431 Class 14

thomaselove.github.io/431

2021-10-07

Today's R Setup

```
knitr::opts_chunk$set(comment=NA) # as always
options(width = 55) # to fit things on the slides

library(broom)
library(Hmisc) # for smean.cl.boot(), mostly
library(janitor)
library(knitr)
library(magrittr)
library(naniar)
library(tidyverse)

source("data/Love-boost.R") # for bootdif() function

theme_set(theme_bw())
```

Today's Agenda

- Comparing Two Population Means
  - In a Study using Independent Samples
    - T tests (Pooled and Welch) and Bootstrap Approaches
  - In a Study using Matched (Paired) Samples
    - T test and Bootstrap Approaches
- New Examples: Two Studies from the Cleveland Clinic

Paired vs. Independent Samples

Suppose you can afford to measure n = 400 outcome values and want to compare the outcome's mean under exposure A to the outcome's mean under exposure B. Consider these two designs:

1. Select a random sample from the population of interest containing 200 people, each of whom provide you with an outcome under exposure A, and then provide you with an outcome under exposure B.
2. Select a random sample from the population of interest containing 400 people and then randomly assign 200 people to receive exposure A and the remaining 200 people to receive exposure B.

- What are the main differences between the studies?

Paired vs. Independent Samples

Suppose you can afford to measure n = 400 outcome values and want to compare the outcome's mean under exposure A to the outcome's mean under exposure B. Consider these two designs:

1. Select a random sample from the population of interest containing 200 people, each of whom provide you with an outcome under exposure A, and then provide you with an outcome under exposure B.
2. Select a random sample from the population of interest containing 400 people and then randomly assign 200 people to receive exposure A and the remaining 200 people to receive exposure B.

- What are the main differences between the studies?
- We'll call Study 1 a **paired samples** study, since each result under exposure A is matched to the exposure B result from the same person. Calculating paired B - A differences by person makes sense.

Paired vs. Independent Samples

Suppose you can afford to measure n = 400 outcome values and want to compare the outcome's mean under exposure A to the outcome's mean under exposure B. Consider these two designs:

1. Select a random sample from the population of interest containing 200 people, each of whom provide you with an outcome under exposure A, and then provide you with an outcome under exposure B.
2. Select a random sample from the population of interest containing 400 people and then randomly assign 200 people to receive exposure A and the remaining 200 people to receive exposure B.

- What are the main differences between the studies?
- We'll call Study 1 a **paired samples** study, since each result under exposure A is matched to the exposure B result from the same person. Calculating paired B - A differences by person makes sense.
- We'll call Study 2 an **independent samples** study, where there is no pairing or matching of individual observations across exposure groups.

A Study Involving Two Independent Samples

The Supraclavicular Data

These data come from the Cleveland Clinic's Statistical Education Dataset Repository, which is a great source of examples for me, but the data there cannot be used for Project B (just to let you know in advance.)

The Supraclavicular data come from Roberman et al. "Combined Versus Sequential Injection of Mepivacaine and Ropivacaine for Supraclavicular Nerve Blocks". *Reg Anesth Pain Med* 2011; 36: 145-50.

```
supra_raw <- read_csv("data/Supraclavicular.csv") %>%
  clean_names()

dim(supra_raw)
```

```
[1] 103  17
```

Supraclavicular Study Objective (in brief)

This study consisted of 103 patients, aged 18 to 70 years, who were scheduled to undergo an upper extremity procedure suitable for supraclavicular anesthesia. These procedures were expected to be associated with considerable postoperative pain.

We tested the hypothesis that sequential supraclavicular injection of 1.5% mepivacaine followed 90 seconds later by 0.5% ropivacaine provides a quicker onset and a longer duration of analgesia than an equidose combination of the 2 local anesthetics.

Patients were randomly assigned to either (1) combined group-ropivacaine and mepivacaine mixture; or (2) sequential group-mepivacaine followed by ropivacaine. The primary outcome was time to 4-nerve sensory block onset.

All quotes here are from the Supraclavicular study description

Supraclavicular Variables We'll Study Today

Variable	Description
`subject`	subject identifier (1-103)
`group`	1 = mixture, 2 = sequential (randomly assigned)
`onset_sensory`	Time to 4 nerve sensory block onset (min.)

```
supra <- supra_raw %>%
  mutate(trt = fct_recode(factor(group),
                          "mixture" = "1",
                          "sequential" = "2")) %>%
  rename(onset = onset_sensory) %>%
  select(subject, trt, onset, group)
```

The supra data

```
supra
```

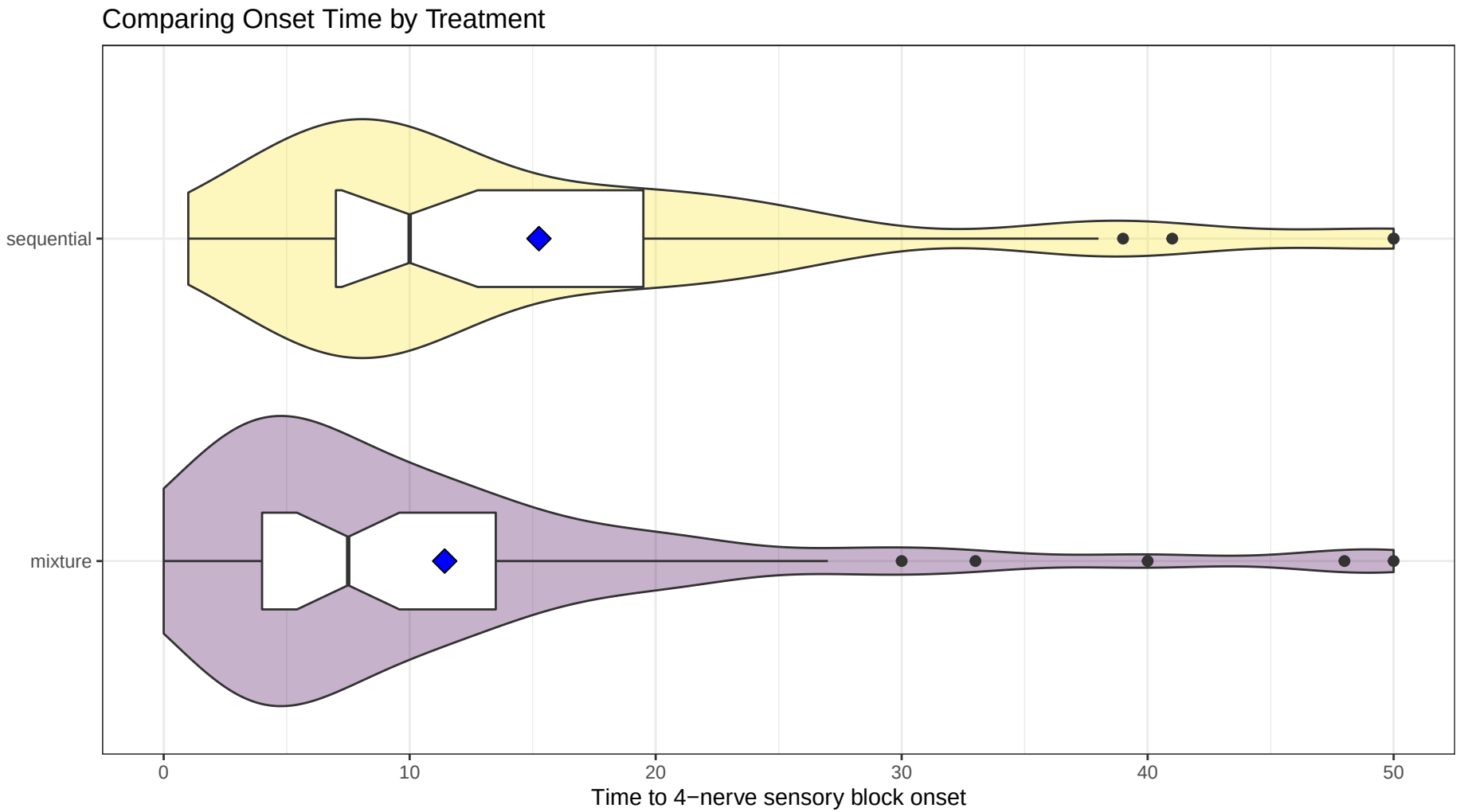
```
# A tibble: 103 x 4
   subject trt         onset group
     <dbl> <fct>       <dbl> <dbl>
 1       1 mixture         0     1
 2       2 sequential      7     2
 3       3 sequential     24     2
 4       4 mixture         4     1
 5       5 mixture        30     1
 6       6 sequential      4     2
 7       7 mixture        12     1
 8       8 sequential     13     2
 9       9 mixture        27     1
10      10 mixture         4     1
# ... with 93 more rows
```

DTDP: Compare onset by treatment

We'll add a blue diamond to indicate the means in each group, too.

```
ggplot(supra, aes(x = trt, y = onset)) +
  geom_violin(aes(fill = trt)) +
  geom_boxplot(width = 0.3, outlier.size = 2, notch = T) +
  stat_summary(fun = "mean", geom = "point",
               shape = 23, size = 4, fill = "blue") +
  guides(fill = "none") +
  scale_fill_viridis_d(alpha = 0.3) +
  coord_flip() +
  labs(y = "Time to 4-nerve sensory block onset",
       x = "",
       title = "Comparing Onset Time by Treatment")
```

DTDP: Compare onset by treatment (Result)



Numerical Summaries of Onset Time by Treatment

```
mosaic::favstats(onset ~ trt, data = supra)
```

```
         trt min Q1 median   Q3 max     mean       sd
1    mixture   0  4    7.5 13.5  50 11.42308 11.45553
2 sequential   1  7   10.0 19.5  50 15.25490 12.08113
   n missing
1 52       0
2 51       0
```

```
supra %>% group_by(trt) %>%
  summarize(n = n(), mean(onset), sd(onset), var(onset)) %>%
  kable(digits = 3)
```

trt	n	mean(onset)	sd(onset)	var(onset)
mixture	52	11.423	11.456	131.229
sequential	51	15.255	12.081	145.954

Study Description

- We selected 103 subjects from the population of all people:
  - ages 18-70 years
  - scheduled to undergo an upper extremity procedure suitable for supraclavicular anesthesia
  - who would have been eligible to participate in the study (details are fuzzy)
- We have randomly allocated subjects to one of two treatments (sequential or mixture.)
- For each subject, we have an outcome (onset time) associated with the treatment they received.
- The subjects were sampled from the population of interest independently of each other, so that the outcomes we see are not matched (or paired) in any way.

Key Question

Does the (true population) mean onset time differ between the two treatments?

Formal Language of Hypothesis Testing

- Null hypothesis H_0
  - H_0: population mean onset time with sequential = population mean onset time with mixture
  - H_0: difference in population means (sequential - mixture) = 0
- Alternative (research) hypothesis H_A or H_1
  - H_1: population mean onset time with sequential $\neq$ population mean onset time with mixture
  - H_1: difference in population means (sequential - mixture) $\neq 0$

Two (related) next steps

1. Given the data, we can then calculate an appropriate test statistic, then compare that test statistic to an appropriate probability distribution to obtain a p value. Small p values favor H_1 over H_0.
2. More usefully, we can use an appropriate probability distribution to help use the data to construct an appropriate **confidence interval** for the difference in population means.

Comparing Two Population Means

If we have independent samples (as we do in this scenario) where the data in the two treatment groups aren't matched or paired in any way, then we have at least four alternatives.

1. Compare population means using a pooled t test or confidence interval
   - This assumes equal population variances of the outcome in the two treatment groups.
   - This also assumes Normality of the outcome in each of the two treatment groups.
   - This is the result of a linear model of outcome ~ treatment.
2. Compare the population means using a Welch's t test or confidence interval
   - This does not assume equal population variances of the outcome.
   - This does assume Normality of the outcome in each of the two treatment groups.

Comparing Two Population Means (continued)

Additional alternatives when working with independent samples:

3. Compare the population means using a bootstrap approach to generate a confidence interval.
   - This does not assume either equal population variances or Normality.
4. Compare the population pseudo-medians (whatever those are) using a Wilcoxon signed rank test or confidence interval
   - This does not assume either equal population variances or Normality, but describes something other than population means, so we'll hold off on this for now.

Using a linear model to obtain pooled t-test results

- Let's start our comparison process with a pooled t test and associated 90% confidence interval, as we can obtain from a linear model.

```
m1 <- lm(onset ~ trt, data = supra)

tidy(m1, conf.int = TRUE, conf.level = 0.90) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	11.423	1.632	6.999	0.000	8.714	14.133
trtsequential	3.832	2.319	1.652	0.102	-0.019	7.682

What can we conclude about the difference in means?

Using a Two-Sample `t.test()` approach

We can obtain the same results for the t test comparing two independent samples, and assuming equal variances, with...

```
t.test(onset ~ trt, data = supra,
       var.equal = TRUE, conf.level = 0.90)
```

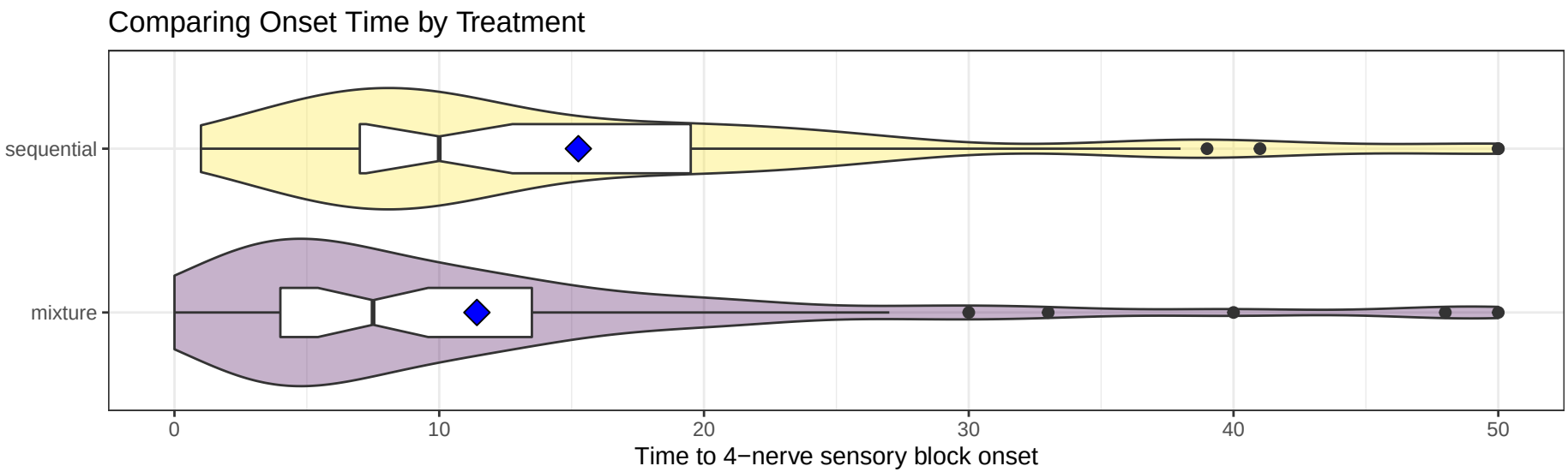
```
    Two Sample t-test

data:  onset by trt
t = -1.652, df = 101, p-value = 0.1016
alternative hypothesis: true difference in means between group
90 percent confidence interval:
 -7.68230689  0.01865682
sample estimates:
   mean in group mixture mean in group sequential
                 11.42308                 15.25490
```

Assessing Pooled T test Assumptions

In preparing a t test with equal variances, we assume that:

- each of the samples (sequential and mixture) are drawn from a Normally distributed population
- each of those populations have the same variance

Do these seem like reasonable assumptions in this case?



Let's first consider dropping the "equal variances" assumption.

The Welch's t test approach

Here is the Welch's t test comparing two independent samples, without assuming equal variances. . .

```
t.test(onset ~ trt, data = supra, conf.level = 0.90)
```

```
    Welch Two Sample t-test

data:  onset by trt
t = -1.6512, df = 100.47, p-value = 0.1018
alternative hypothesis: true difference in means between group
90 percent confidence interval:
 -7.6845015  0.0208514
sample estimates:
   mean in group mixture mean in group sequential
                 11.42308                 15.25490
```

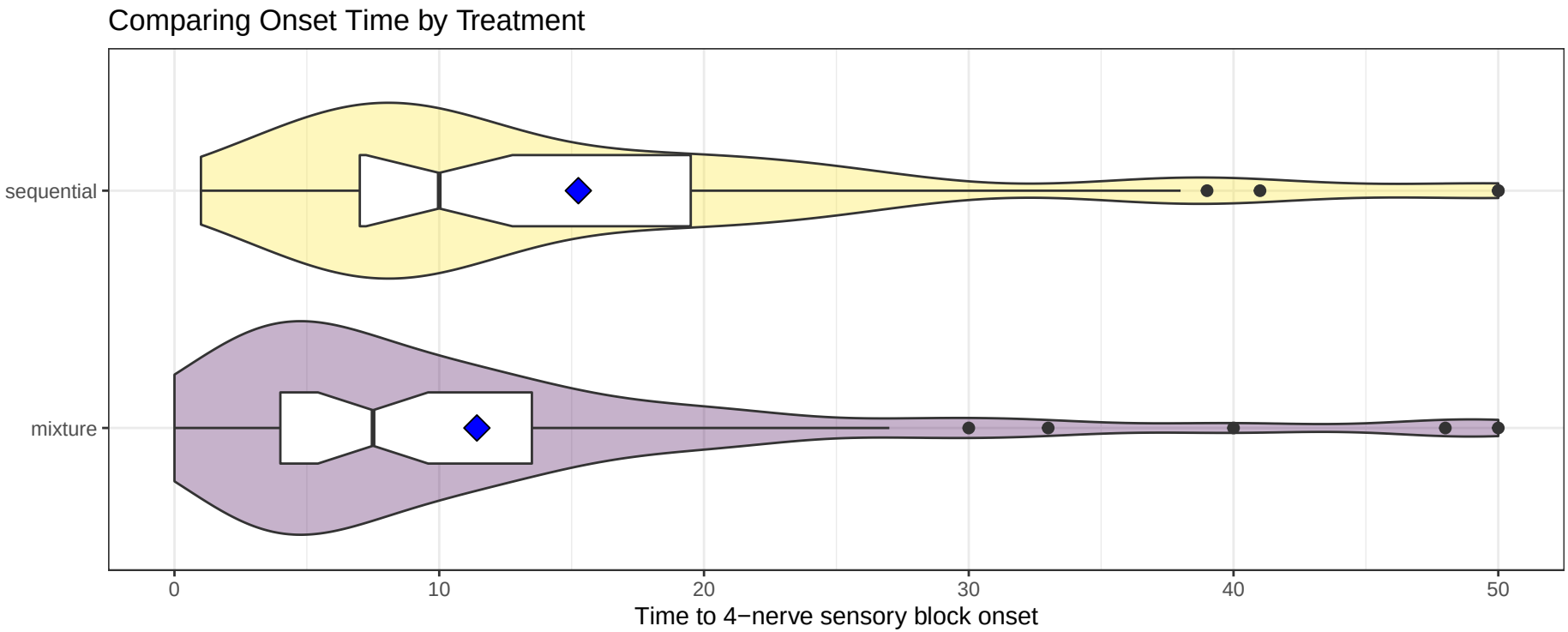
Comparing the two "T tests"

```
t1 <- t.test(onset ~ trt, data = supra, conf.level = 0.90,
             var.equal = TRUE)
w1 <- t.test(onset ~ trt, data = supra, conf.level = 0.90)

bind_rows(tidy(t1), tidy(w1)) %>%
  select(method, estimate, conf.low, conf.high, p.value) %>%
  kable(digits = 3)
```

method	estimate	conf.low	conf.high	p.value
Two Sample t-test	-3.832	-7.682	0.019	0.102
Welch Two Sample t-test	-3.832	-7.685	0.021	0.102

- Note: If we have a **balanced design** with equal sample sizes in the two groups, then these two approaches will yield essentially the same results. Here we have 51 sequential and 52 mixture subjects.

What about the Normality assumption?



- Does it seem reasonable to assume that the onset times are Normally distributed across the populations of sequential and mixed subjects, based on these samples of data?

Using a bootstrap approach

Consider the **bootstrap**, without assuming the population distributions are Normal, or have the same variance, at the expense of requiring some random sampling, which can lead to some conflicts.

- We'll use the `bootdif()` function I've provided in the Love-boost.R script.

```
set.seed(20211006)
supra %$% bootdif(y = onset, g = trt, conf.level = 0.90,
                  B.reps = 2000)
```

```
Mean Difference            0.05            0.95
     3.83182504      0.07592383      7.52083333
```

Using a bootstrap approach

- If we'd set a different seed or selected a different number of bootstrap replications, we'd get a different result.

```
set.seed(431)
supra %$% bootdif(y = onset, g = trt, conf.level = 0.90,
                  B.reps = 2000)
```

```
Mean Difference            0.05            0.95
     3.83182504     -0.08301282      7.55837104
```

```
set.seed(431)
supra %$% bootdif(y = onset, g = trt, conf.level = 0.90,
                  B.reps = 10000)
```

```
Mean Difference            0.05            0.95
     3.83182504      0.04935897      7.61973982
```

- This doesn't mean to suggest that we "shop around" until we find an appealing result, of course.

The Gathered Set of Estimates

Method	Estimate and 90% CI for $\mu_{Seq} - \mu_{Mix}$
Pooled Two-Sample T	3.83 (-0.02, 7.68)
Welch Two-Sample T	3.83 (-0.02, 7.69)
Bootstrap	3.83 (0.08, 7.52)

All of these results are in minutes (recall 0.08 minutes = 4.8 seconds) so are these **clinically meaningful** differences in this context?

- Do these data involve random sampling?
- What population(s) do these data represent?
- What can we say about the p values associated with these approaches?

A Study Involving Two Matched (Paired) Samples (to be discussed in Class 15)

The Hypoxia MAP Data

These data also come from the Cleveland Clinic's Statistical Education Dataset Repository.

Source: Turan et al. "Relationship between Chronic Intermittent Hypoxia and Intraoperative Mean Arterial Pressure in Obstructive Sleep Apnea Patients Having Laparoscopic Bariatric Surgery" *Anesthesiology* 2015; 122: 64-71.

```
hypox_raw <- read_csv("data/HypoxiaMAP.csv") %>%
  clean_names() %>%
  mutate(subject = row_number())

dim(hypox_raw)
```

```
[1] 281  37
```

Background and Study Description

[The Hypoxia MAP study] retrospectively examined the intraoperative blood pressures in 281 patients who had laparoscopic bariatric surgery between June 2005 and December 2009 and had a diagnosis of OSA within two preoperative years.

Time-weighted average (TWA) intraoperative MAP was the main outcome in the study. MAP (or mean arterial pressure) is a term used to describe an average blood pressure in a subject.

MAP is normally between 65 and 110 mmHg, and it is believed that a MAP $>$ 70 mmHg is enough to sustain the organs of the average person. If the MAP falls below this number for an appreciable time, vital organs will not get enough oxygen perfusion, and will become hypoxic, a condition called ischemia.

Our Objective with these Data

We will focus today on two measurements of MAP for each subject (outside of some missing data).

- MAP1 = time-weighted average mean arterial pressure from ET intubation to trocar insertion, in mm Hg.
- MAP2 = time-weighted average mean arterial pressure from trocar insertion to the end of the surgery, in mm Hg.

We are interested in estimating the **difference** between the two MAP levels, across a population of subjects like those enrolled in this study.

Our Key Variables

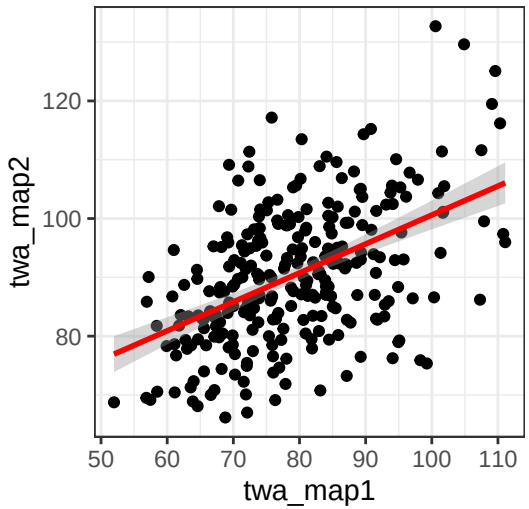
- For each subject, we have two outcomes to compare: their MAP1 and their MAP2.

```
hypox <- hypox_raw %>%
  select(subject, twa_map1, twa_map2) %>%
  mutate(map_diff = twa_map2 - twa_map1)

hypox %>% head(., 4)
```

```
# A tibble: 4 x 4
  subject twa_map1 twa_map2 map_diff
    <int>    <dbl>    <dbl>    <dbl>
1       1     67.9     87.4     19.5
2       2     67.0     83.3     16.3
3       3     91.6     83.0     -8.59
4       4     67.1     79.9     12.8
```

We have Paired Samples in this setting

- Every MAP1 value is connected to the MAP2 value for the same subject. We say that the MAP1 and MAP2 are paired by subject.



Each subject provides a MAP1 and a MAP2

- The pairing is fairly strong here. The Pearson correlation of MAP1 and MAP2 across the subjects with complete data is 0.494.
- It makes sense to calculate the (paired) difference in MAP values for each subject, so long as there aren't any missing data.

Are there any missing values?

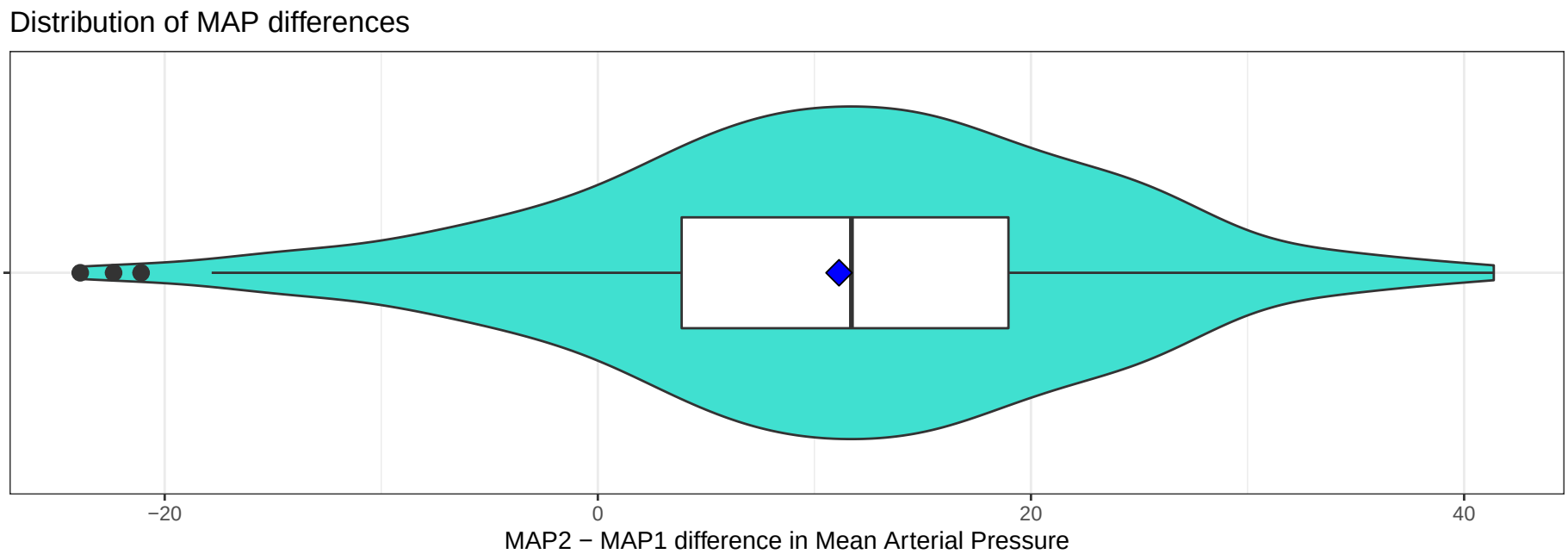
```
miss_var_summary(hypox)
```

```
# A tibble: 4 x 3
  variable n_miss pct_miss
  <chr>     <int>    <dbl>
1 twa_map1      4     1.42
2 map_diff      4     1.42
3 subject       0     0
4 twa_map2      0     0
```

```
hypox <- hypox %>% filter(complete.cases(map_diff))
```

Boxplot of the MAP differences

```
ggplot(data = hypox, aes(x = map_diff, y = "")) +
  geom_violin(fill = "turquoise") +
  geom_boxplot(width = 0.3, outlier.size = 3) +
  stat_summary(fun = "mean", geom = "point",
               shape = 23, size = 4, fill = "blue") +
  labs(x = "MAP2 - MAP1 difference in Mean Arterial Pressure",
       y = "", title = "Distribution of MAP differences")
```



Numerical Summaries

```
res1 <- as_tibble(bind_rows(
  mosaic::favstats(~ twa_map1, data = hypox),
  mosaic::favstats(~ twa_map2, data = hypox),
  mosaic::favstats(~ map_diff, data = hypox))) %>%
  mutate(item = c("map1", "map2", "map_diff")) %>%
  select(item, n, mean, sd, min, median, max)

res1 %>% kable()
```

item	n	mean	sd	min	median	max
map1	277	79.24274	11.73903	51.96	78.02	111.10
map2	277	90.37921	11.69104	66.17	89.74	132.71
map_diff	277	11.13646	11.78064	-23.90	11.71	41.37

- Is the mean of `map_diff` equal to the difference between the mean of `map2` and the mean of `map1`? Other summaries?

Hypothesis Testing Comparing Paired Samples

- Null hypothesis H_0
  - H_0: population mean of paired differences (MAP2 - MAP1) $= 0$
- Alternative (research) hypothesis H_A or H_1
  - H_1: population mean of paired differences (MAP2 - MAP1) $\neq 0$

Two (related) next steps

1. Given the data, we can then calculate the paired differences, then an appropriate test statistic based on those differences, which we compare to an appropriate probability distribution to obtain a p value. Again, small p values favor H_1 over H_0.
2. More usefully, we can calculate the paired differences, and then use an appropriate probability distribution to help use the data to construct an appropriate **confidence interval** for the population of those differences.

Paired T test via Linear Model

```
m3 <- lm(map_diff ~ 1, data = hypox)

summary(m3)$coef
```

```
             Estimate Std. Error  t value     Pr(>|t|)
(Intercept) 11.13646  0.7078298 15.73325 2.971357e-40
```

```
confint(m3, conf.level = 0.90)
```

```
               2.5 %   97.5 %
(Intercept) 9.743031 12.52989
```

```
summary(m3)$r.squared
```

```
[1] 0
```

Tidied Regression Model

```
tidy(m3, conf.int = TRUE, conf.level = 0.90) %>%
  select(term, estimate, conf.low, conf.high) %>%
  kable(digits = 3)
```

term	estimate	conf.low	conf.high
(Intercept)	11.136	9.968	12.305

```
tidy(m3, conf.int = TRUE, conf.level = 0.90) %>%
  select(term, estimate, std.error, statistic, p.value) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	11.136	0.708	15.733	0

Paired T test via t.test

```
hypox %$% t.test(map_diff, conf.level = 0.90)
```

```
    One Sample t-test

data:  map_diff
t = 15.733, df = 276, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
  9.968265 12.304660
sample estimates:
mean of x
 11.13646
```

Paired T Confidence Interval yet another way

```
hypox %$%
  smean.cl.normal(map_diff, conf = 0.90)
```

```
     Mean     Lower     Upper
11.136462  9.968265 12.304660
```

The function `smean.cl.normal` (and that's an L, not a 1 after C) comes from the `Hmisc` package.

So does the `smean.cl.boot` function we'll see on the next slide, which will let us avoid the key assumption of Normality for the population of paired differences.

Bootstrap for Comparing Paired Means

```
set.seed(20211006)
hypox %$%
  Hmisc::smean.cl.boot(map_diff, conf = 0.90, B = 1000)
```

```
     Mean      Lower      Upper
11.136462  9.932226 12.323684
```

```
set.seed(123431)
hypox %$%
  Hmisc::smean.cl.boot(map_diff, conf = 0.90, B = 5000)
```

```
    Mean     Lower     Upper
11.13646 10.00769 12.30044
```

Gathered Estimates from our Paired Samples

Method	Estimate and 90% CI	Assumes Normality?
Paired t	11.14 (9.97, 12.30)	Yes
Bootstrap	11.14 (9.93, 12.32)	No

We estimate that the time-weighted average mean arterial pressure is 11.14 mm Hg higher (90% CIs shown above) after trocar insertion than it is during the period from ET intubation to trocar insertion, based on our sample of 277 subjects with complete data in this study.

- Does it matter much whether we assume Normality here?
- What can we say about the p values here?
- Is this a random sample of subjects?
- What population do these data represent?

Paired vs. Independent Samples

If you can afford to obtain n = 400 observations to compare means under exposure A to means under exposure B, and you could either:

1. select a random sample from the population of interest containing 400 people and then randomly assign 200 people to receive exposure A and the remaining 200 people to receive exposure B (thus doing an independent samples study), or
2. select a random sample from the population of interest containing 200 people and then randomly assign 100 of them to get exposure A first, and then, a little later, when the effects have worn off, to then receive exposure B, while the other 100 people are assigned to receive B first, then A (thus doing a paired samples study)

Which do you think would be the more powerful study design?