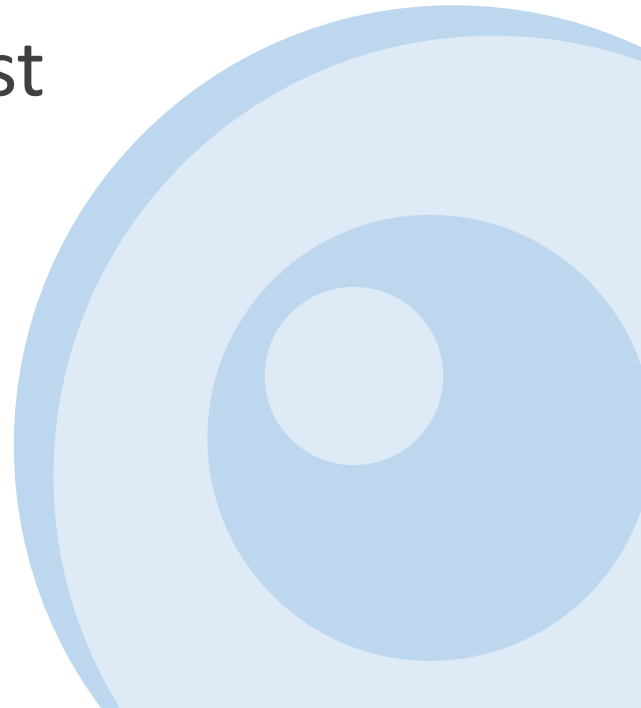


PSYC 193R:

Data Analysis & Visualization with R

Lecture 10: Independent-samples t-test

Jonas S.-H. Lau [silau -et- ucsd.edu]
psyc193r.ucsd.edu

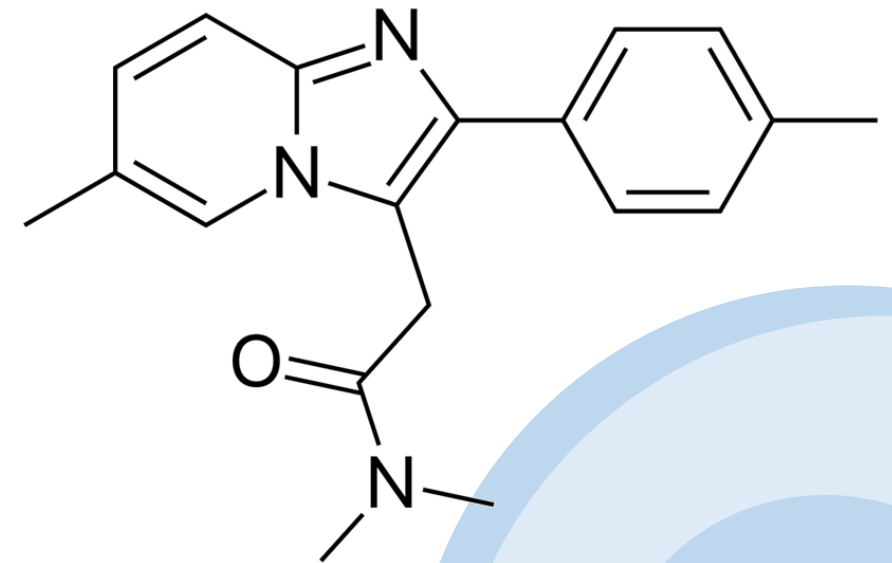


Sleeping Pills on Sleep Quality

People with insomnia often find it hard to fall asleep. Sleeping pills are often their "solution". A group of scientists wanted to test the effectiveness of sleeping pills (non-benzodiazepine hypnotics, *a.k.a.* Z drugs) beyond placebo effects.

A group of clinically diagnosed insomnia patients were invited to take part in a 7-day test. They were put into nice hotel rooms while their brain waves, heart and breathing rates were monitored during sleep. They were randomly assigned to either the vitamin (placebo) or Z drugs groups. Each night, they were given the drug 15 minutes before scheduled bed time.

The time each participant took to fall asleep (polysomnographic sleep latency), and their self-reported quality of the sleep in the morning, were recorded.



Sleeping Pills on Sleep Quality: The Design

- What was the research question?
- What was the target population?
- What were the IV(s) and DV(s)?
- How were the IV(s) and DV(s) operationalized?
(also, scale of measurement?)
- Was it a between / within-subjects design?
- What are the expected results?
If the manipulation does not work:
If the manipulation works:

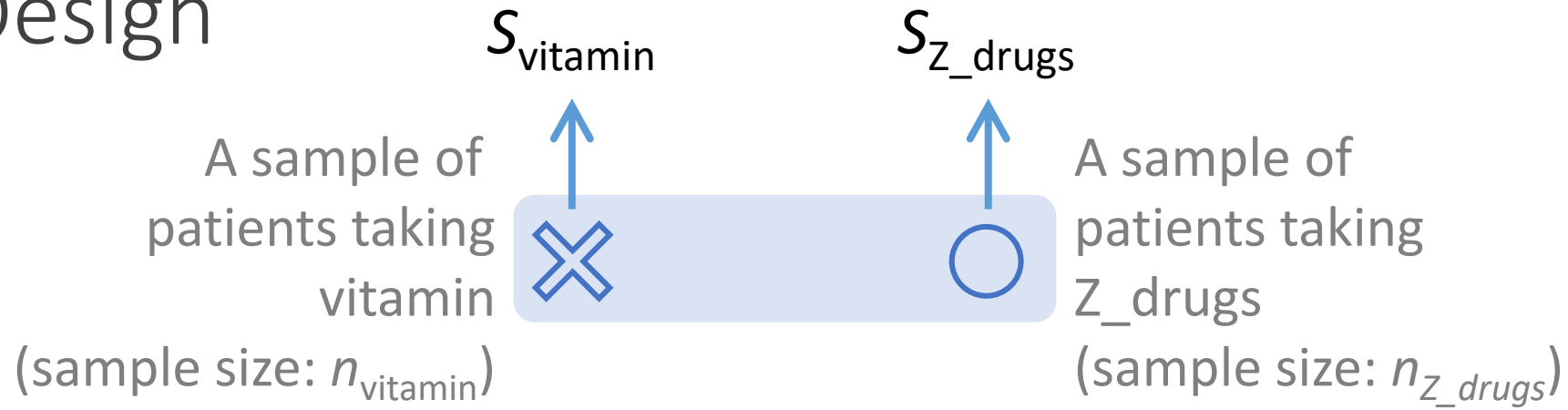


Why between-subjects design?

- Carry over effect: drugs might have an effect lasting longer than a day
- Expectation: Participant may know if you gave them different medications on alternative days

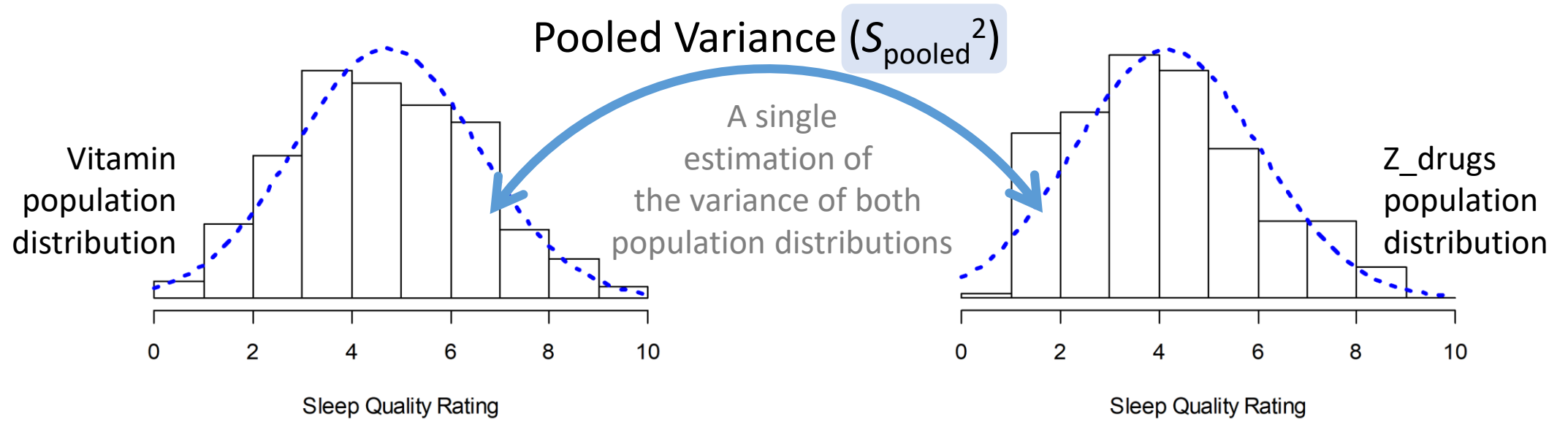


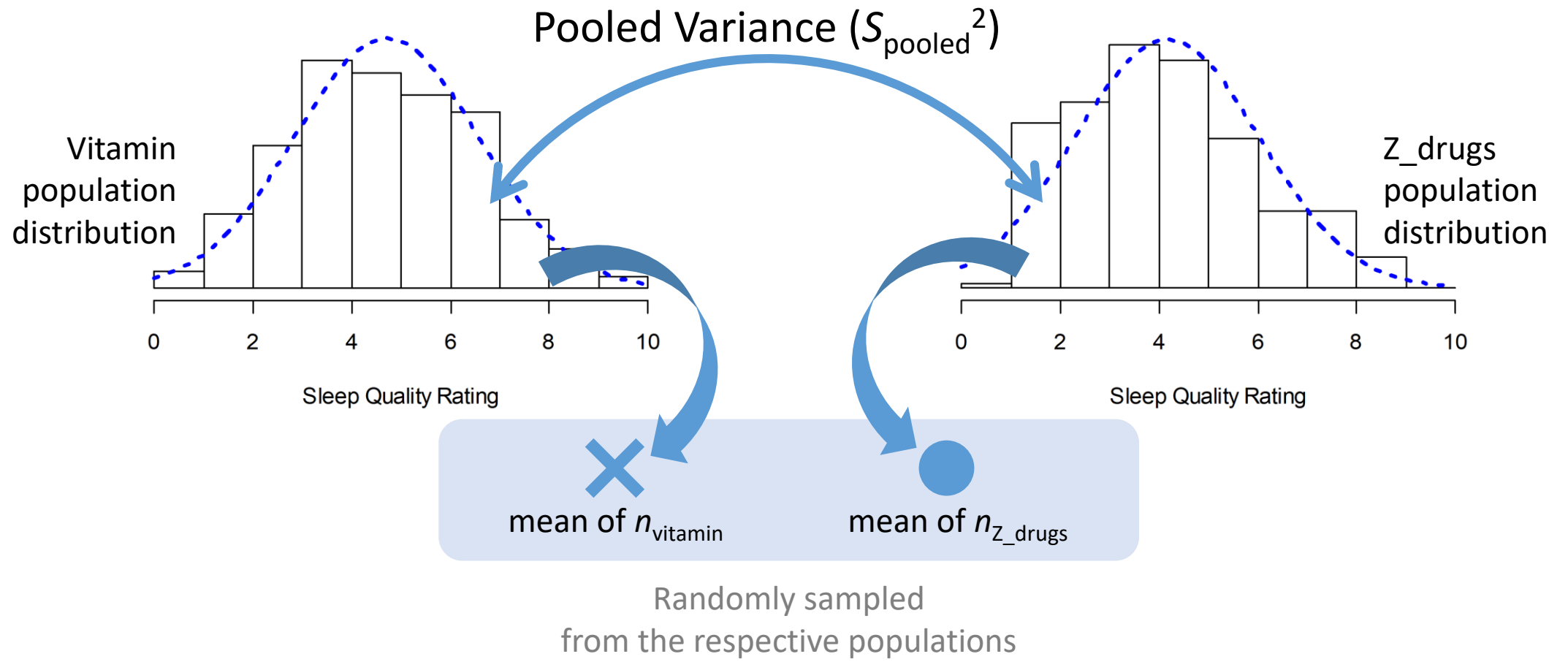
The Design

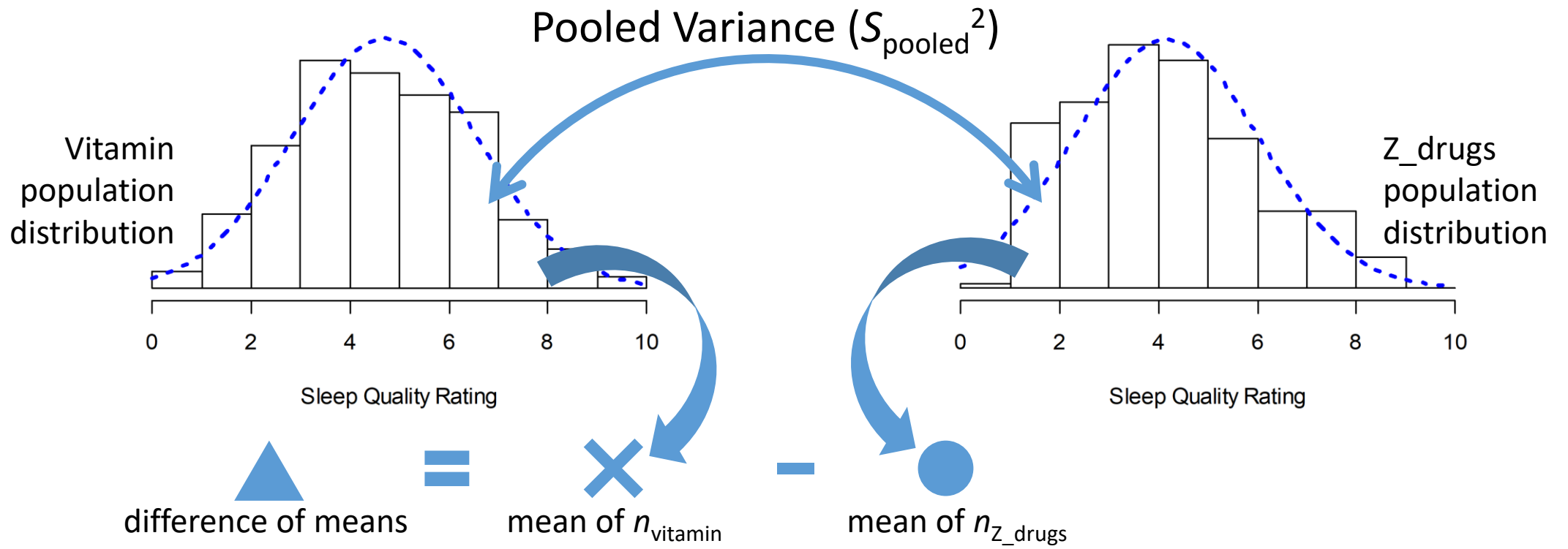


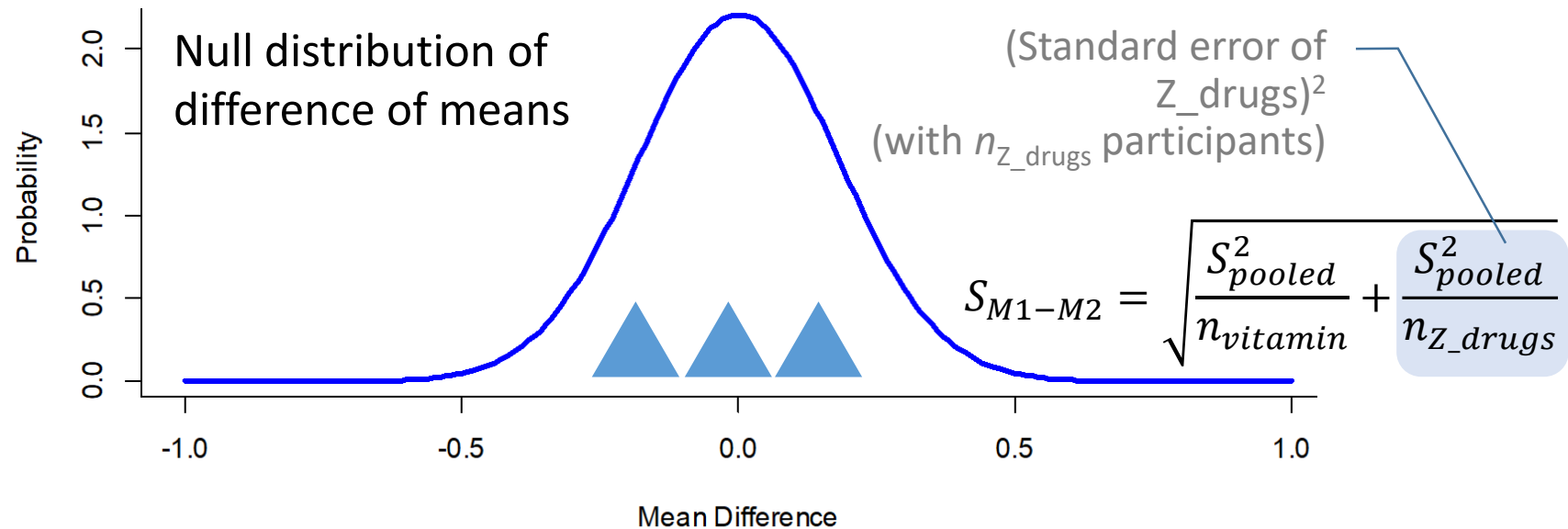
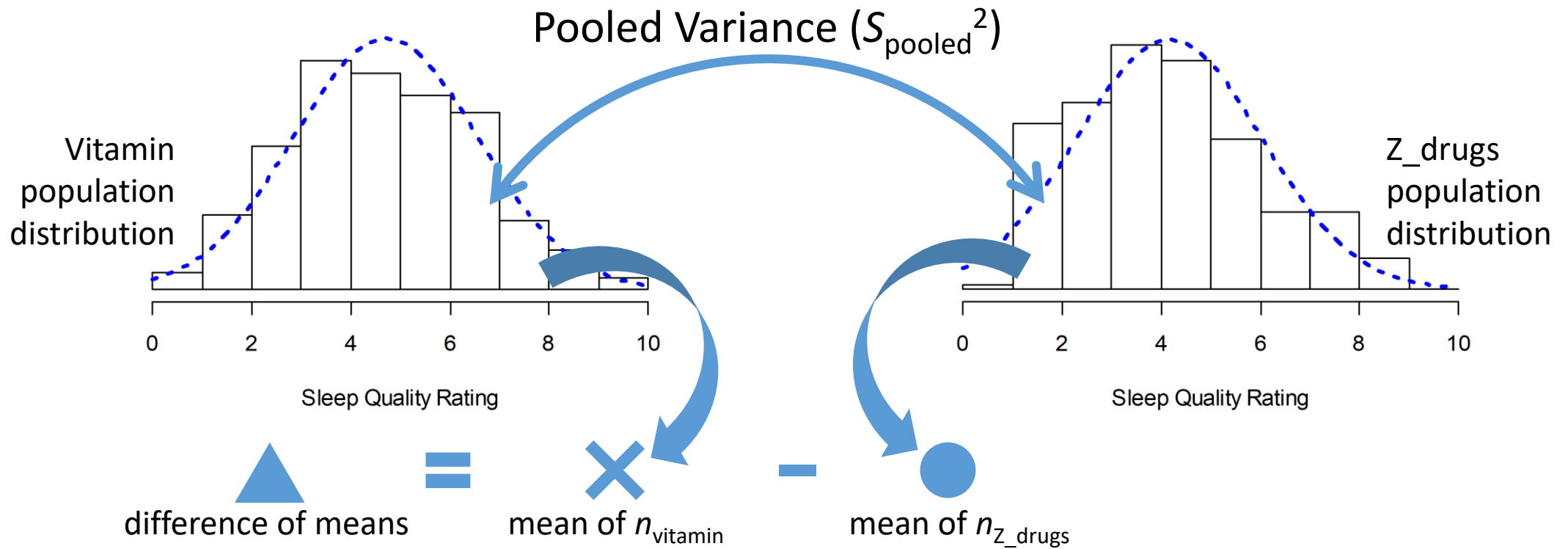
Research Question

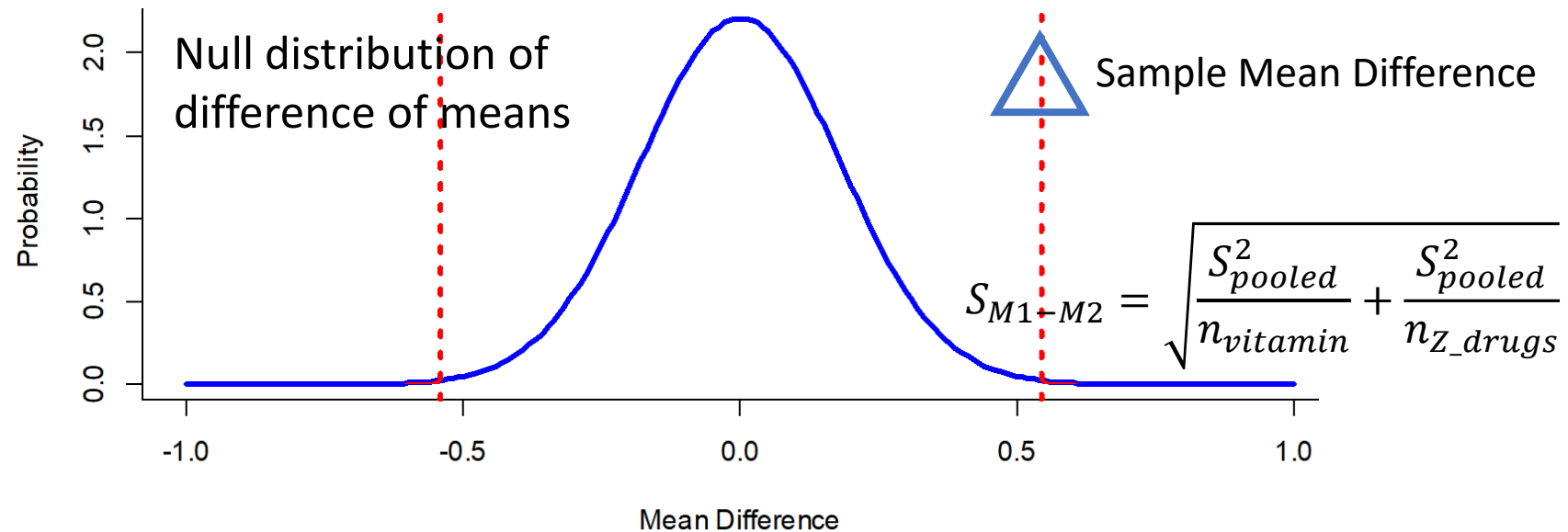
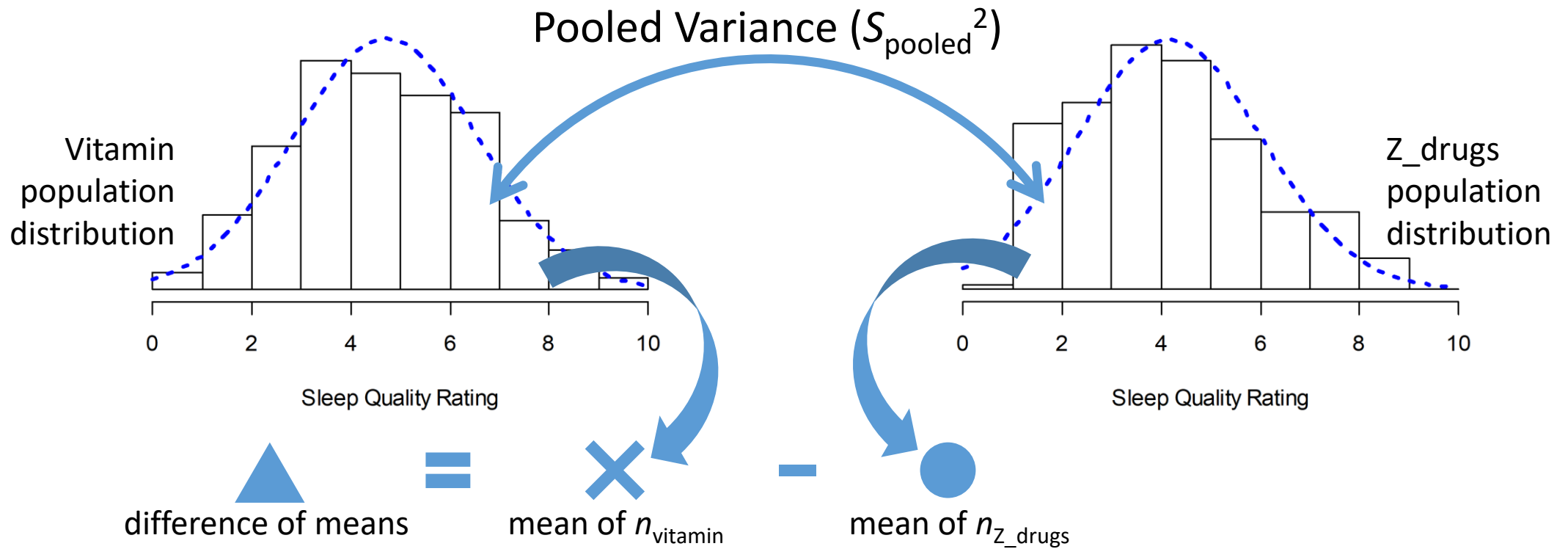
whether the difference between the means
is reliably different from zero







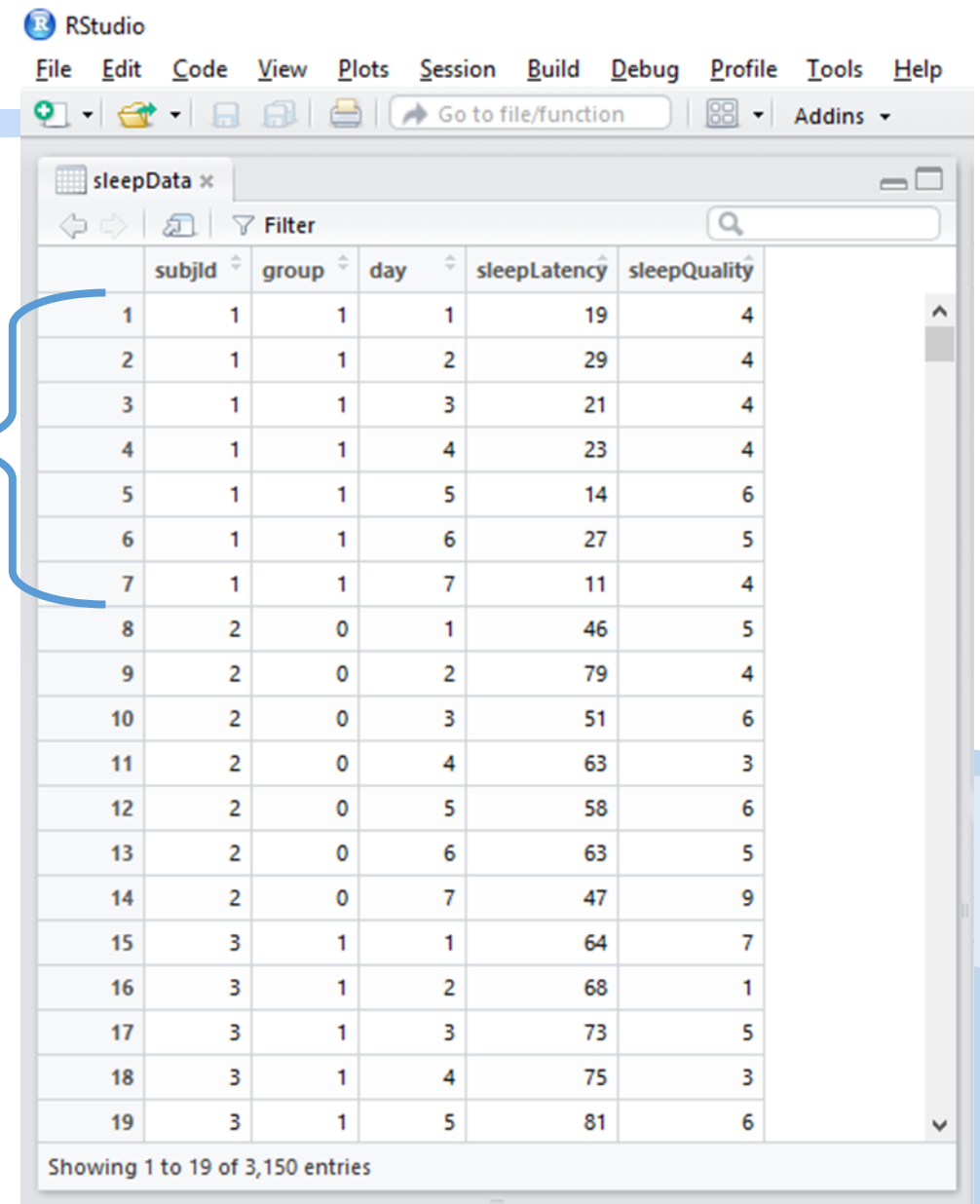




Data format

The data set has a "long" format: Each row represents an entry / a day for each participant

- `subjId`: participant identification
- `group`: 0: vitamin; 1: Z_drugs
- `day`: The day they were in the program
- `sleepLatency`: Time needed to fall asleep (in minutes)
- `sleepQuality`: Self-report quality of that night's sleep



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

sleepData x

Filter

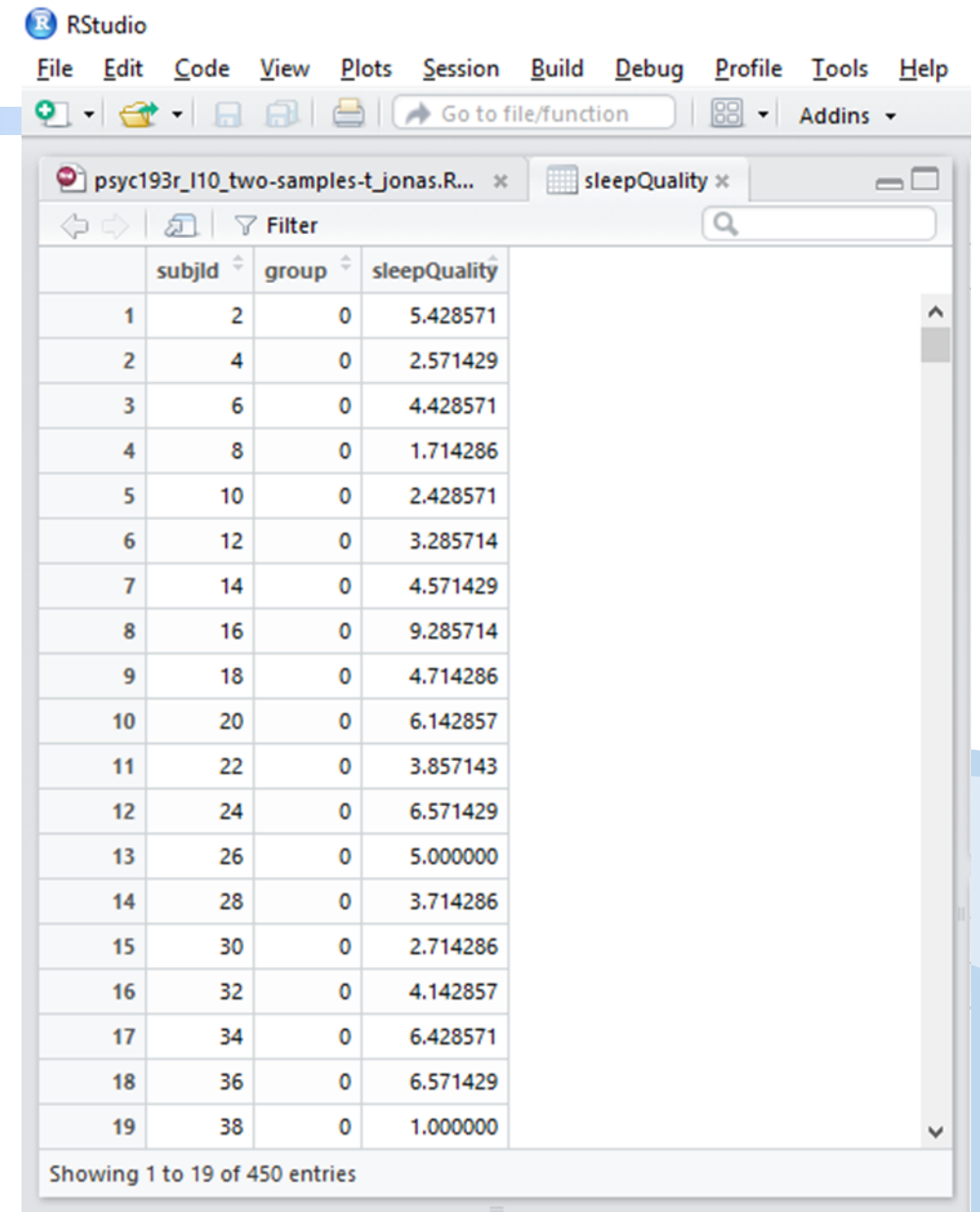
	subjId	group	day	sleepLatency	sleepQuality
1	1	1	1	19	4
2	1	1	2	29	4
3	1	1	3	21	4
4	1	1	4	23	4
5	1	1	5	14	6
6	1	1	6	27	5
7	1	1	7	11	4
8	2	0	1	46	5
9	2	0	2	79	4
10	2	0	3	51	6
11	2	0	4	63	3
12	2	0	5	58	6
13	2	0	6	63	5
14	2	0	7	47	9
15	3	1	1	64	7
16	3	1	2	68	1
17	3	1	3	73	5
18	3	1	4	75	3
19	3	1	5	81	6

Showing 1 to 19 of 3,150 entries

Ideal Data Frame for Analysis

We want to have a data frame with 3 variables:

- `subjId`
- `group`
- `sleepQuality`: mean self-reported sleep quality scores for each participant in the 7 days



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

psyc193r_I10_two-samples-t_jonas.R... sleepQuality

Filter

	subjId	group	sleepQuality
1	2	0	5.428571
2	4	0	2.571429
3	6	0	4.428571
4	8	0	1.714286
5	10	0	2.428571
6	12	0	3.285714
7	14	0	4.571429
8	16	0	9.285714
9	18	0	4.714286
10	20	0	6.142857
11	22	0	3.857143
12	24	0	6.571429
13	26	0	5.000000
14	28	0	3.714286
15	30	0	2.714286
16	32	0	4.142857
17	34	0	6.428571
18	36	0	6.571429
19	38	0	1.000000

Showing 1 to 19 of 450 entries

Import Data

- Download the data file:
https://psyc193r.ucsd.edu/data/psyc193r_sleep_data.csv
- Tell R where your data file is:
Session → Set Working Directory → Choose Directory
`setwd()`
- Check the *comma separated values* (csv) file with MS Excel to ensure correct format
- Import the file into R as a data frame, saving it to “sleepData” variable
`... = read.csv("file_name.csv")`

Data Formatting / Processing / Cleaning

- Factorize subjId

```
sleepData$subjId = ...
```

- Factorize group, and label the two levels

```
sleepData$group = factor(sleepData$group, labels =  
  c("...", "..."))
```

- Coerce sleepQuality to a numeric variable

```
sleepData$sleepQuality = ...
```

- Create a new data frame that aggregates the mean “sleep quality” for each participant, saving it to a new variable "sleep"

```
sleep = aggregate(sleepQuality ~ ____ * ____ , data = sleepData,  
  FUN = ____ )
```

Summary statistics

- Use aggregate to find the descriptive statistics for each group

- Means

```
aggregate(sleepQuality ~ ___, data = ___, FUN = ___)
```

- Standard Deviations

```
aggregate(sleepQuality ~ ___, data = ___, FUN = ___)
```

- Numbers of participants

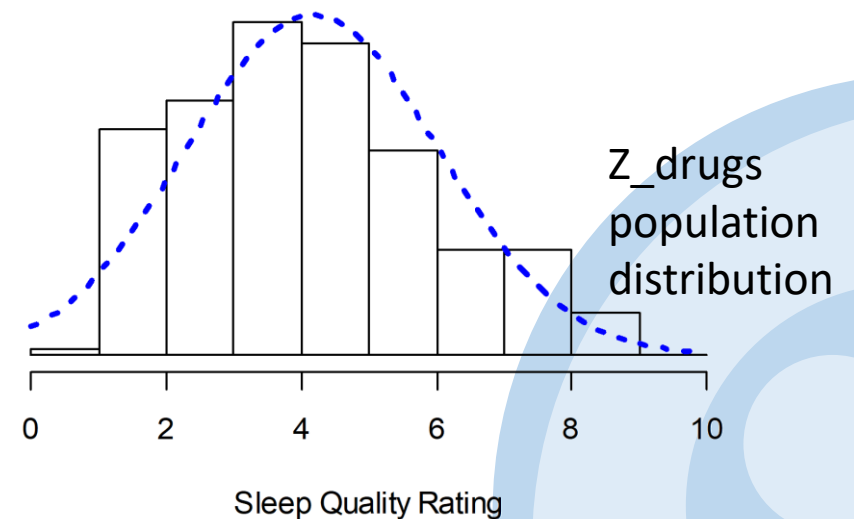
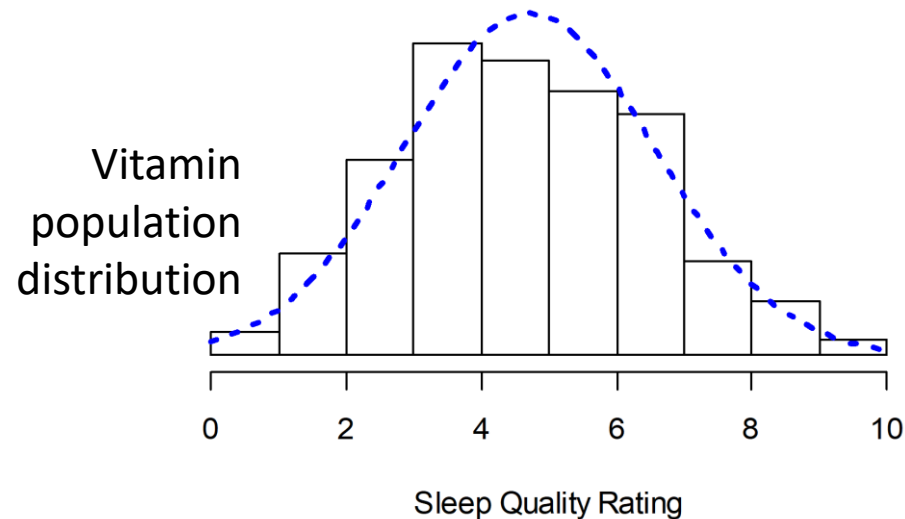
```
aggregate(sleepQuality ~ ___, data = ___, FUN = ___)
```



Checking assumptions

- Normality Assumption
Plot a histogram for each group

```
hist(sleep[ sleep$__ == "__", ]$sleepQuality)  
hist(sleep[ sleep$__ == "__", ]$sleepQuality)
```

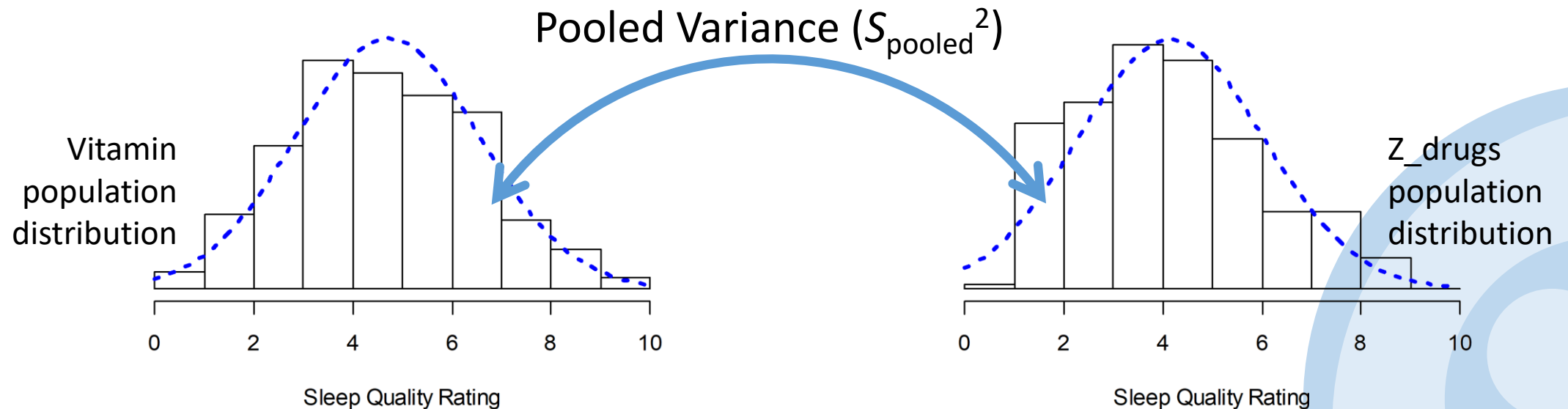


Checking assumptions

- Normality Assumption
Plot a histogram for each group

```
hist(sleep[ sleep$__ == "__", ]$sleepQuality)  
hist(sleep[ sleep$__ == "__", ]$sleepQuality)
```

- Equal variance assumption



Checking assumptions

- Normality Assumption

Plot a histogram for each group

```
hist(sleep[ sleep$___ == "___", ]$sleepQuality)
hist(sleep[ sleep$___ == "___", ]$sleepQuality)
```

- Equal variance assumption [$F(244,244) = 0.93$, $p > 0.05$]

```
var.test(sleep[sleep$group == "Placebo",]$sleepQuality,
        sleep[sleep$group == "Z_drugs",]$sleepQuality)
```

F test to compare two variances

data: Placebo\$sleepQuality and Z_drugs\$sleepQuality

$F = 0.93151$, num df = 224, denom df = 224, p-value = 0.5959

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.716396 1.211229

sample estimates:

ratio of variances

0.9315148

`t.test()` for independent samples t test

```
t.test( dv ~ iv, data = __ , mu = 0, var.equal = TRUE, paired = FALSE)
```



Hypothesis Testing – Step-by-step

- Step 0: Gather all the information you need

- Check assumptions
- Data cleaning

- Step 1: Set up the hypotheses

- H_0 (null hypothesis)
- H_1 (Research hypothesis)

- Step 2: Calculate the test statistic

- Step 3: Find the probability associated with the test statistic

(i.e., the probability of getting such a score or a more extreme score x 2, given H_0 is true)

- Step 4: Decision & Conclusion
(compare p against α)

$\mu_{\text{baseline_difference}} = 0$ (theoretical)

$M_{\text{vitamin}} ; M_{Z_drugs} ; S_{\text{vitamin}} ; S_{Z_drugs}$

$n_{\text{vitamin}} ; n_{Z_drugs} ; df = n_1 + n_2 - 2$

kurtosis of population ? $n > 30$ ✓

$H_0: \mu_{Z_drugs} - \mu_{\text{vitamin}} = 0$

$H_1: \mu_{Z_drugs} - \mu_{\text{vitamin}} \neq 0$

See `t.test()` output

See `t.test()` output

Report baseline & sample means, standard deviations, sample sizes, sample mean difference, test statistics & p-value, decision/ conclusion/ notes

Reliability Measure: Confidence Interval

When you repeat the same experiment (with the same condition and n), for ___% of the time, the sample mean difference will lie between
[__ , __]

$$(M_1 - M_2) \pm t_{\frac{\alpha}{2}, df} \times S_{M1-M2}$$

Or, obtain in `t.test()` output



Effect Size

The effect size tells us how big the average difference between the two conditions is, in standard deviation

$$Cohen's\ d = \frac{|(\mu_{Z_drugs} - \mu_{vitamin}) - \mu_{baseline_difference}|}{\sqrt{S_{pooled}^2}}$$

Install the effsize package:

```
install.packages("effsize")
```

Load the package, run the analysis

```
library("effsize")  
cohen.d( ___ ~ ___, data = ___ , pooled = TRUE, paired = FALSE)
```

Power analysis with pwr package

- Load the package each time after restarting R, if you need the package

```
library("pwr")
```

- Usage

```
?pwr.t.test
```

```
pwr.t.test(n = NULL, d = NULL, sig.level = 0.05,  
power = NULL, type = c("two.sample", "one.sample", "paired"),  
alternative = c("two.sided", "less", "greater"))
```

- Leave the argument that you are looking for as NULL

- E.g.,

```
pwr.t.test(n = 50, d = 0.5, sig.level = 0.05,  
power = NULL, type = "two.sample")
```

n denotes sample sizes for each group