ECE750: Usable Security and Privacy

Statistical TestsStatistical Tests

Dr. Kami Vaniea
Electrical and Computer Engineering
kami.vaniea@uwaterloo.ca

UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING

First, the news...

- First 5 minutes we talk about something interesting and recent
- You will not be tested on the news part of lecture
- You may use news as an example on tests
- Why do this?
    1. Some students show up late for various good reasons
    2. Reward students who show up on time
    3. Important to see real world examples

STATISTICAL TESTS

Comparing	Dependent	Independent	Parametric (Dependent variable is mostly normally distributed)	Non-parametric
The means of two independent groups	Continuous / scale	Categorical / nominal	Independent t-test	Mann-Whitney test
The means of 2 paired (matched) samples	Continuous / scale	Time variable (before/after)	Paired t-test	Wilcoxon signed rank test
The means of 3+ independent groups	Continuous / scale	Categorical / nominal	One-way ANOVA	Kruskal-Wallis test
3+ measurements on the same subject	Continuous / scale	Time variable	Repeated measures ANOVA	Friedman test
Relationship between 2 continuous variables	Continuous / scale	Continuous / scale	Pearson's Correlation Coefficient	Spearman's Correlation Co-efficient
Predicting the value of one	Continuous / scale	Any	Simple Linear Regression	

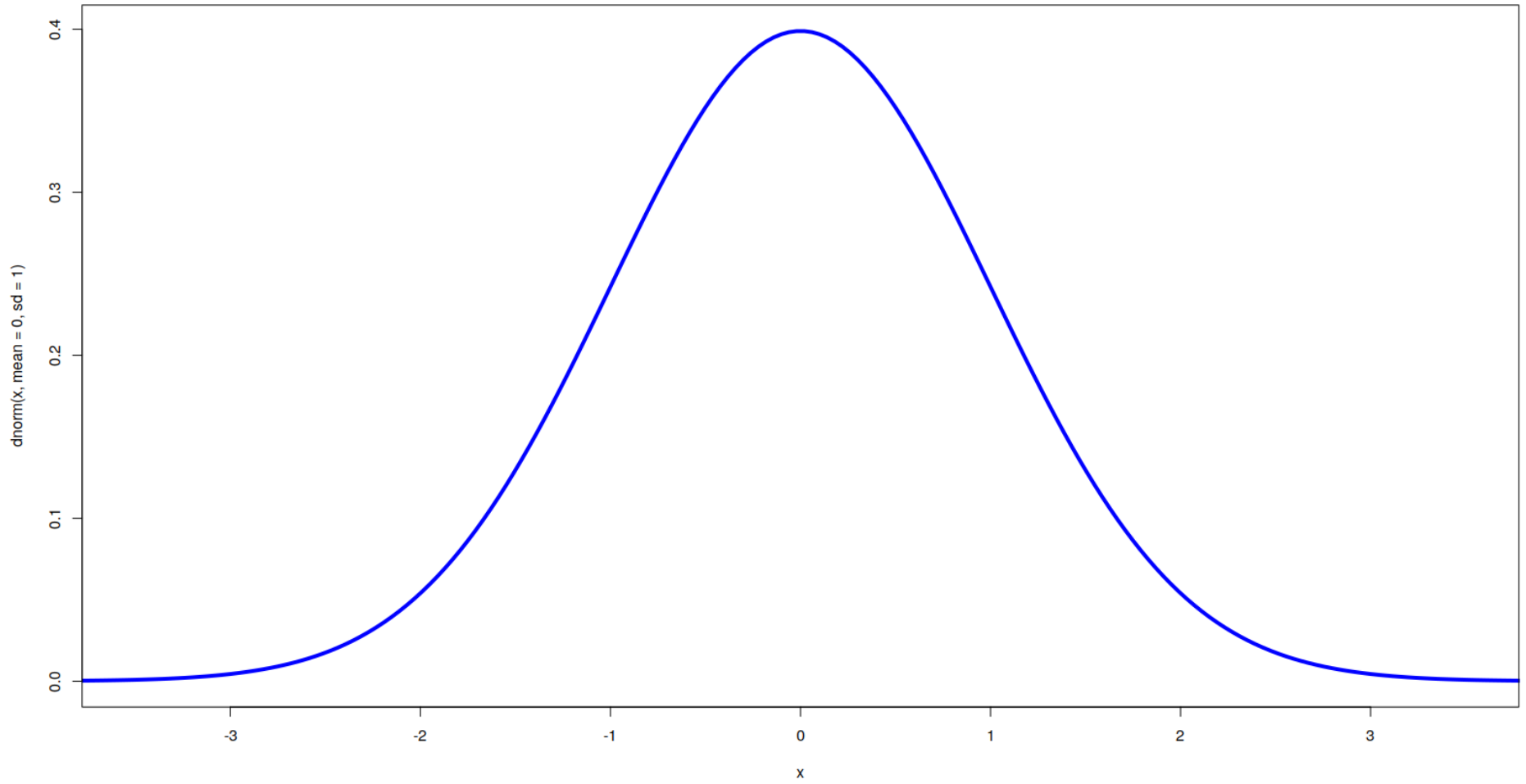
t-test: Test if two groups have the same mean (average)

T-test requires:

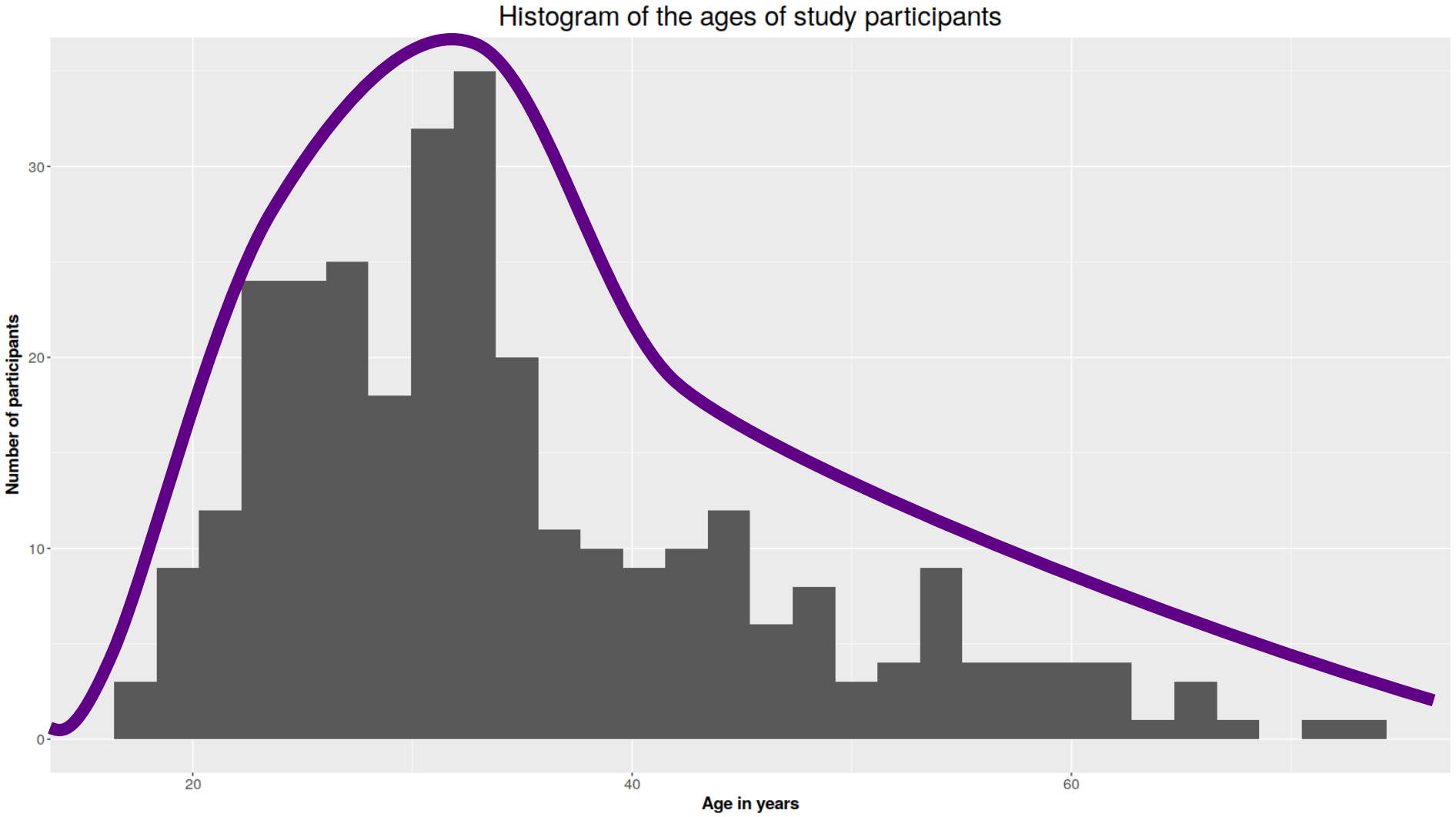
- Independent variable: categorical binary
- Dependent variable: numeric (continuous or discrete)

- Data must be normally distributed

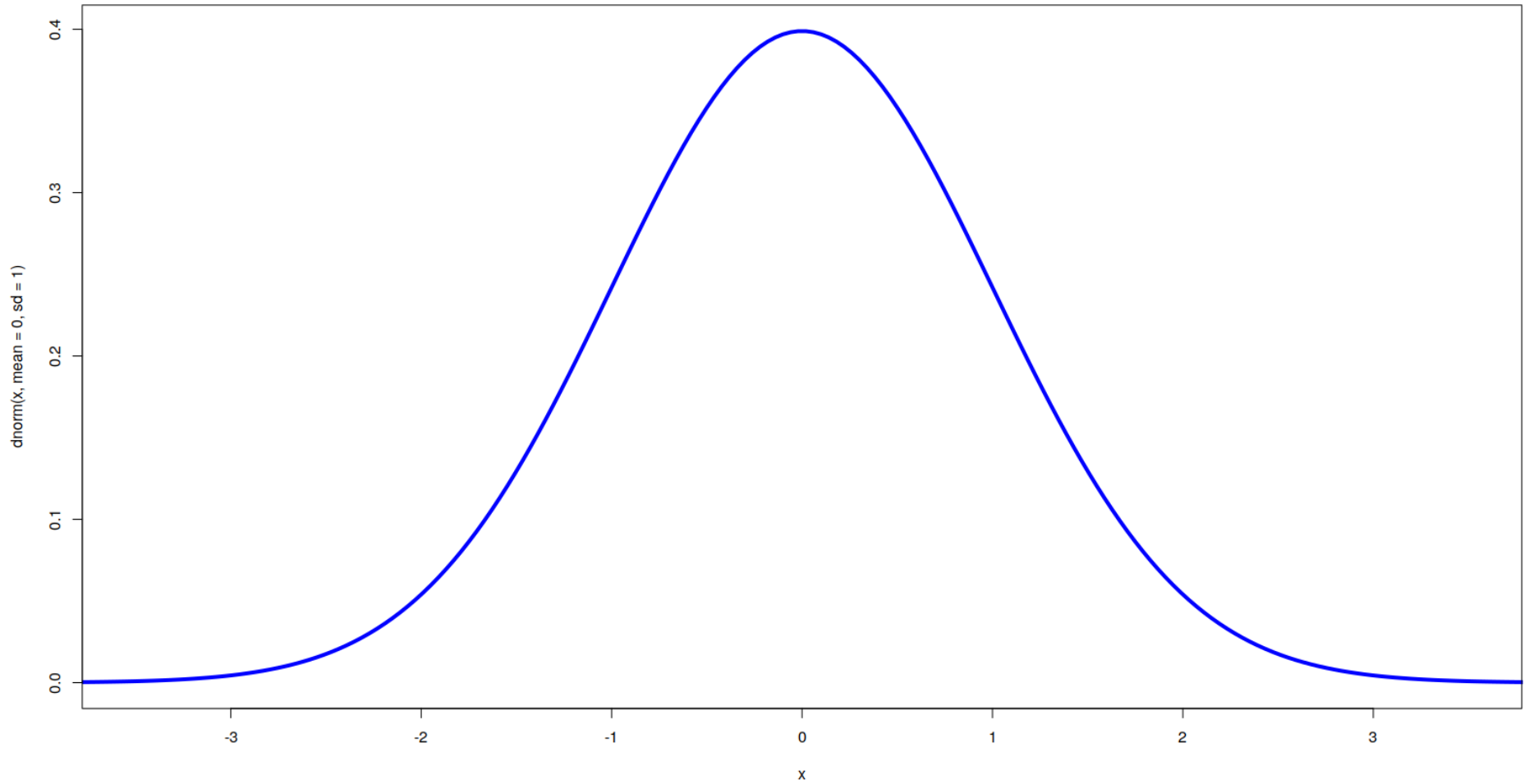
Normal distribution



Real data is messy

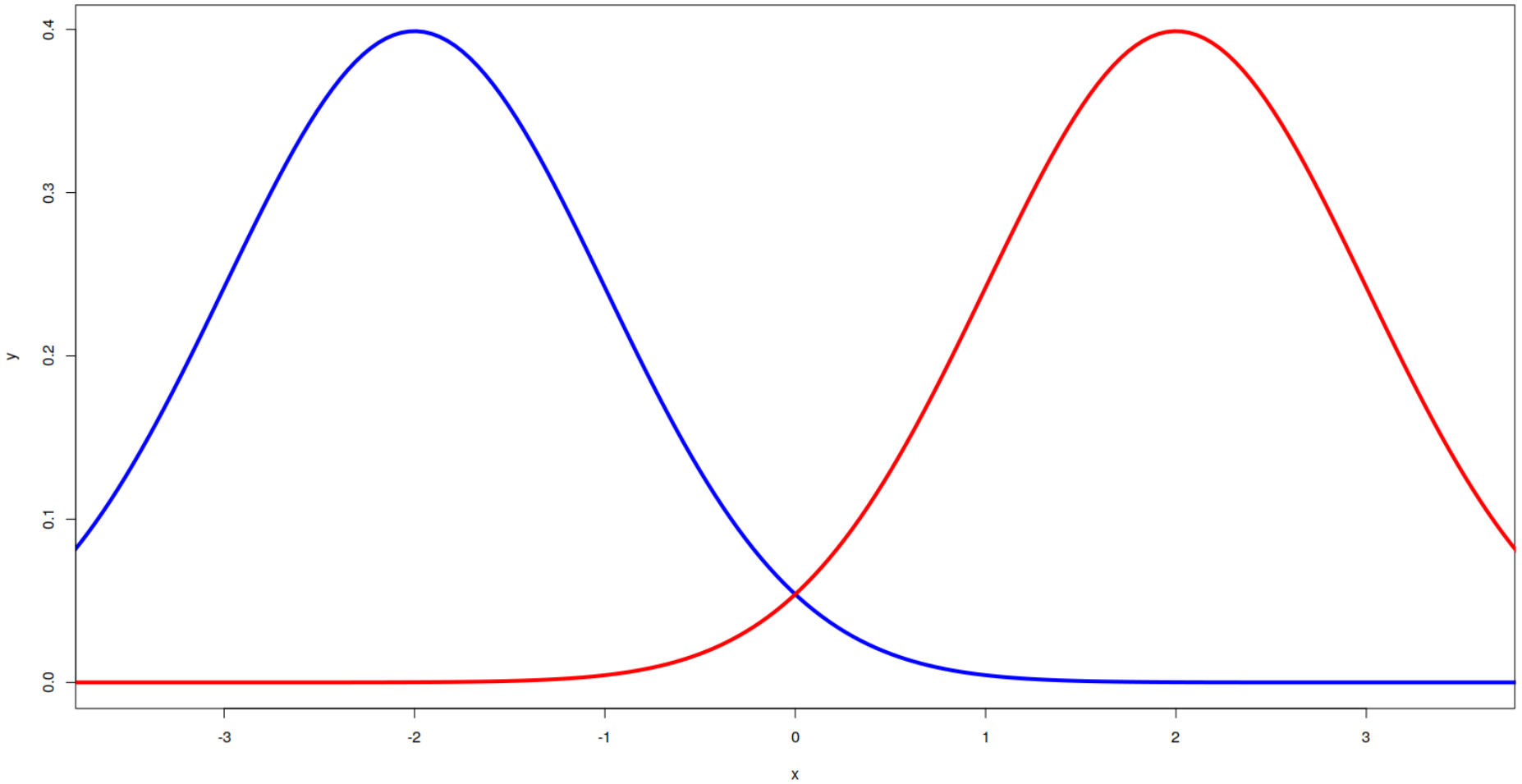


Normal distribution

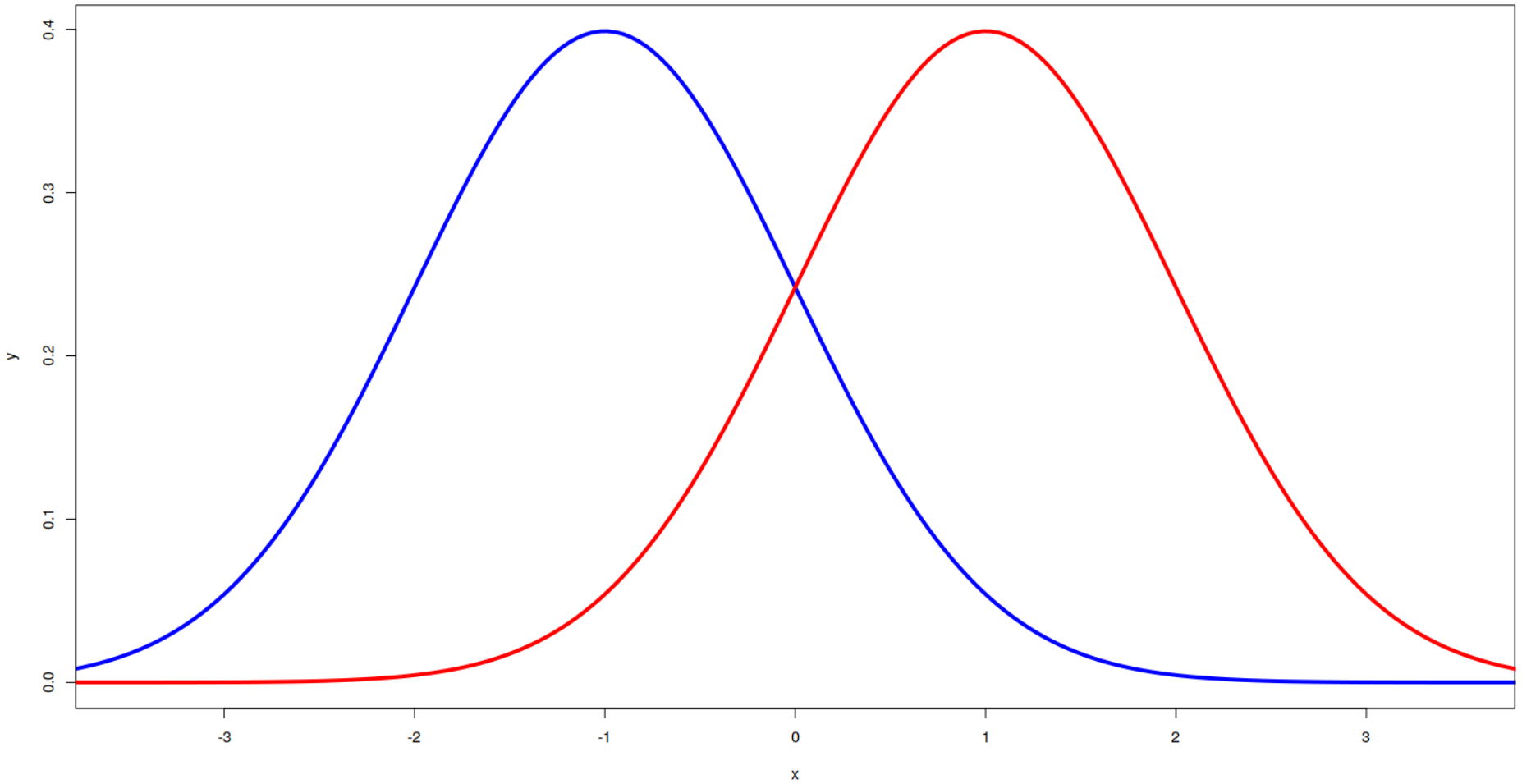


T-test: Do two populations have the same mean?

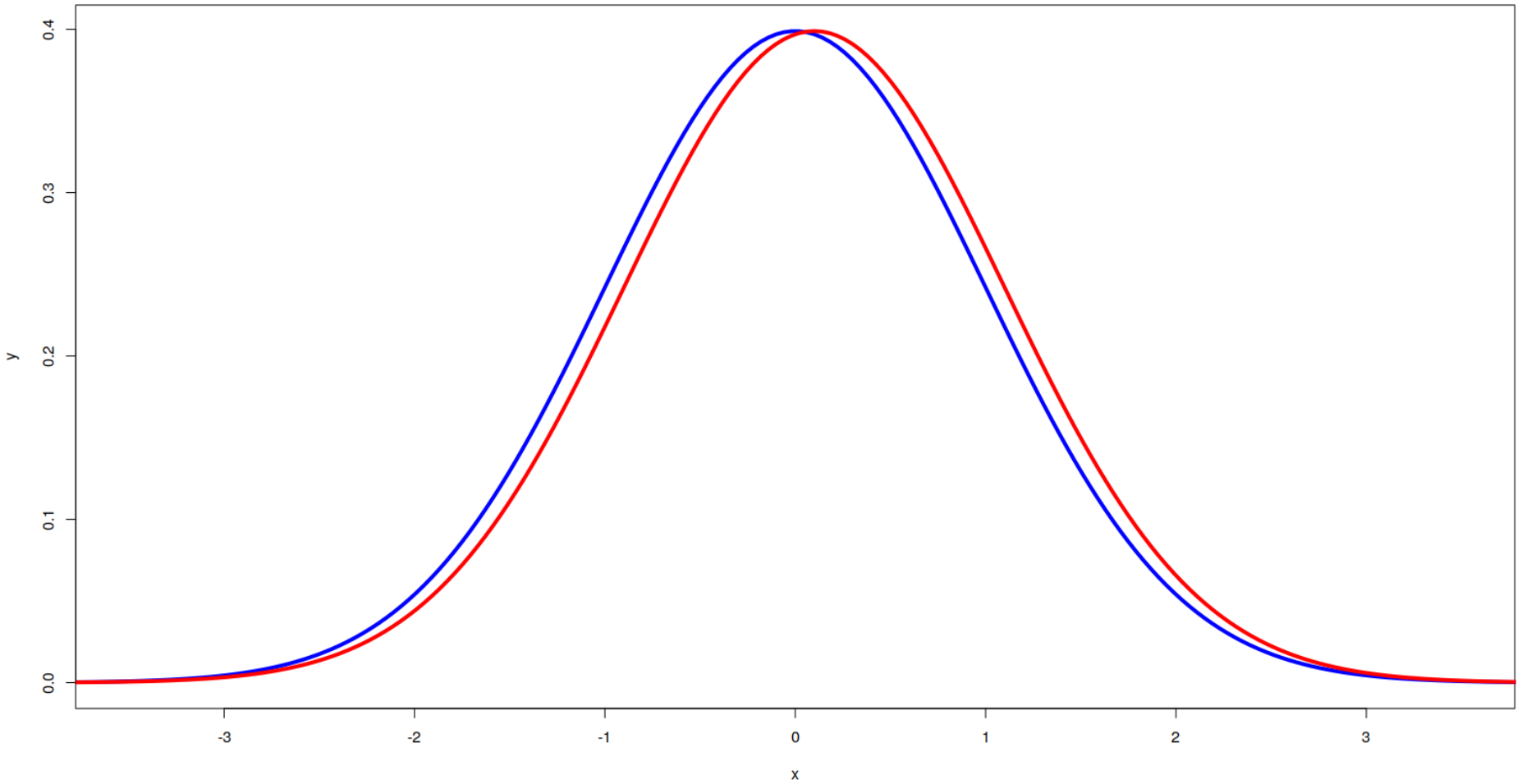
Different means



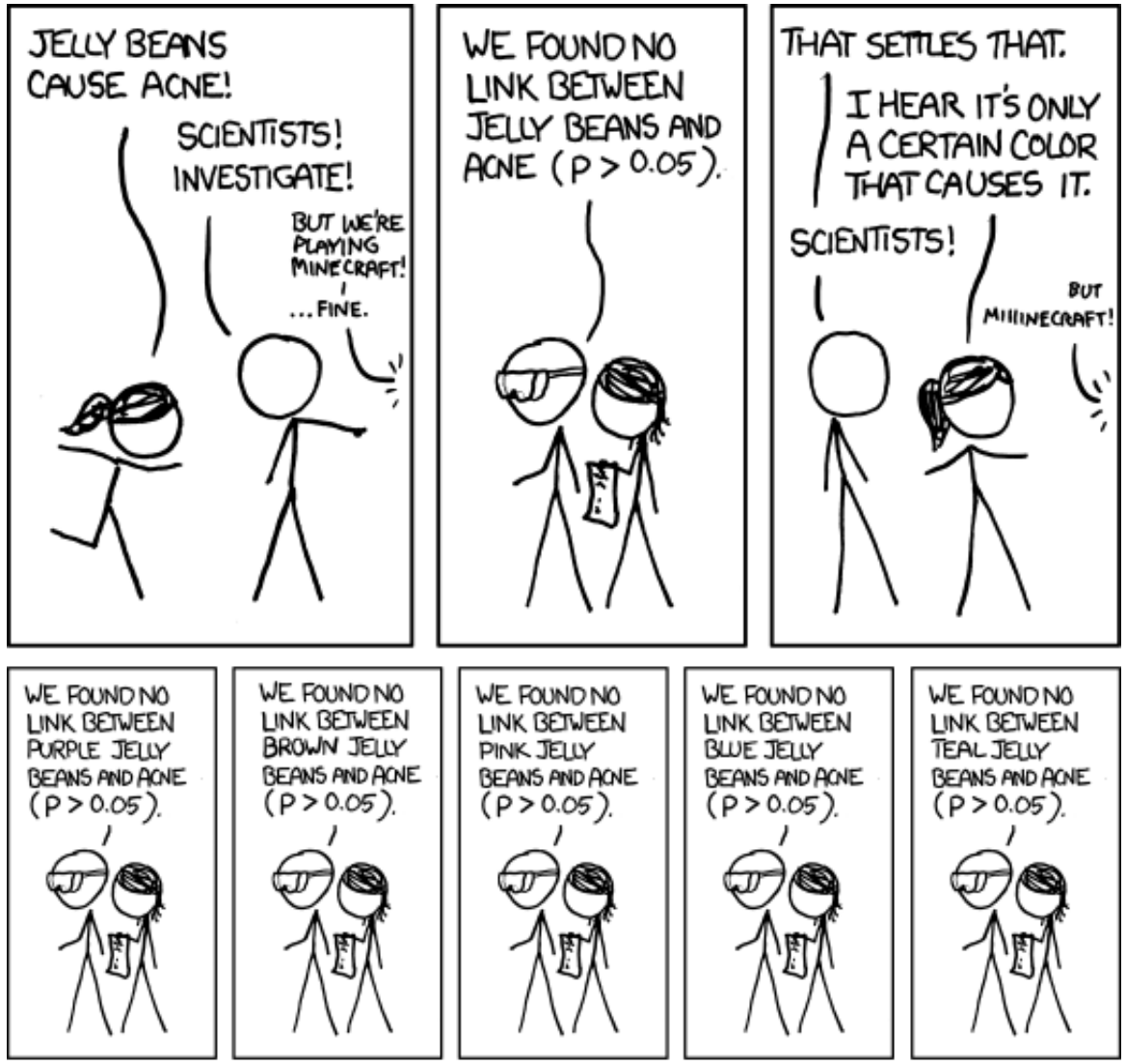
Maybe? different means

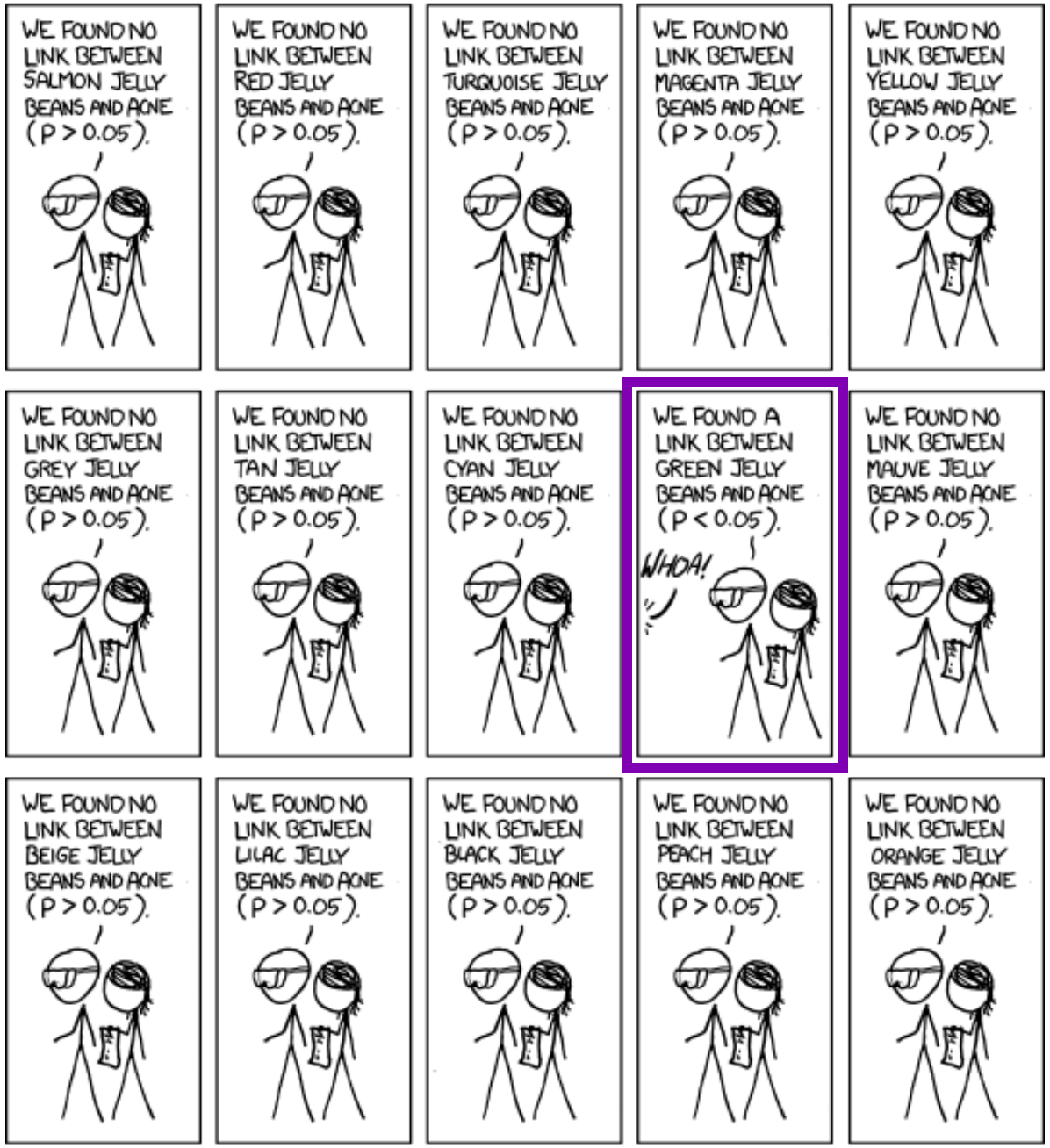


Likely true mean is the same

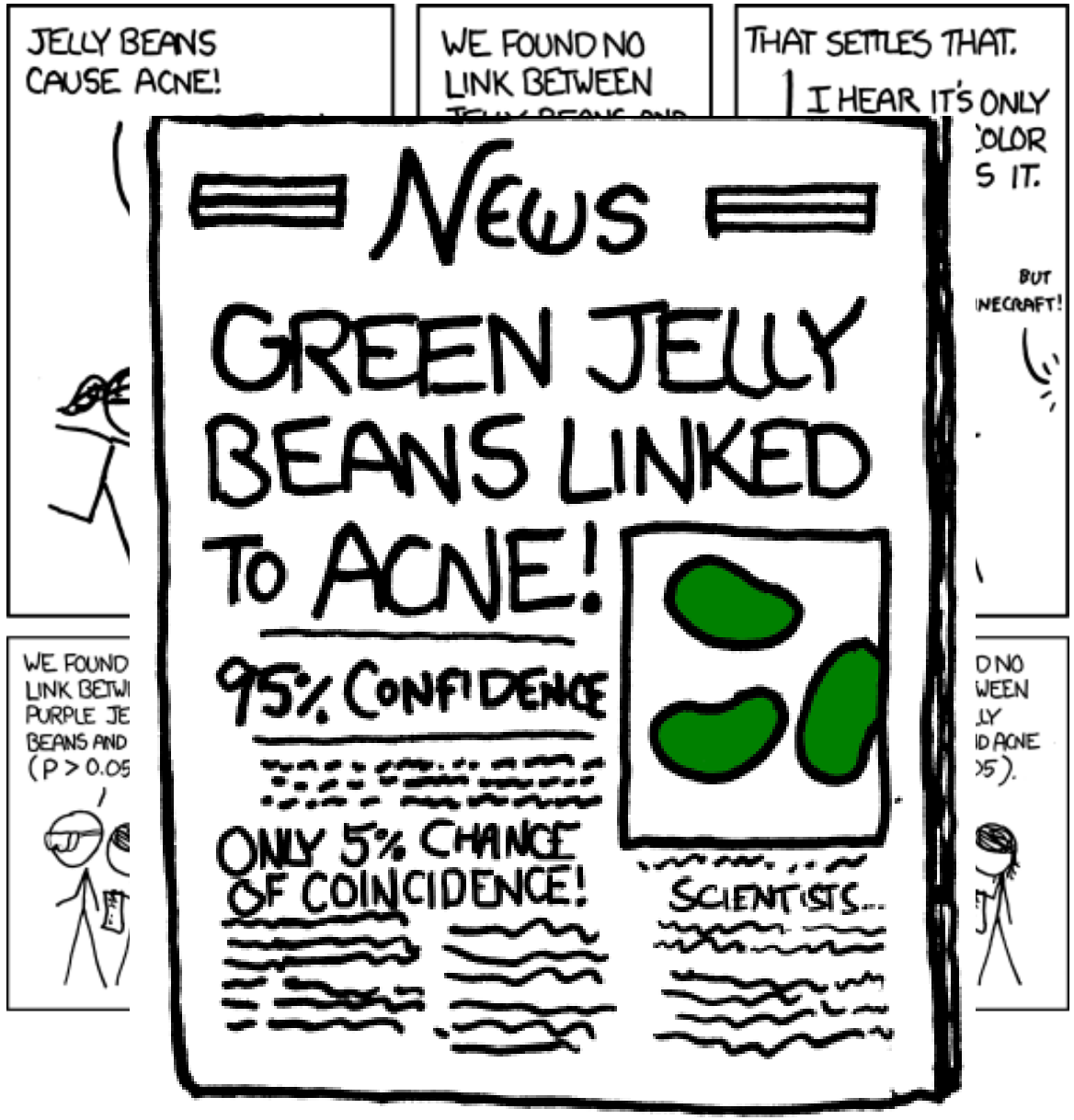


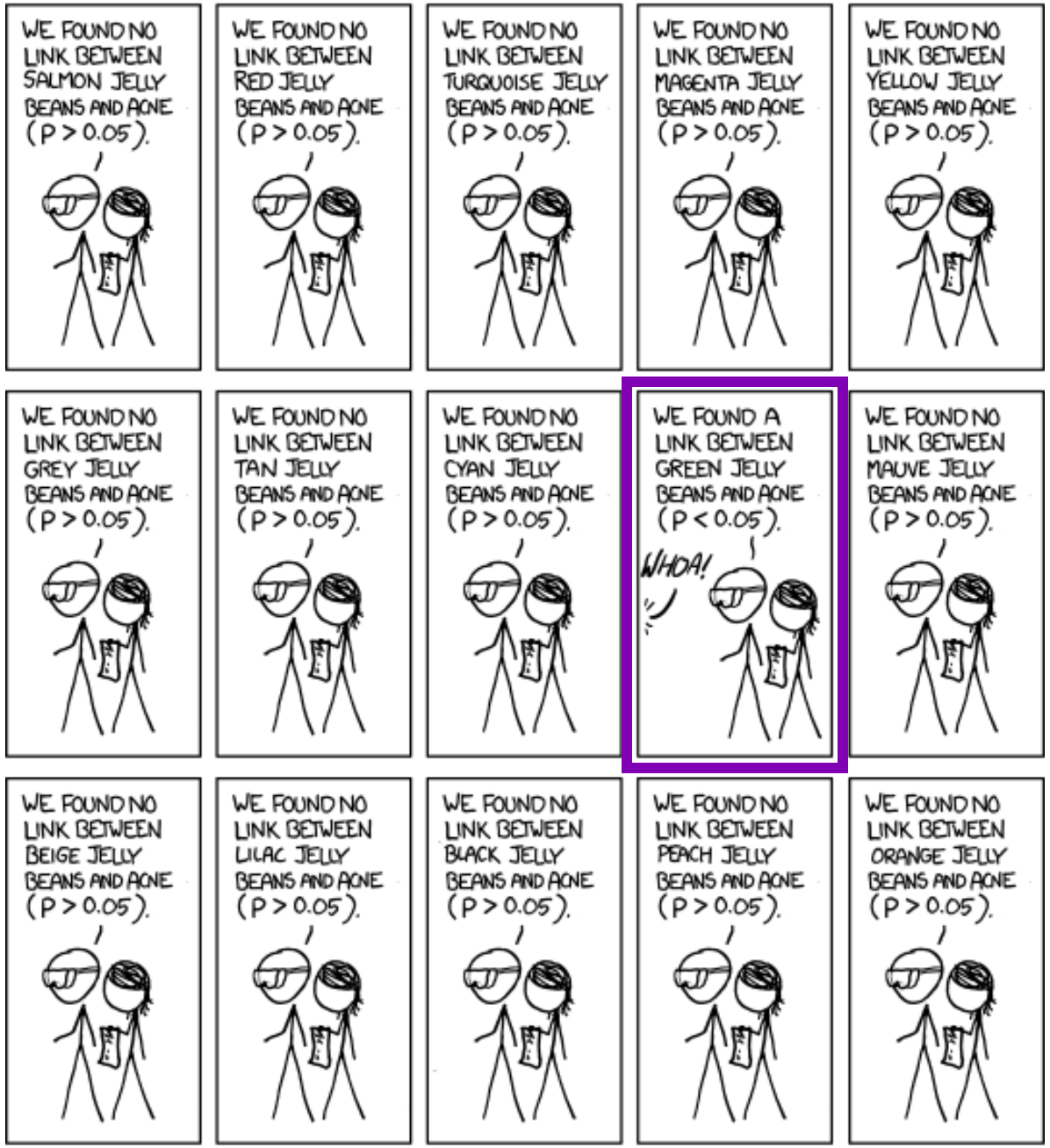
Statistics sometimes concludes incorrectly that two means are different when they are really the same.





https://imgs.xkcd.com/comics/significant.png





https://imgs.xkcd.com/comics/significant.png

I showed participants 4 code samples and asked them what the code would do. I then asked them how confident they were in their answer.

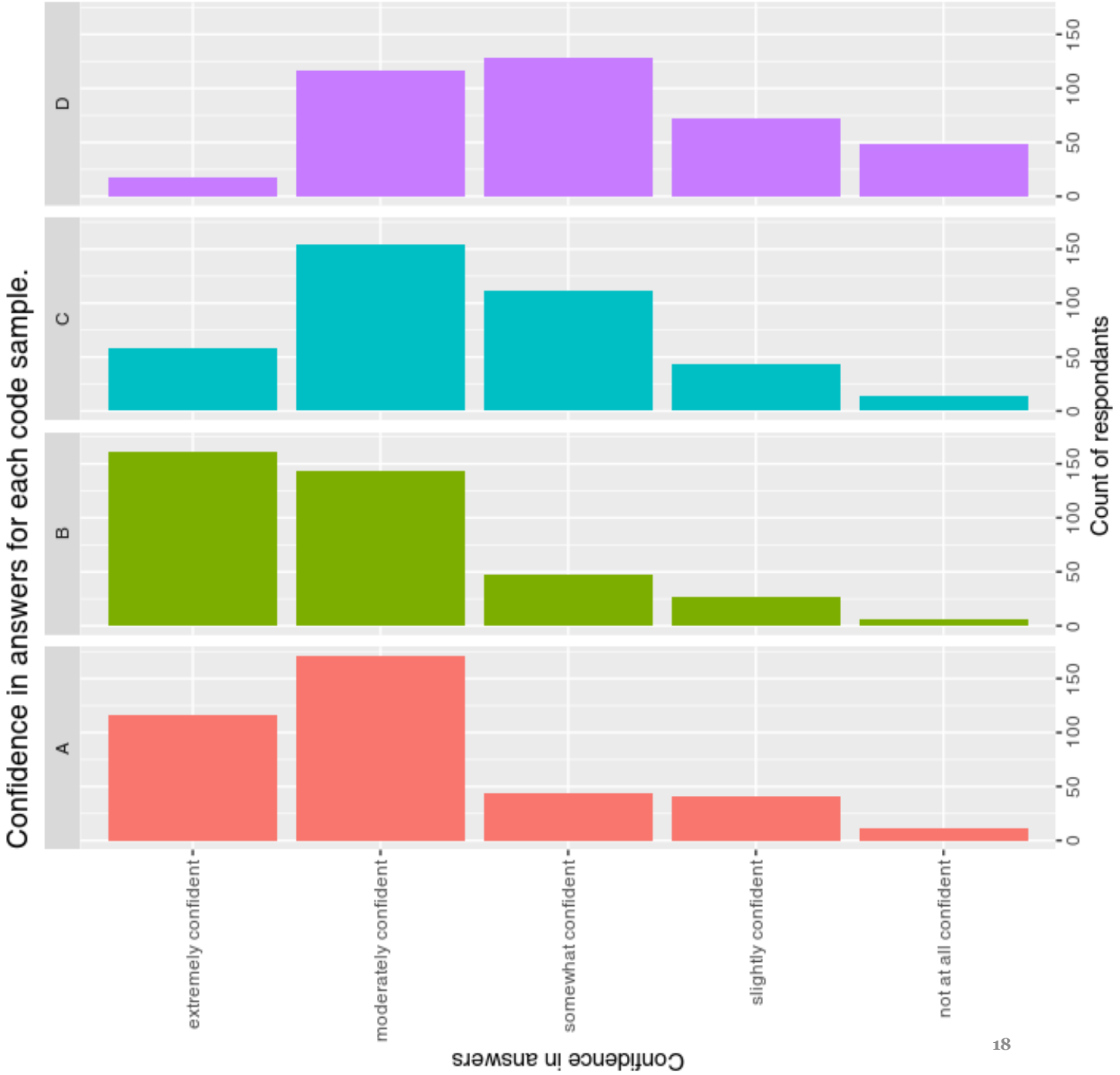
Research Question: Does the code sample shown impact confidence in their answer?

Research Question:
Does the code sample shown impact confidence in their answer?

Within-subjects

Independent:
Which code sample shown

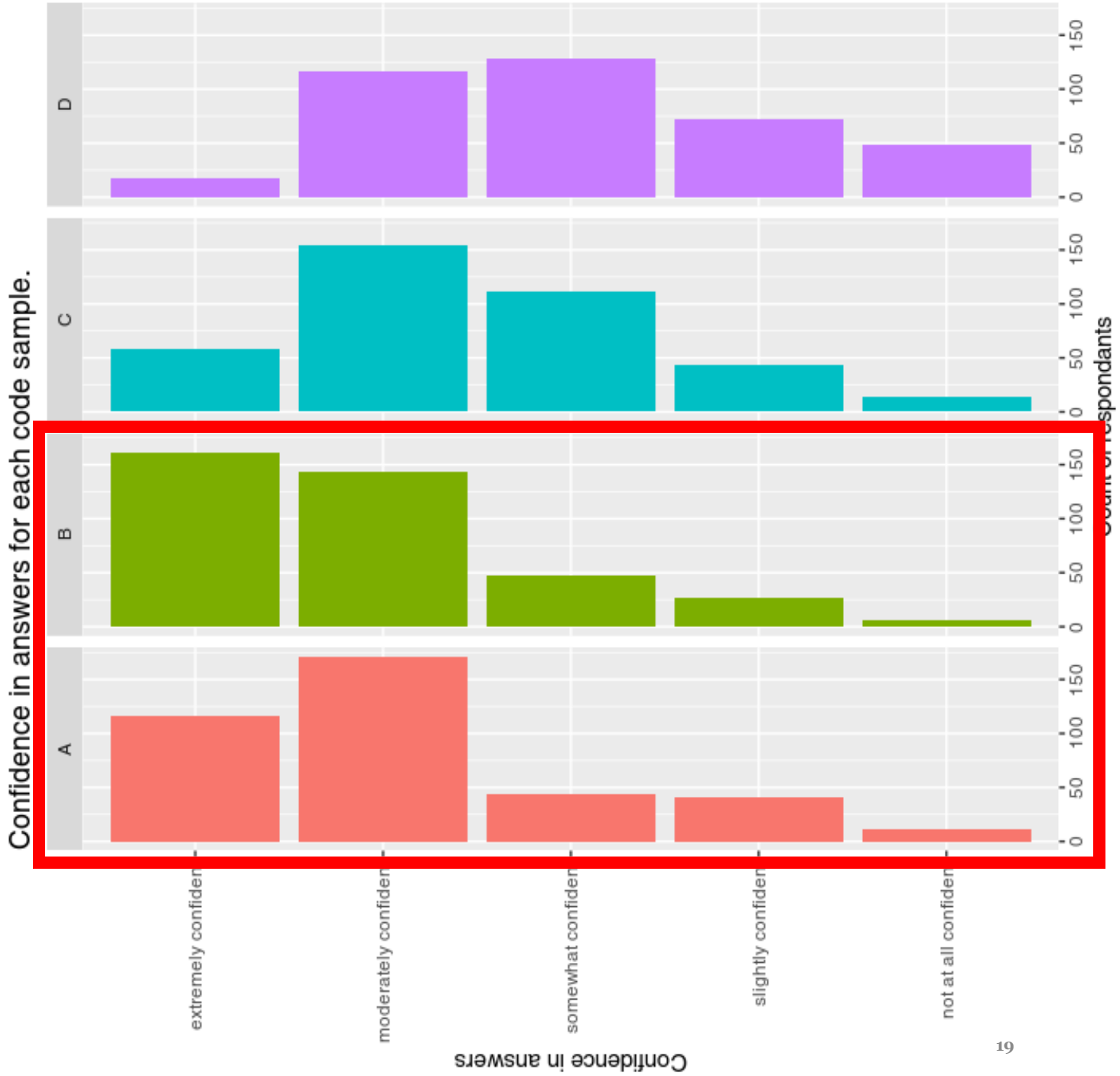
Dependent: Confidence



Problem: My categorical variable (code sample) is not binary, there are 4 levels.

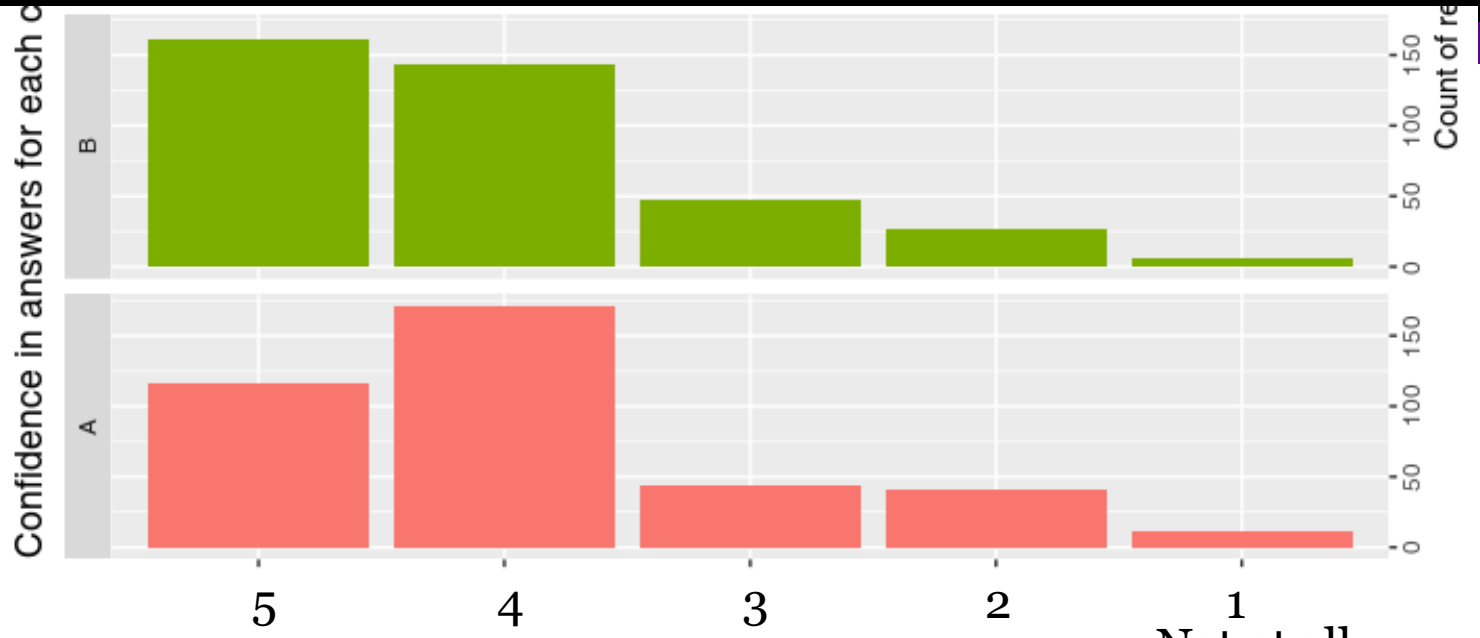
Solution: Run the t-test on each pair. So test A vs B, A vs C, …. C vs D.

Real solution: Use an ANOVA (not covered in this class)



Running the t-test

- This is a "**within** subjects" test where one person gave a confidence answer for both Code Sample A and Code Sample B
  - So we use a **Paired t-test**
- Create two arrays (or Excel columns) one with Code Sample A confidence, the other with Code Sample B confidence
- Two-sided (tailed)
  - For now, just do this. I don't have time to explain.
- Alpha of 0.05
  - p-value needs to be less than 0.05 to show that the two code samples produce different levels of confidence
  - Means that 5% of the time we will get the wrong answer from the statistical test

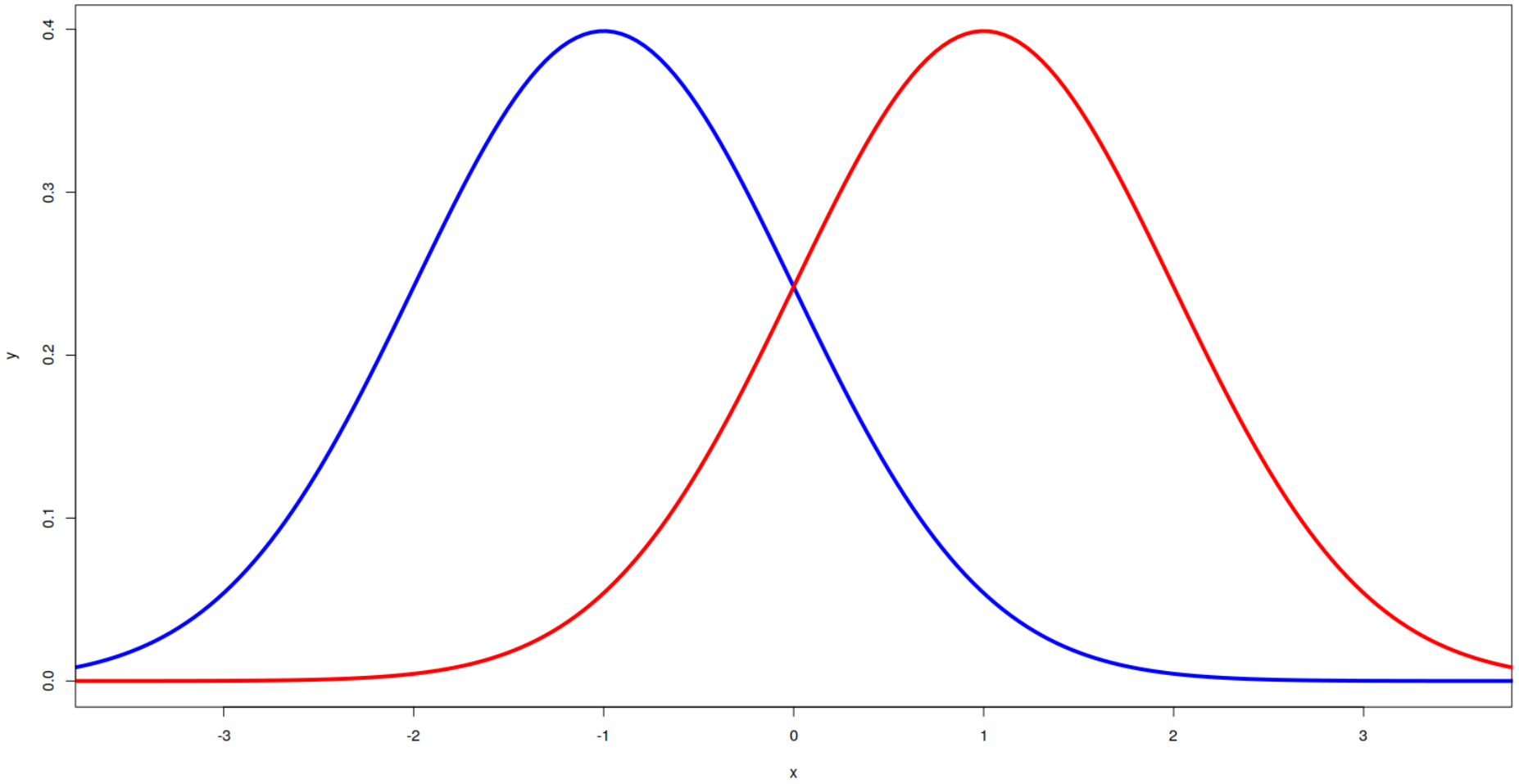



```
> t.test(a.confidence,b.confide

        Paired t-test

data:  a.confidence and b.confidence
t = -5.2699, df = 383, p-value = 2.285e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3218198 -0.1469302
sample estimates:
mean of the differences
                -0.234375
```

Different means, small difference



I ran a survey to learn about software update behaviors.

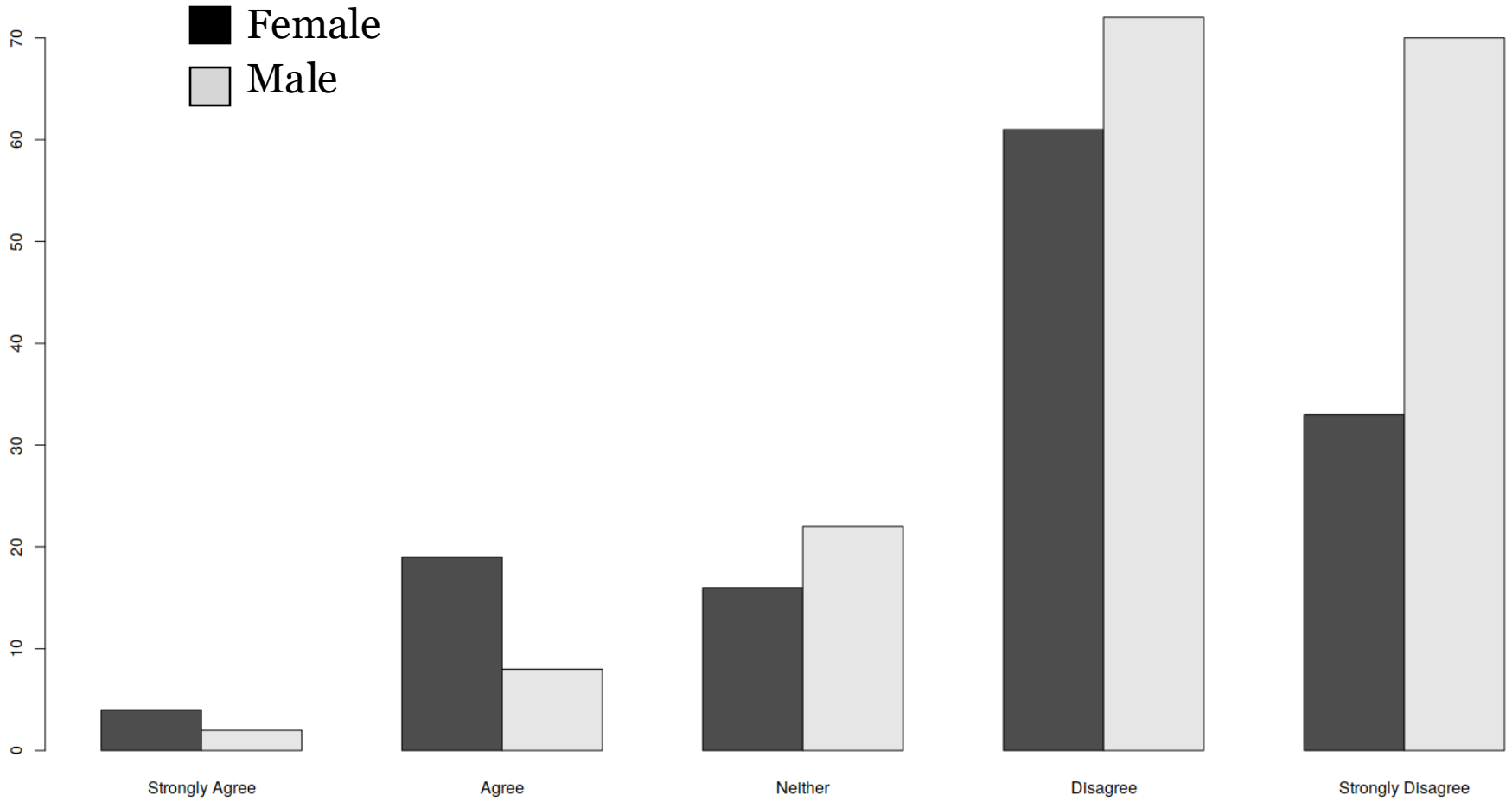
Research Question: Do women and men feel like they ask others for technical help with different frequency?

Research Question:
Do women and men feel like they ask others for help with different frequency?

Between-subjects

Independent:
Gender

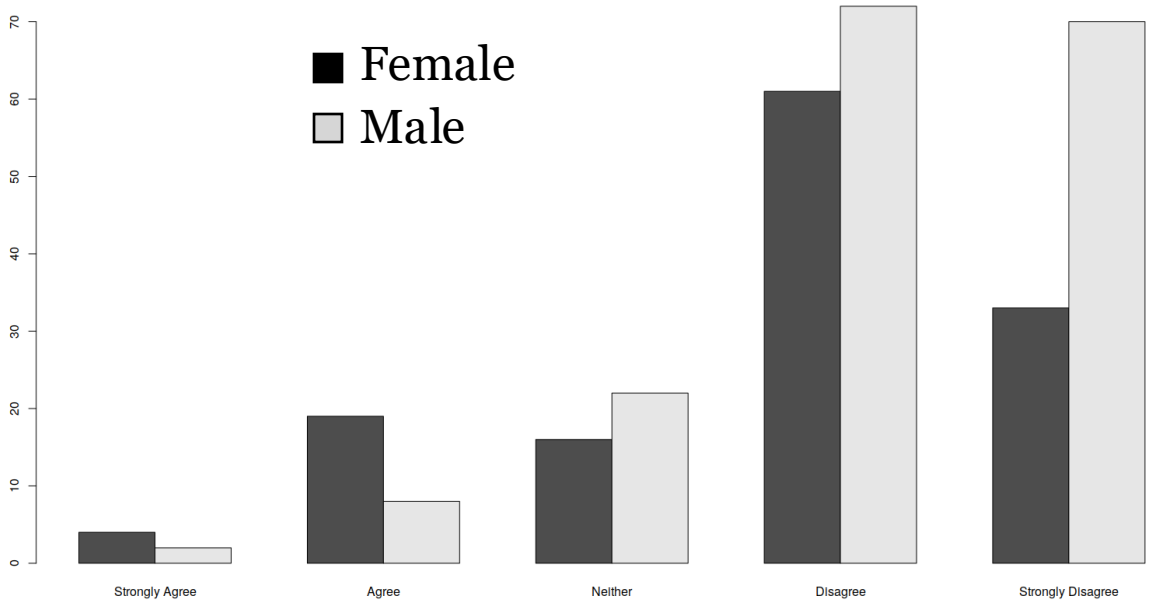
Dependent:
Agreement



I often ask others for help with technical questions

Running the t-test

- This is a "**between** subjects" test where each person gave only one answer
  - So we use a **normal t-test** (not paired)
- Create two arrays one with women's responses, one with men's
- Two-sided (tailed)
  - For now, just do this. I don't have time to explain.
- Alpha of 0.05
  - p-value needs to be less than 0.05 to show that the two genders produce different levels of confidence
  - This choice means that 5% of the time we will get the wrong answer from the statistical test




```
> t.test(as.numeric(d$i_ask_others_for_he
```

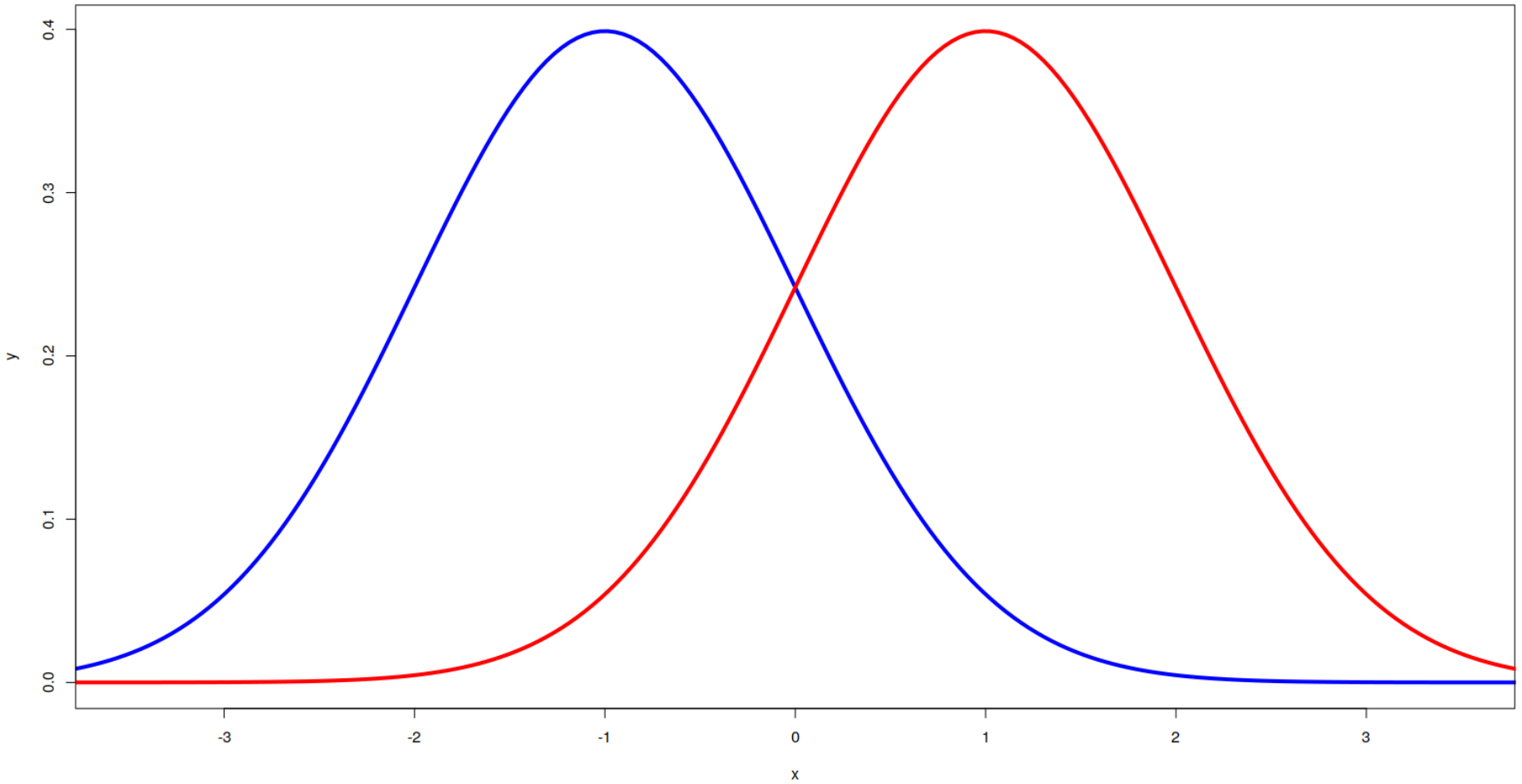
1

5

```
        Welch Two Sample t-test

data:  as.numeric(d$i_ask_others_for_help[d$gender == "Female"]) and
t = -3.4481, df = 253.99, p-value = 0.0006606
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.6245978 -0.1704934
sample estimates:
mean of x mean of y
 3.751880  4.149425
```

Different means, but difference is not large

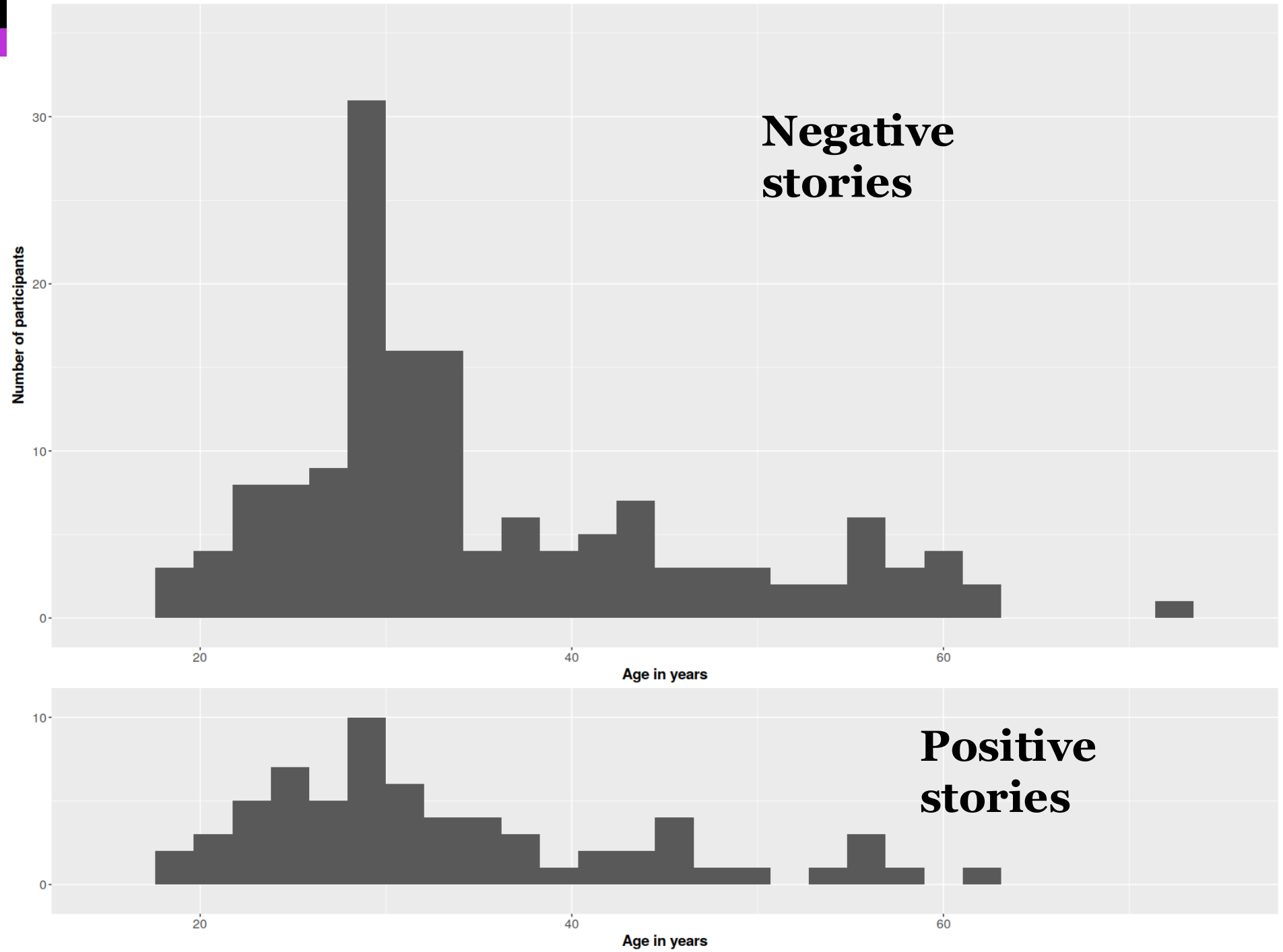


I asked participants to tell me a story about a prior software update.

Research Question: Are people who relate positive stories older or younger?

Research Question:
Are people who relate positive stories older or younger?

Between-subjects
Dependent:
- Age
- Numerical
Independent:
- Negative or Positive
- Binary

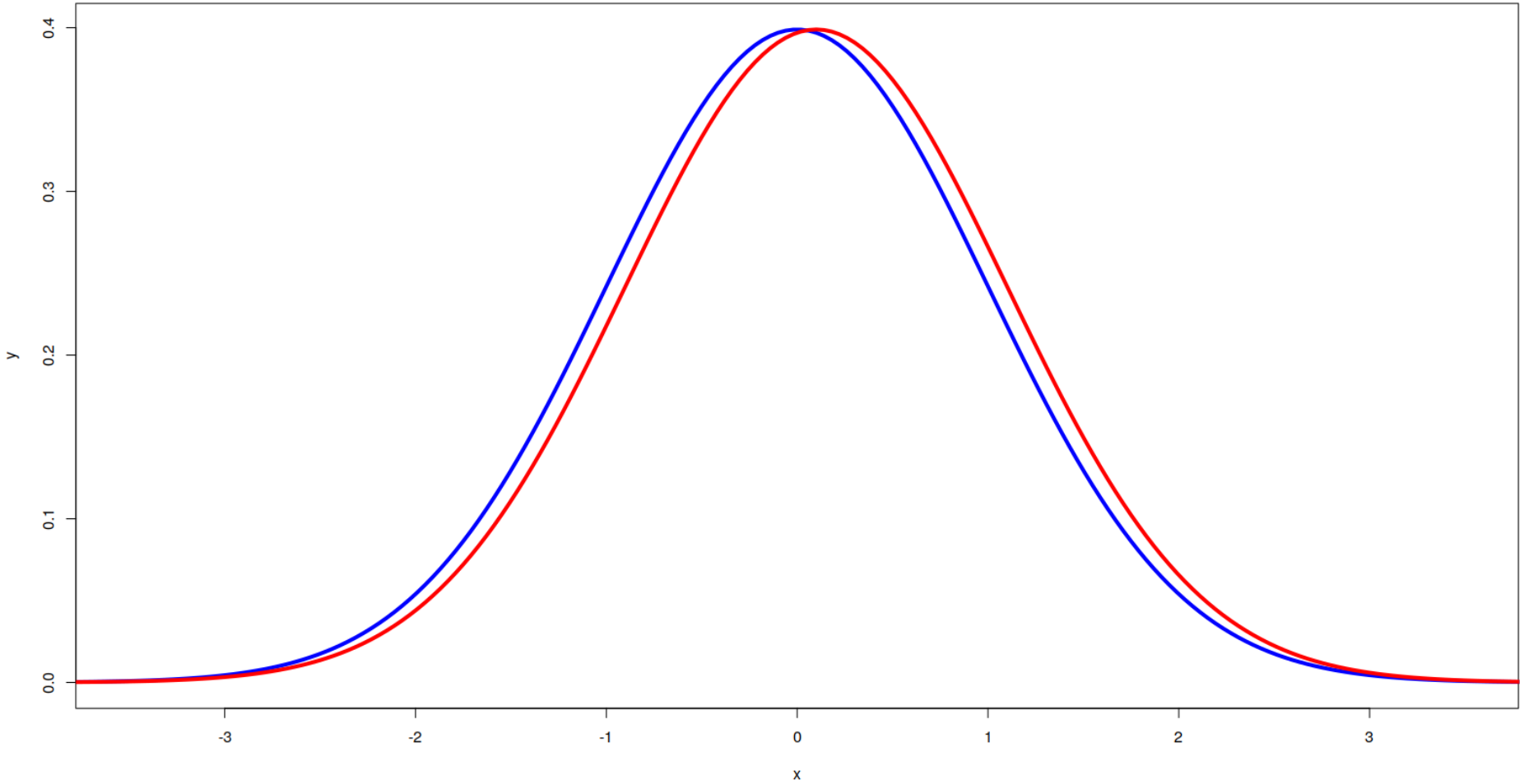


```
> t.test(s_neg$age, s_pos$age)

        Welch Two Sample t-test

data:  s_neg$age and s_pos$age
t = 0.75677, df = 123.07, p-value = 0.4506
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.063833  4.618658
sample estimates:
mean of x mean of y
 35.42667  34.14925
```

No statistical difference between means



Questions