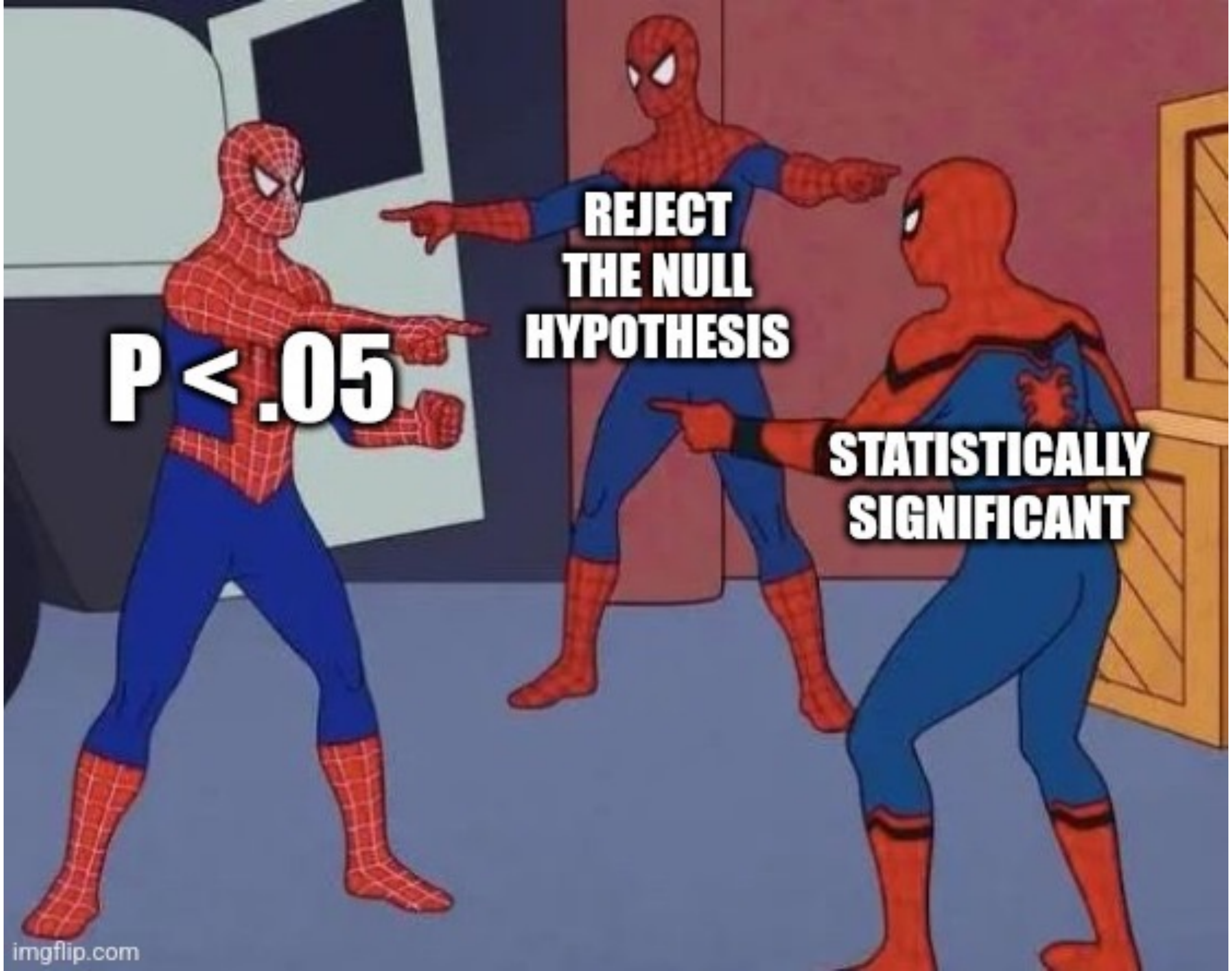
P-value

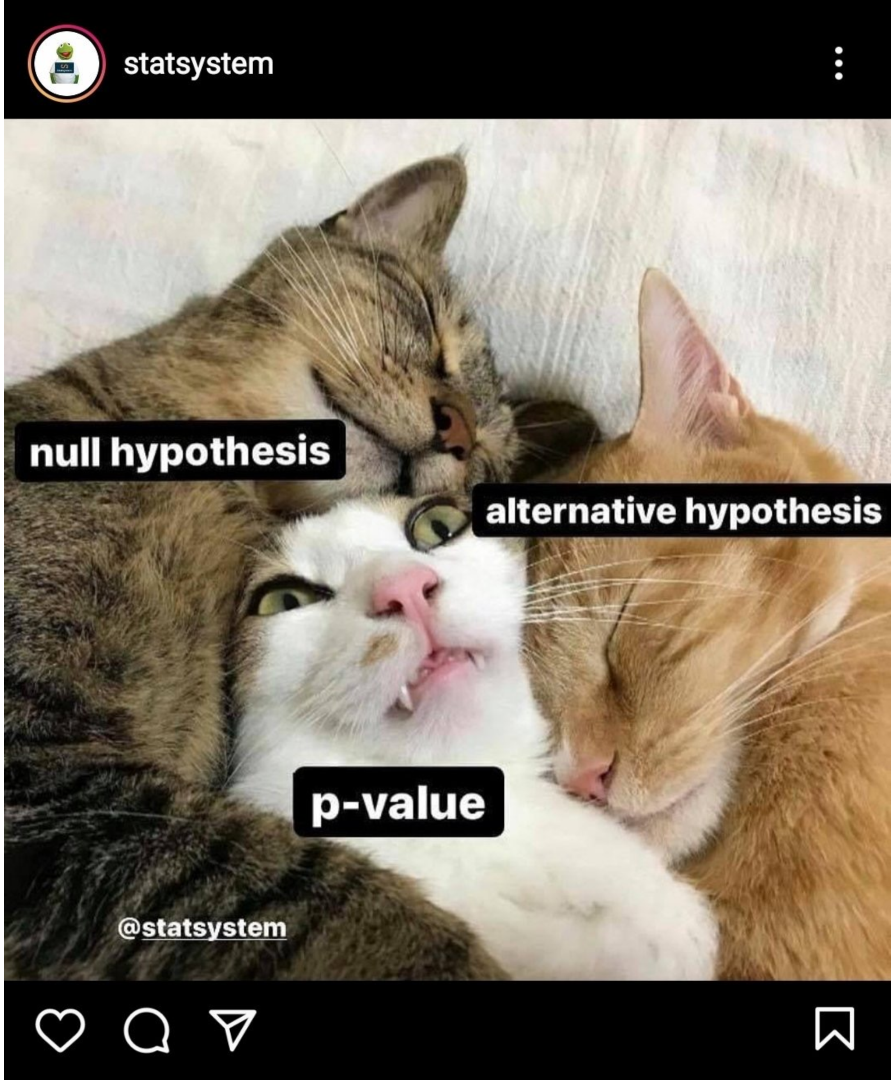
Central tendency measures, measures of variability

Basic graphs

Chi-squared test of independence

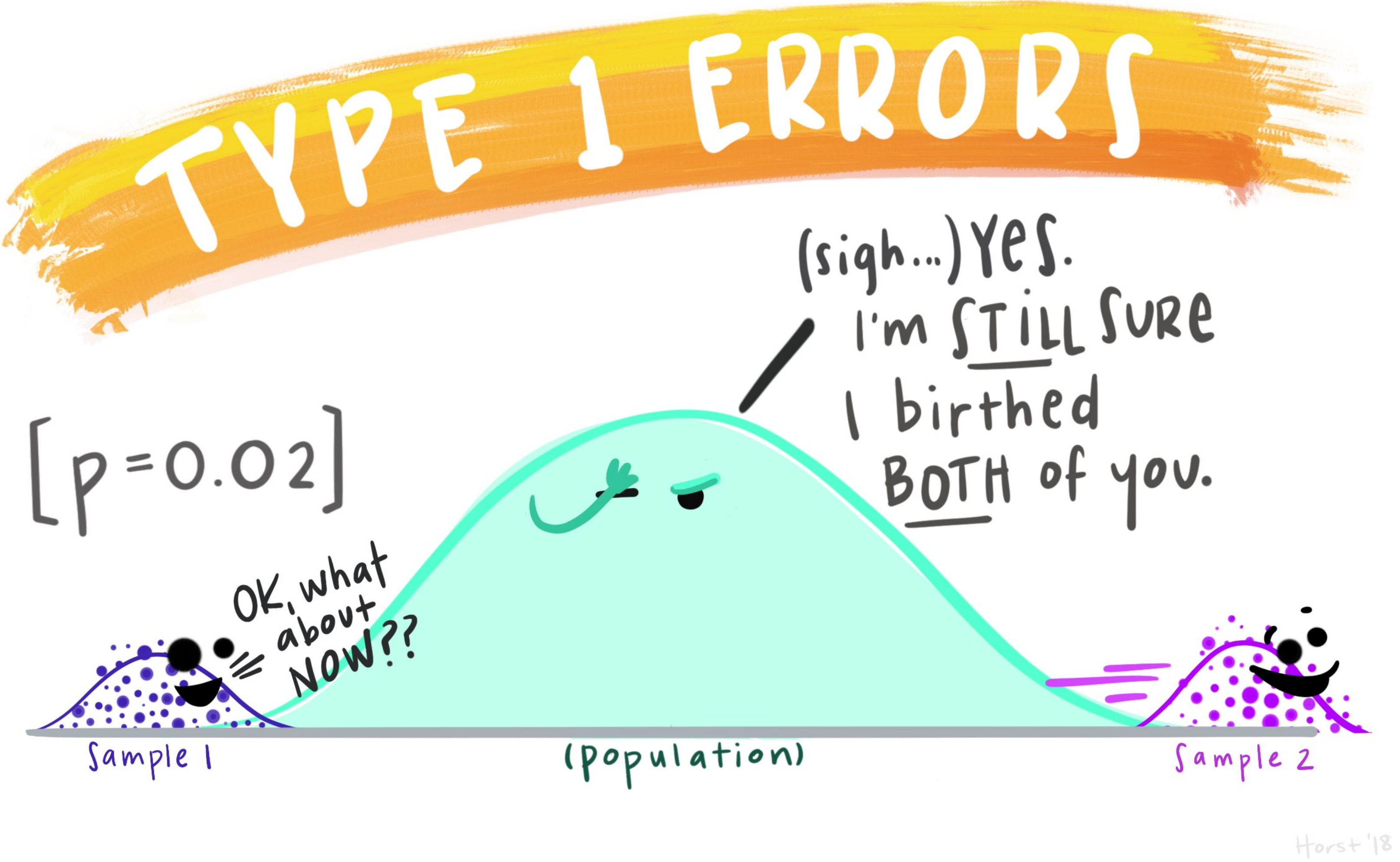

Reject,
or fail to reject.
There is no accept.


REJECT
THE NULL
HYPOTHESIS
P < .05
STATISTICALLY
SIGNIFICANT
imgflip.com

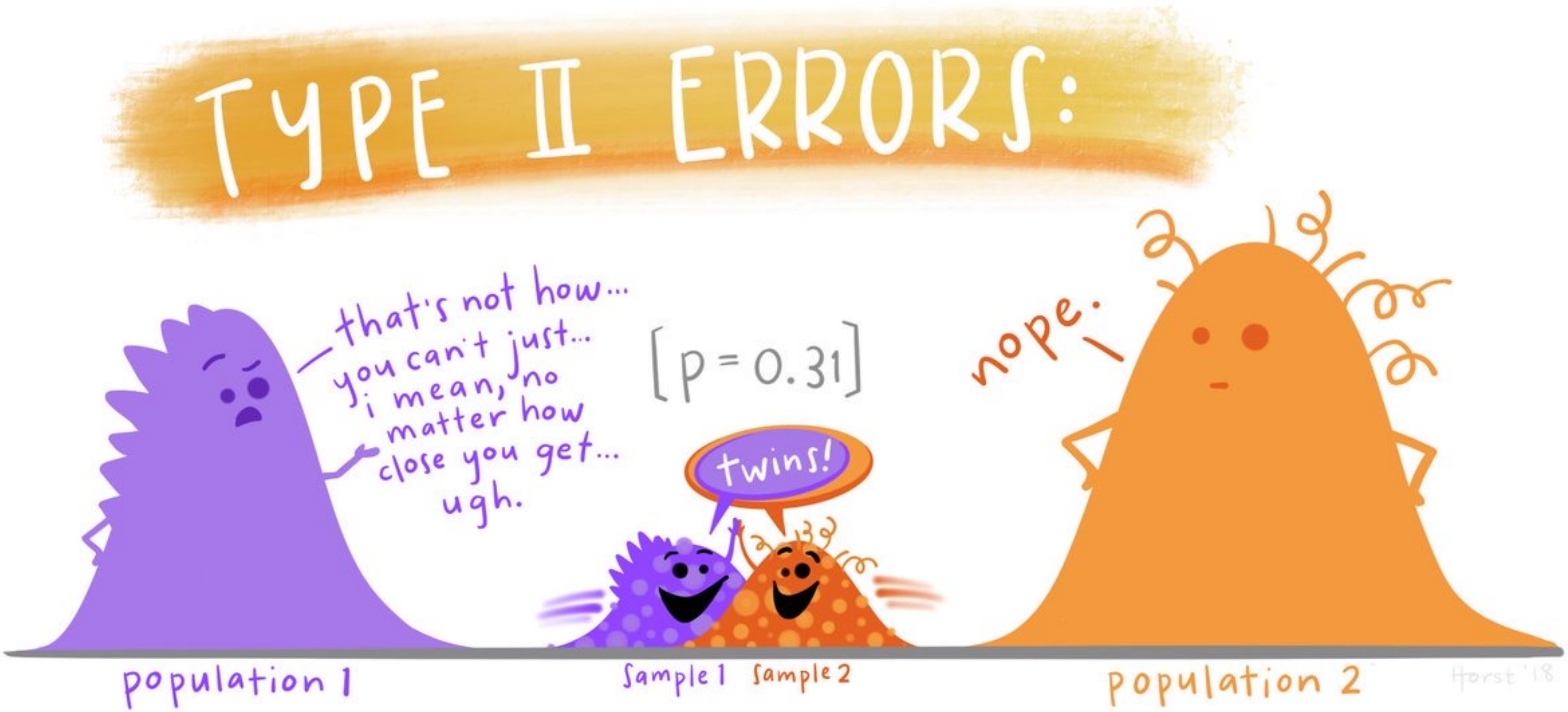

statsystem
null hypothesis
alternative hypothesis
p-value
@statsystem


CHI-SQUARES AREN'T FOR CONTINUOUS DATA
CHALLENGE ACCEPTED
memegenerator.net

Artwork by @allison_horst



Artwork by @allison_horst



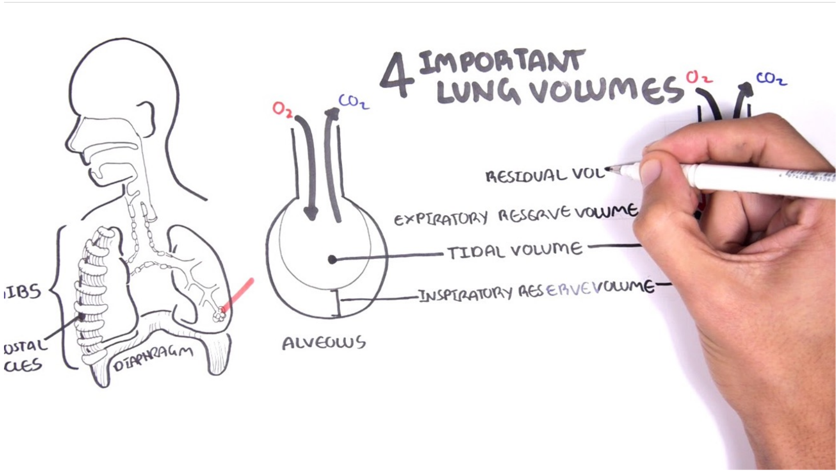
How to find relationship between different types of variables?

	Categorical	Numeric
Categorical	**Chi-square test**	
Numeric		

Solve a problem:

Do smokers have smaller lung capacity compared to non-smokers?

What graph can help?



Can the chi-square test solve it?

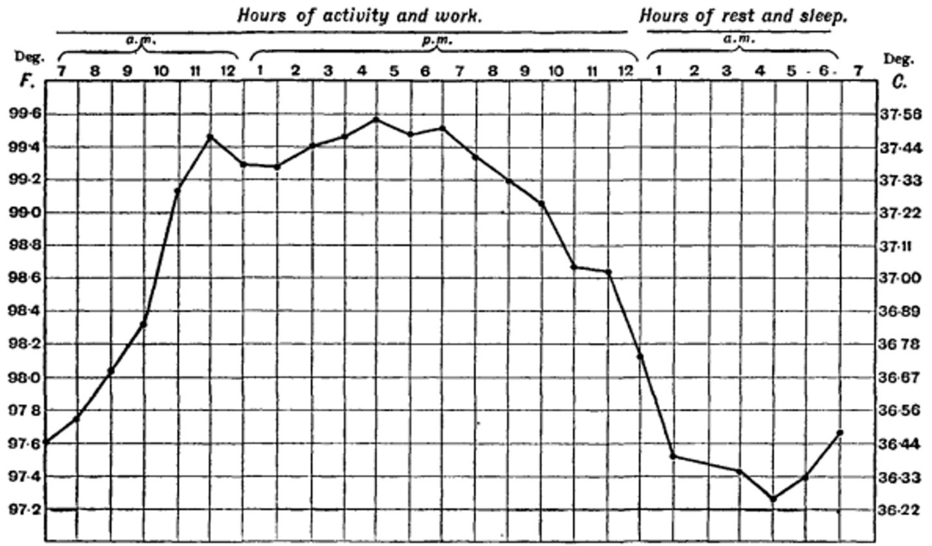
Another problem

Is body temperature lower in the mornings than in the afternoon?

How many observations are needed?

What graph to use?

How is this problem different from the lung capacity one?



Yet another problem

At a beer factory, supply of barley is delivered. Does it match the brewery’s standards?

How to decide?

How to answer the question, “Is there a difference between the groups?”

1. ***Is this a question about proportions or about means?***
   How many variables?
   What is the scale of measurement of the variables?

2. To compare **categorical** variables, use the chi-squared test.
   To compare **two groups by a continuous variable**, use the t-test.

All these problems deal with means

Lung capacity problem:

two groups, different size → Independent-samples t-test

Temperature problem:

same observations,
'before and after', same size → Paired-samples t-test

Barley problem:

a sample is compared to
population parameters → One-sample t-test

Varieties of the t-test

Independent samples t-test (two groups coming from two entities)

The most common type

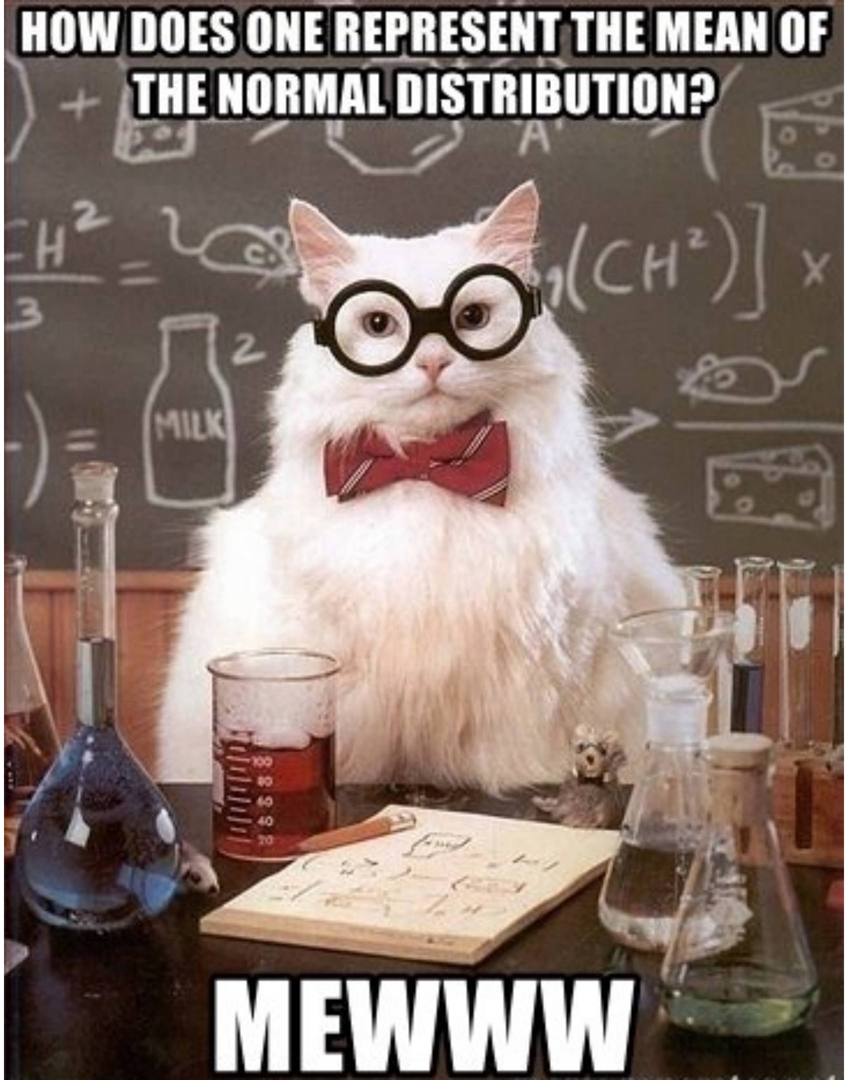
Also known as *unpaired samples, unrelated samples*

The **t statistic** is the difference between group means divided by the estimate of the standard error of that difference (square root of the sum of variances).

H0: $\mu1 = \mu2$ ($mu_1 = mu_2$)

H1 (two-sided): $\mu1 \neq \mu2$

H1 (one-sided): $\mu1 > \mu2$ or $\mu1 < \mu2$


HOW DOES ONE REPRESENT THE MEAN OF THE NORMAL DISTRIBUTION?
MEWWW

Independent samples t-test (two groups coming from two entities)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$df = n_1 + n_2 - 2$$

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_{X_1}^2 + (n_2 - 1)\, s_{X_2}^2}{n_1 + n_2 - 2}}$$

statistic

degrees of freedom

pooled standard deviation

In the case of equal group sizes and equal variances: df = 2n-n, and

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{2}{n}}}$$

$$s_p = \sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{2}}.$$

Independent samples t-test (two groups coming from two entities)

We only need to know the means, variances, and sample sizes to calculate the t statistic!

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$df = n_1 + n_2 - 2$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_{X_1}^2 + (n_2 - 1)s_{X_2}^2}{n_1 + n_2 - 2}}$$

statistic **degrees of freedom** **pooled standard deviation**

In the case of equal group sizes and equal variances: *df=2n-n*, and

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{2}{n}}}$$

$$s_p = \sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{2}}.$$

Independent samples t-test - when variances are not equal (Welch's t-test)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

$$\mathrm{d.f.} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1} + \frac{\left(s_2^2/n_2\right)^2}{n_2-1}}.$$

Paired-samples t-test

Also known as *dependent samples, related samples, matched samples, within-subject t-test, repeated-measures t-test*

Applies to: (1) repeated measures, *e.g., before vs after*, and (2) matched subjects, *e.g. twins*.

$\mathbf{t_{paired}}$ is the difference between the sample means minus the differences between population means, taking into account the standard error of the differences.

H0: $\mu_D = 0$

H1 (two-sided): $\mu_D \neq 0$

H1 (one-sided): $\mu_D > 0$ or H1: $\mu_D < 0$

Paired-samples t-test

$$t = \frac{\bar{X}_D - \mu_0}{\frac{s_D}{\sqrt{n}}}$$

df = n - 1

One-sample t-test (sample as part of the population)

$$t = \frac{\overline{X} - \mu}{S / \sqrt{N-1}}$$

df = n - 1

t is the difference between the sample mean and a specified value μ

H0: x - μ = 0

H1 (two-sided): x - μ ≠ 0

H1 (one-sided): x - μ > 0 or H1: x - μ < 0

Assumptions of the t-test

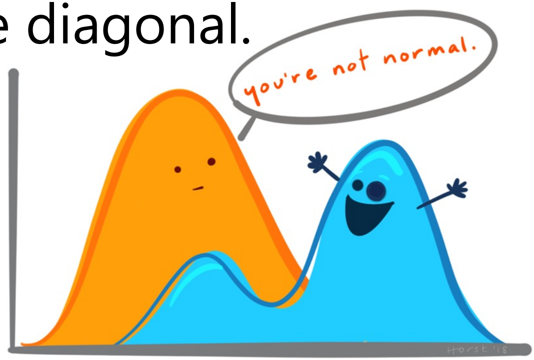
Assumptions of the two-sample t-test

- Observations are independent (“my response does not depend on yours”)
- Normally distributed sampling means in the two groups/residuals
  - less strict in large samples (e.g., n>300)
  - for paired t-test, the normally distributed difference
- <u>Equal variances across two groups</u>
  - if not, use Welch’s t-test

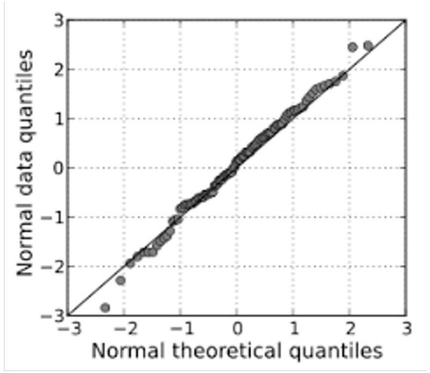
How to check normality of the variable distribution

There are many correct ways:

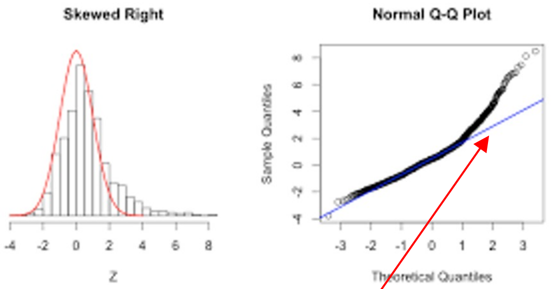
(1) Look at the histogram or a density plot of the residuals or the variable in each group - it should be bell-shaped.
(2) Look at the skew and kurtosis. Normal skew is up to +-0.5, normal kurtosis is within +-1 from zero.
(3) Look at the Q-Q plot -- the line should follow the diagonal.
(4) Formal tests: Shapiro-Wilk, Kolmogorov-Smirnov.



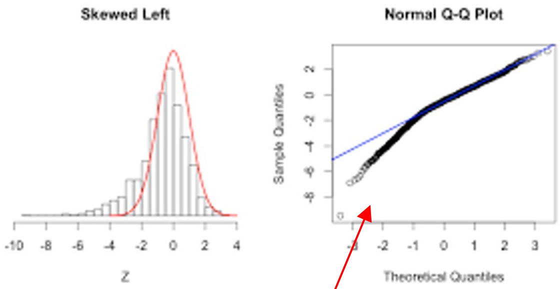
How to check normality of the variable



Q-Q plot
normal



Heavy right tail
(right skew)



Heavy left tail
(left skew)

How to check normality of the variable

```
> shapiro.test(normal)

        Shapiro-Wilk normality test

data:  normal
W = 0.97471, p-value = 0.5847
```

```
> shapiro.test(skewed)

        Shapiro-Wilk normality test

data:  skewed
W = 0.88499, p-value = 0.0016
```

normal

skewed

https://www.sheffield.ac.uk/polopoly_fs/1.885202!/file/95_Normality_Check.pdf

How to check equality of variances

There are many correct ways:

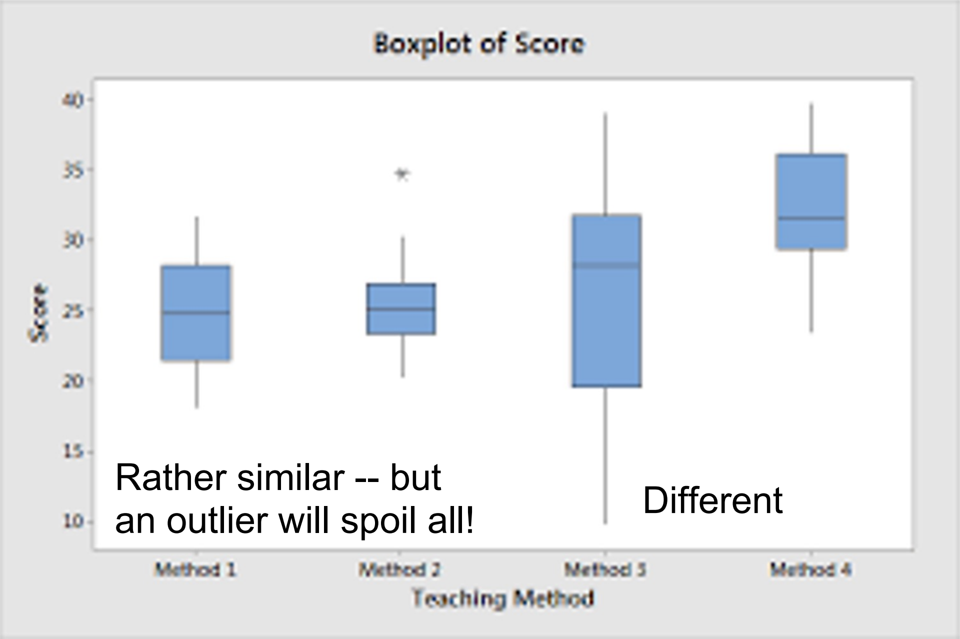
(1) Boxplots -- the boxes and the spread should be similar (approximate)

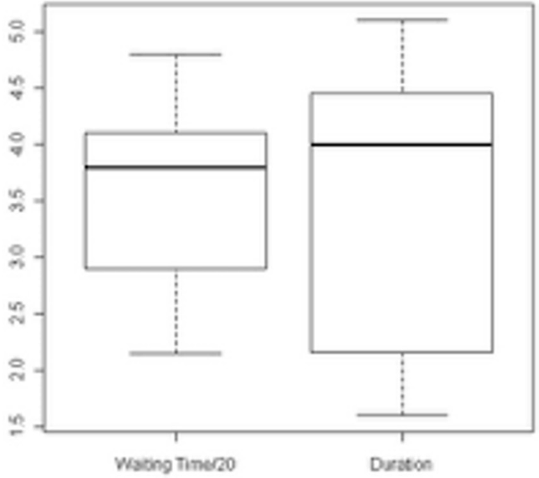
(2) Calculate and compare variances directly (approximate)

(3) Levene's test is the best -- robust to non-normality

(4) Bartlett's test (more sensitive to non-normality)

How to check equality of variances





Rather similar

How to check equality of variances

```
Levene's Test for Homogeneity of Variance (center = mean)
        Df F value    Pr(>F)
group    3  25.733 < 2.2e-16 ***
      5140
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Variances are not equal

Variances are equal

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value  Pr(>F)
group   3   2.304 0.07552 .
      887
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What if the assumptions are not met?

1. The t-test is highly robust to non-normality, especially if the samples are large (n > 100 or more). *It is Okay if you see non-normality*.For ordinal data or really small samples, Type I risk increases and the test's power is smaller. You can use the non-parametric Mann-Whitney-Wilcoxon test (it has no assumptions about the data). MW test converges raw scores into ranks.
2. If variances are unequal, use Welch's t-test that corrects for this (a special type of t-test).

How ranks work?

Raw observations: 1.1, 1.6, 10.8, 12, 14, 100, 1000, 100000

Ranks: 1, 2, 3, 4, 5, 6, 7, 8

Effect size for the t-test — 'okay, the difference is significant, but how large is it?'

Cohen's *d* is the difference between two means divided by a standard deviation for the data

$$d = \frac{M_E - M_C}{\text{Sample } SD \text{ pooled}} \times \boxed{\left(\frac{N-3}{N-2.25}\right) \times \sqrt{\frac{N-2}{N}}}$$

correction factor for small samples <50

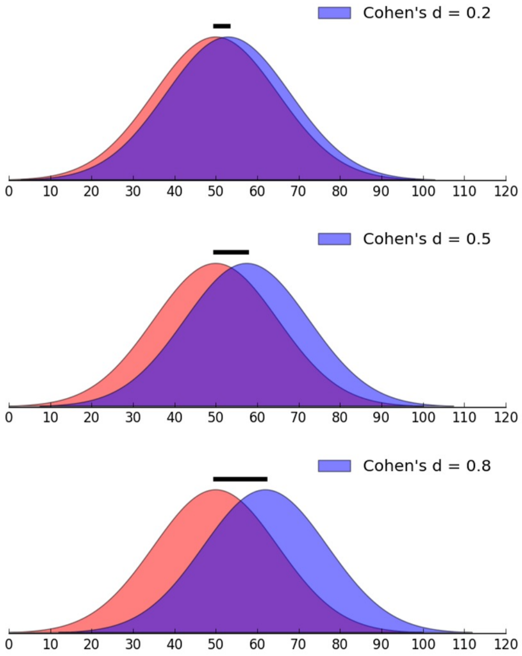
$$d = \frac{\bar{X}_1 - \bar{X}_2}{\text{SD}} = \frac{t}{\sqrt{N}}$$

Cohen's d for paired t-test

A d = 1 indicates the two groups differ by 1 SD, a d = 2 indicates they differ by 2 SDs

https://rpsychologist.com/cohend/ - explore

Effect size for the t-test



A d = 1 indicates the two groups differ by 1 SD,

a d = 2 indicates they differ by 2 SDs

https://rpsychologist.com/cohend/ - explore

https://scientificallysound.org/2017/07/27/cohens-d-how-interpretation/

Effect size for the t-test - Cohen's d

Very small = 0.1

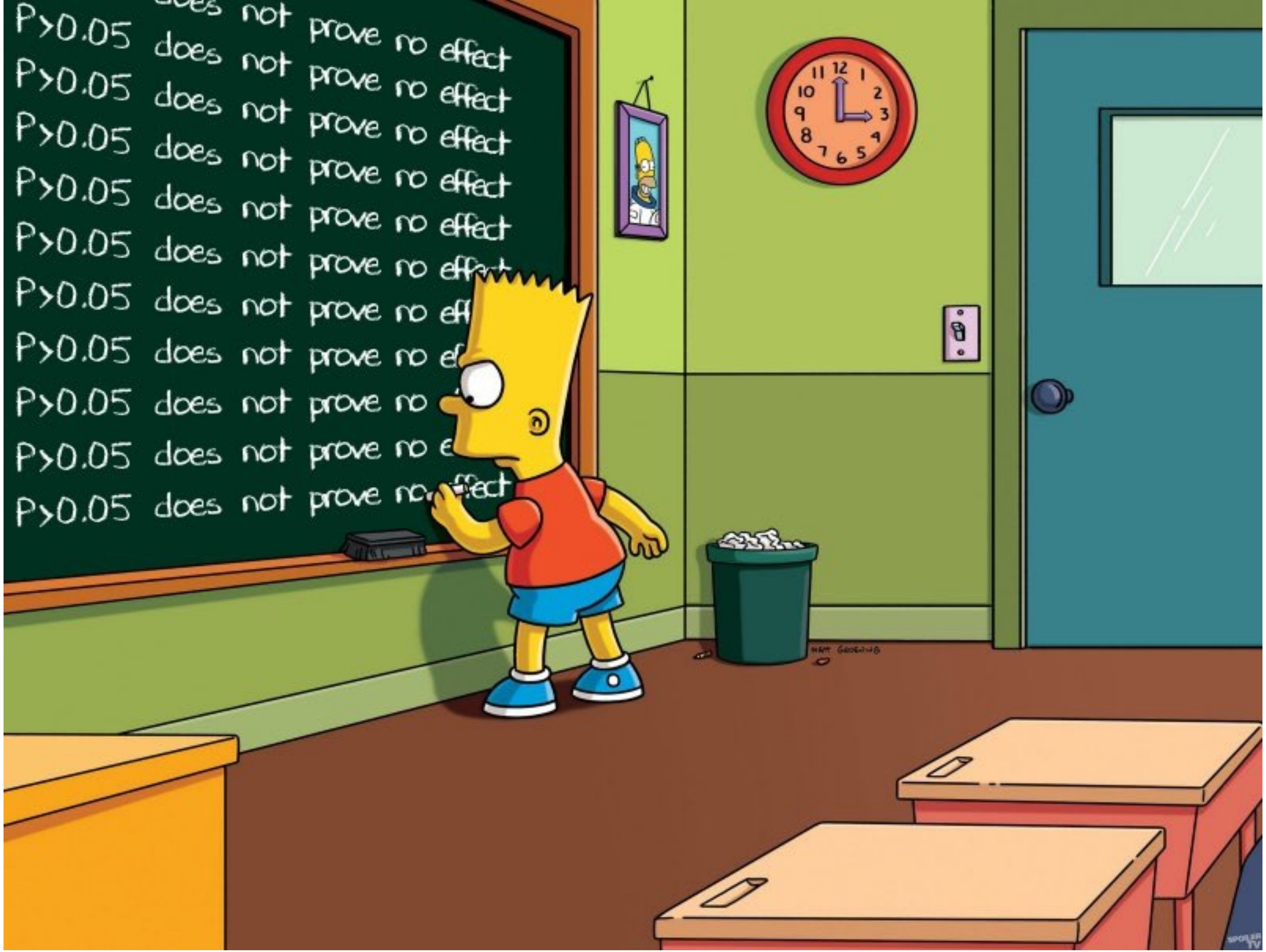
Small effect = 0.2

Medium Effect = 0.5,

Large Effect = 0.8

Relative size	Effect size	% of control group below the mean of experimental group
	0.0	50%
Small	0.2	58%
Medium	0.5	69%
Large	0.8	79%
	1.4	92%

Note! This is a rule of thumb, the threshold depends on the scientific field.


P>0.05 does not prove no effect
P>0.05 does not prove no effect
P>0.05 does not prove no effect
P>0.05 does not prove no effect
P>0.05 does not prove no effect
P>0.05 does not prove no
P>0.05 does not prove no
P>0.05 does not prove no
P>0.05 does not prove no
P>0.05 does not prove no
P>0.05 does not prove no


RESEARCHER
SEEKING PUBLICATION
TINY
EFFECT SIZE
TINY P-VALUE
imgflip.com


Report effect
size
for
all of the things!
quickmeme.com

Artwork by @allison_horst

R functions

```
shapiro.test(x)  # check normality, but not a strict assumption
under large samples
bartlett.test(y~x)  # check variance equality
library(car)
leveneTest(y~x)  # equal variances test; H0: variances are equal.

t.test(X1, X2, mu=0, paired = TRUE/FALSE, var.equal = TRUE/FALSE,
       alternative = “two.sided”/“less”/”greater”, conf.level=0.95)
wilcox.test(x1, x2, paired = TRUE/FALSE)

library(lsr)
cohensD(x1, x2)  # effect size of the t-test
library(wBoot); boot.two.bca() # bootstrapped confidence intervals*
(advanced)
```

How to find relationship between different types of variables?

	Binary	Categorical (3+ cats)	Numeric
Binary	Chi-square test (Yates correction)	Chi-square test	**T-test** **MW U-test**
Categorical (3+ cats)	Chi-square test	Chi-square test	
Numeric	**T-test** **MW U-test**		

Resources

Annotated script: http://www.sthda.com/english/wiki/unpaired-two-samples-t-test-in-r

More about t-test in R:

https://statistics.berkeley.edu/computing/r-t-tests

https://www-users.york.ac.uk/~er13/17C%20-%202018/pracs/05OneAndTwoSampleTests.html

Think of the following situations:

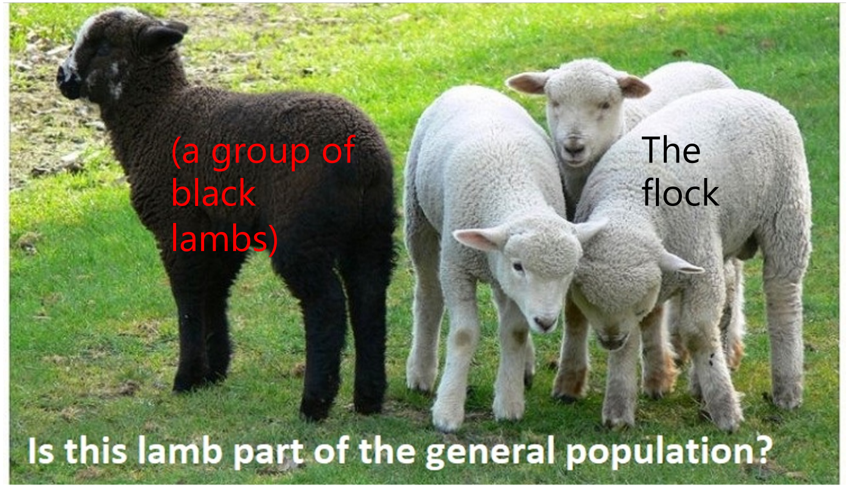
Many more people here

Many more people here

You need to compare the satisfaction with organisational culture between two departments of an organisation. There are 40 and 33 people in them. What do you do?

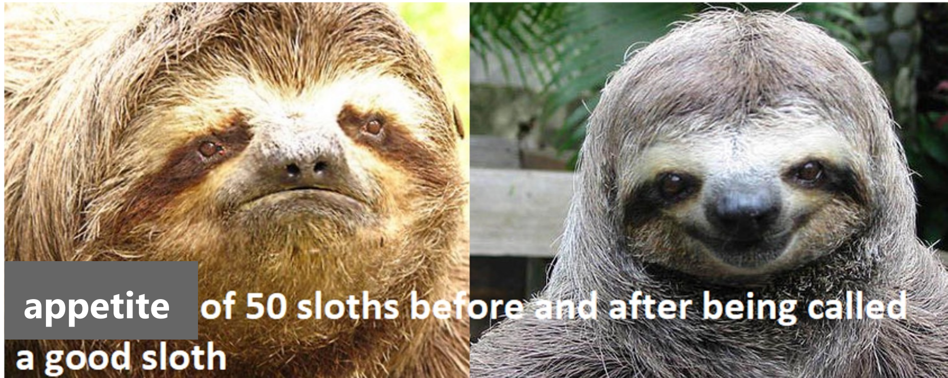
Think of the following situations:

You are in another dream where a shepherd asks you to check whether a group of lambs belongs to his flock.



Think of the following situations:

You work at a zoo and your boss asks you to measure the appetite of 50 sloths you happen to have there (with a sophisticated chemical test that measures happiness) before and after they are called a good sloth.



t-test can help!

T-test estimates whether the mean values are equal or not equal.
The measured variable (means) should be continuous.

Example	What is compared	test	Dependent variable
2 large departments	Satisfaction between two departments	Independent samples t-test	Satisfaction with organisational culture
50 sloths	Appetite before and after being called a good sloth	Paired-samples t-test	Appetite
A few lambs and a flock	Do these lambs belong to this flock?	One-sample t-test	Height

Independent samples t-test - Example

Satisfaction (0-100) with organization culture among two departments.

H0: $\mu 1 = \mu 2$

H1 (two-sided): $\mu 1 \neq \mu 2$

Dep't 1:

mean1 = 68, SD1 = 15, n1 = 40

Dep't 2:

mean2 = 76, SD2 = 15, n2 = 33

```
set.seed(1234)
x <- rnorm(n=40, mean=68, sd=15)
y <- rnorm(n=33, mean=76, sd=15)
t.test(x,y, var.equal = T)
```

```
> t.test(x,y, var.equal = T)

        Two Sample t-test

data:  x and y
t = -3.5778, df = 71, p-value = 0.0006292
alternative hypothesis: true difference
in means is not equal to 0
95 percent confidence interval:
 -19.623399  -5.578252
sample estimates:
mean of x mean of y
 61.79200  74.39282
```

Independent samples t-test - Example

Satisfaction (0-100) with organization culture among two departments.

If alpha = 0.01, t(71) = -3.6,
p = 0.001.
H0 can be rejected.

The levels of satisfaction are different between the two departments.

```
> t.test(x,y, var.equal = T)

        Two Sample t-test

data:  x and y
t = -3.5778, df = 71, p-value = 0.0006292
alternative hypothesis: true difference
in means is not equal to 0
95 percent confidence interval:
 -19.623399  -5.578252
sample estimates:
mean of x mean of y
 61.79200  74.39282
```

Independent samples t-test - Example

Satisfaction (0-100) with organization culture among two departments.

What if the departments are really small, say 6 and 8? Use the non-parametric Wilcoxon-Mann-Whitney rank sum test:

```
> wilcox.test(x1, y1)

        Wilcoxon rank sum test

data:  x1 and y1

W = 17, p-value = 0.4136

alternative hypothesis: true location shift is not equal to 0
```

At W = 17 and p-value = 0.41, H0 cannot be rejected.
Satisfaction between the departments is not different.

```
x1 <- rnorm(n=6, mean=68, sd=15)
y1 <- rnorm(n=8, mean=76, sd=15)
wilcox.test(x1, y1)
```

Paired-samples t-test - Example

Did the sloths eat more after an adoration session? There were 50 sloths.

H0: $\mu_D = 0$

H1 (one-sided): $\mu_D < 0$

$t(49)=-3.32$, $p=0.0009$,

H0 can be rejected.
The sloths ate more
after the session.

```
> t.test(x = x2, y = y2, paired = T,
alternative = "less")

        Paired t-test
data:  x2 and y2
t = -3.3179, df = 49, p-value =
0.0008577
alternative hypothesis: true
difference in means is less than 0
95 percent confidence interval:
     -Inf -4.89194
sample estimates:
mean of the differences
              -9.888738
```

```
set.seed(1234)
x2 <- rnorm(n=50, mean=68, sd=15)
y2 <- rnorm(n=50, mean=69, sd=15)
t.test(x=x2, y=y2, paired=T, alternative="less")
```

Paired-samples t-test - Example

What is the zoo was small and there were 6 sloths? Use Wilcoxon signed-rank:

```
> wilcox.test(x=x3, y=y3, paired=T, alternative="less")

	Wilcoxon signed rank test

data:  x3 and y3
V = 15, p-value = 0.8438
alternative hypothesis: true location shift is less than 0
```

H0: $\mu_D = 0$, H1 (one-sided): $\mu_D < 0$

V=15, p=0.84, H0 cannot be rejected.

The sloths did not eat more after the session.

```
set.seed(1234)
x3 <- rnorm(n=6, mean=68, sd=15)
y3 <- rnorm(n=6, mean=69, sd=15)
wilcox.test(x=x3, y=y3, paired=T, alternative="less")
```

One-sample t-test - Example

Do the sheep you see belong to a particular population with mu = 36.6?

H0: x - μ = 0 (H0: x = μ)

H1 (two-sided): x - μ ≠ 0

The sheep in question:

n = 30, mean = 38

t(29) = 0.94, p = 0.36

H0 is retained, they are not different!

```
set.seed(1234)
x3 <- rnorm(n=30, mean=38, sd=15)
t.test(x3, mu = 36.6, alternative = "two.sided")
```

```
> t.test(x3, mu = 36.6, alternative = 
"two.sided")

        One Sample t-test

data:  x3
t = 0.93619, df = 29, p-value = 0.3569
alternative hypothesis: true mean is 
not equal to 36.6
95 percent confidence interval:
 33.70535 44.38164
sample estimates:
mean of x 
  39.0435
```

Effect size for the t-test - Cohen's d

- A d of 1 indicates the two groups differ by 1 SD, a d of 2 indicates they differ by 2 SDs, and so on
- Small effect = 0.2, Medium Effect = 0.5, Large Effect = 0.8

The satisfaction between departments example, t(71)=-3.6, p = 0.001:

```
> cohensD(x, y)
[1] 0.8413757 # large effect
```

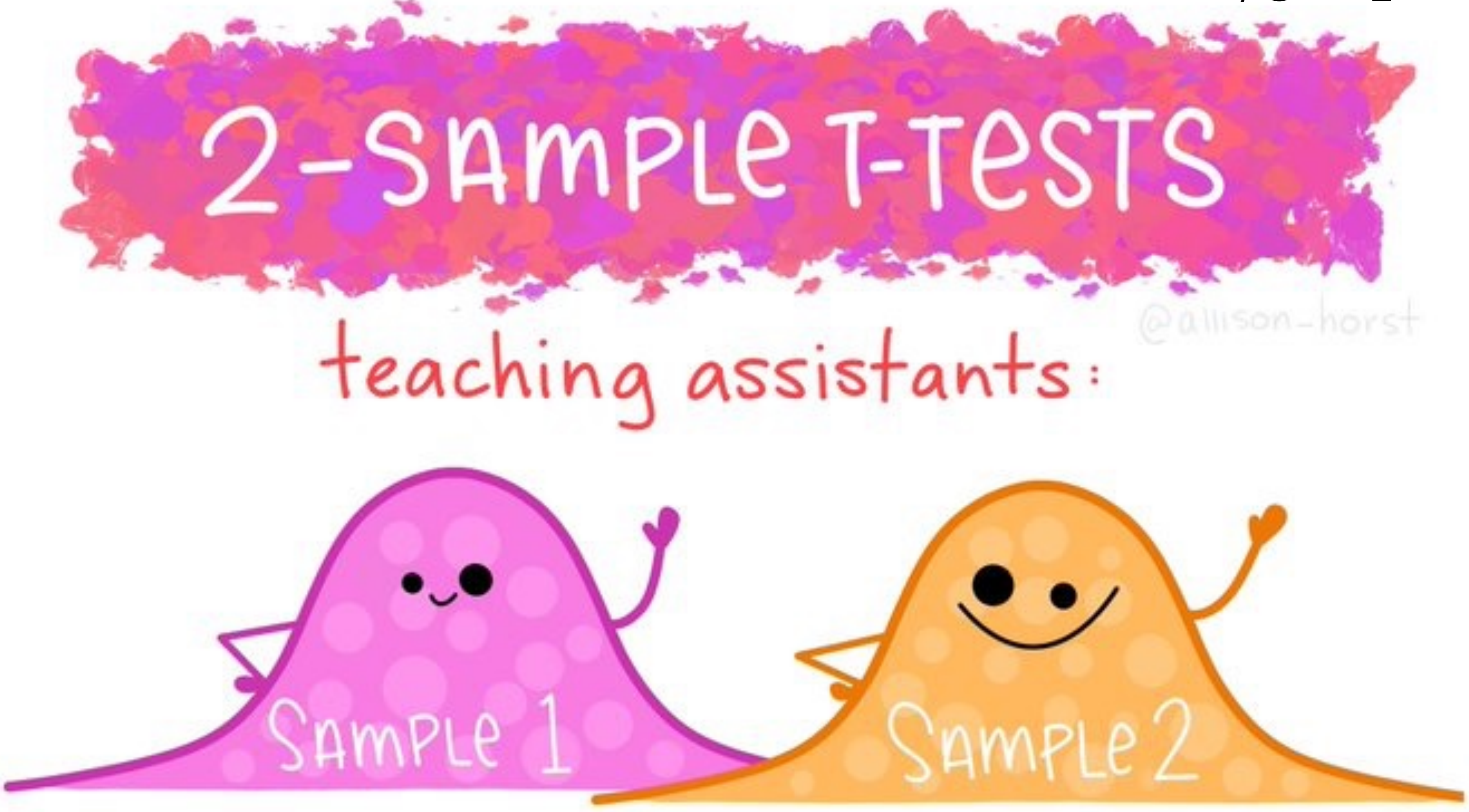
The sloths' example, t(49)=-3.32, p=0.0009:

```
> cohensD(x2, y2, method = "paired")
[1] 0.4692246 # medium effect
```

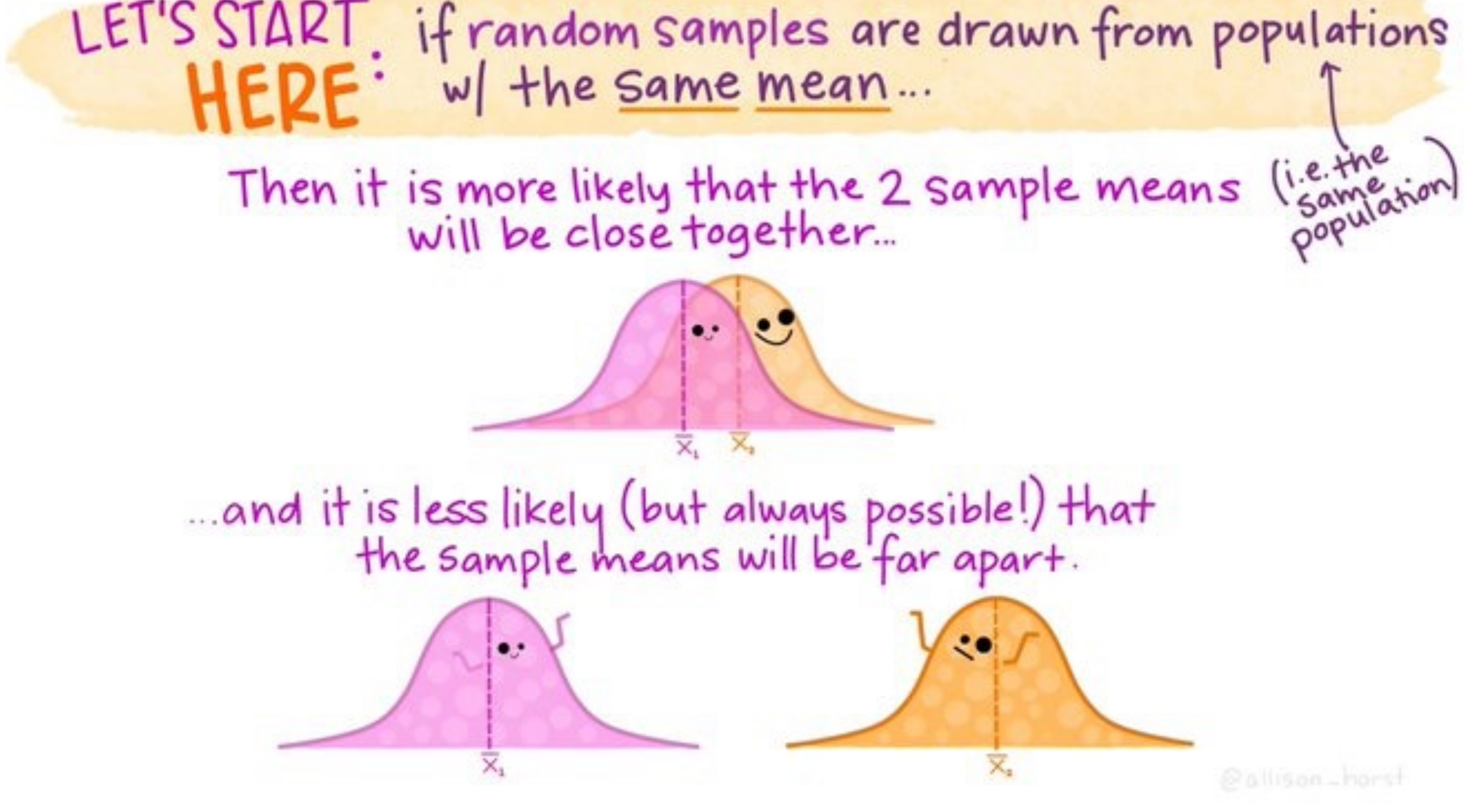
IgNobel prize examples

- GARCÍA-HERNÁNDEZ, S., & Machado, G. (2022). Short-and long-term effects of an extreme case of autotomy: does “tail” loss and subsequent constipation decrease the locomotor performance of male and female scorpions?. Integrative Zoology, 17(5), 672-688.
- Giacomin, Miranda & Rule, Nicholas. (2017). Eyebrows Cue Grandiose Narcissism. Journal of Personality. 87. 10.1111/jopy.12396.
- E A Beseris, S E Naleway, D R Carrier, Impact Protection Potential of Mammalian Hair: Testing the Pugilism Hypothesis for the Evolution of Human Facial Hair, Integrative Organismal Biology, Volume 2, Issue 1, 2020, obaa005
- Schötz, S., & Eklund, R. (2011). A comparative acoustic analysis of purring in four cats. Proceedings from Fonetik 2011, Quarterly Progress and Status Report TMH-QPSR, Volume 51, 2011, 5–8.

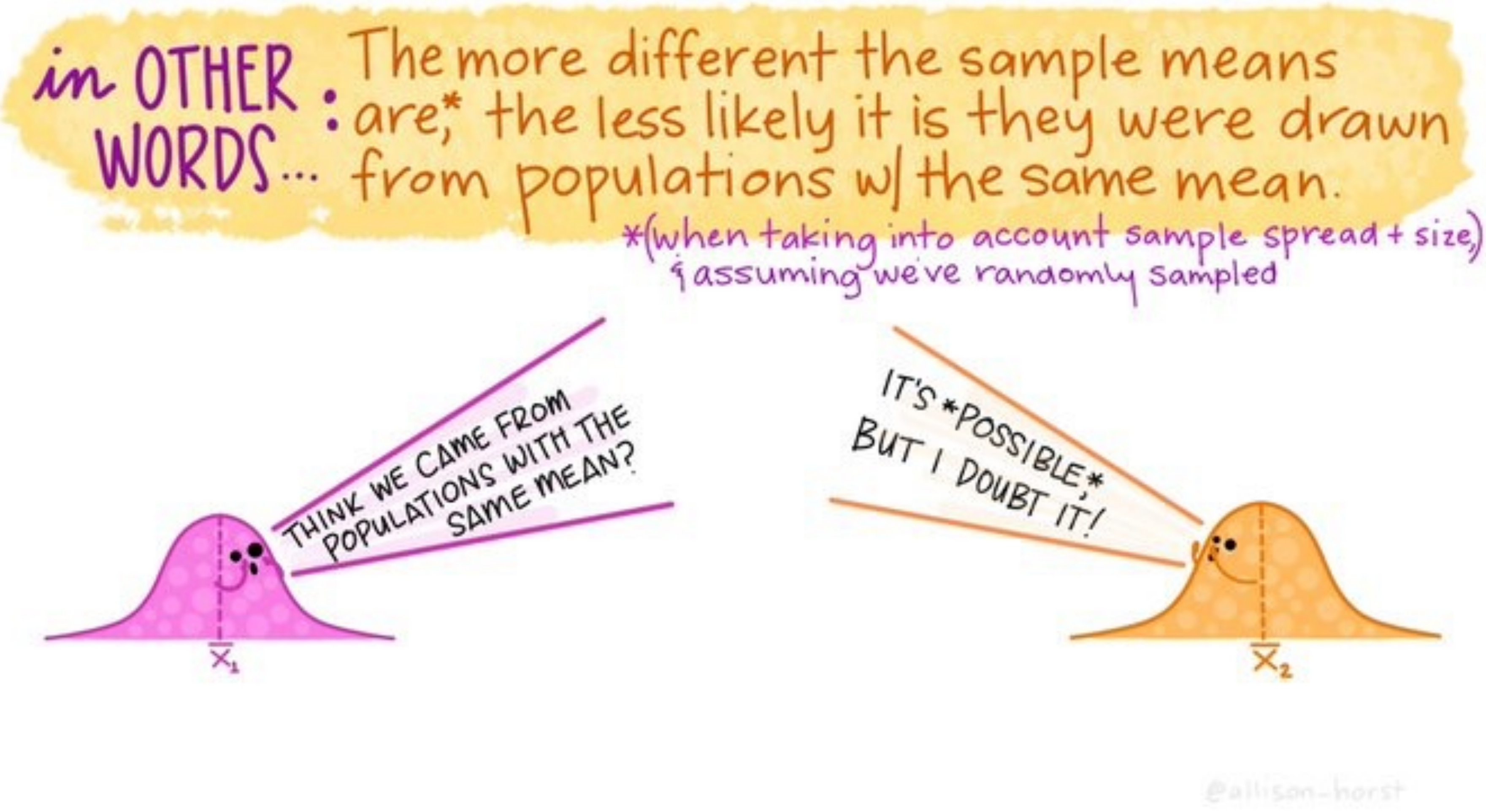
Artwork by @allison_horst



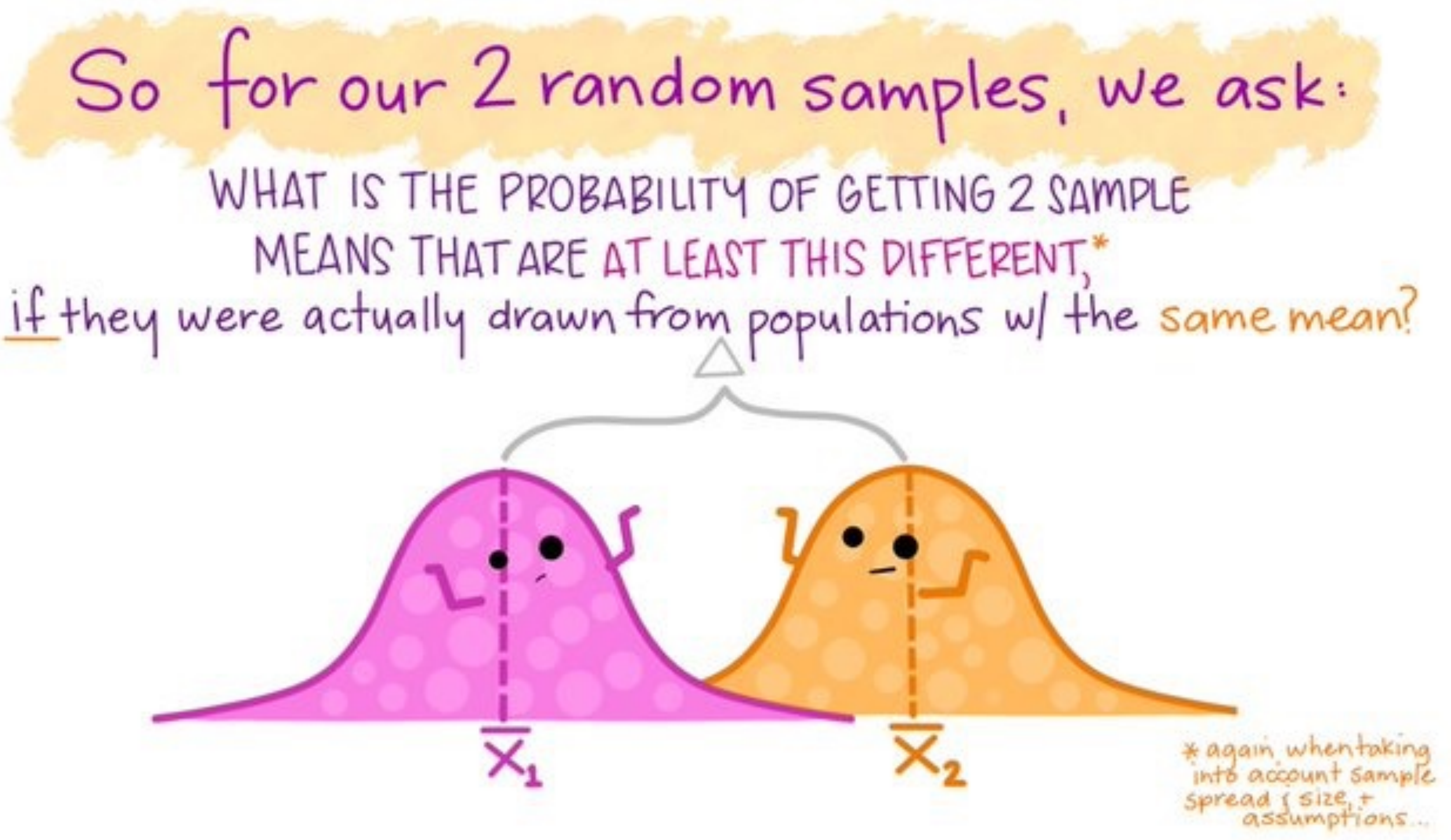
Artwork by @allison_horst



Artwork by @allison_horst



Artwork by @allison_horst



Artwork by @allison_horst