The two-sample bootstrap

Stat 250

Click here for PDF version

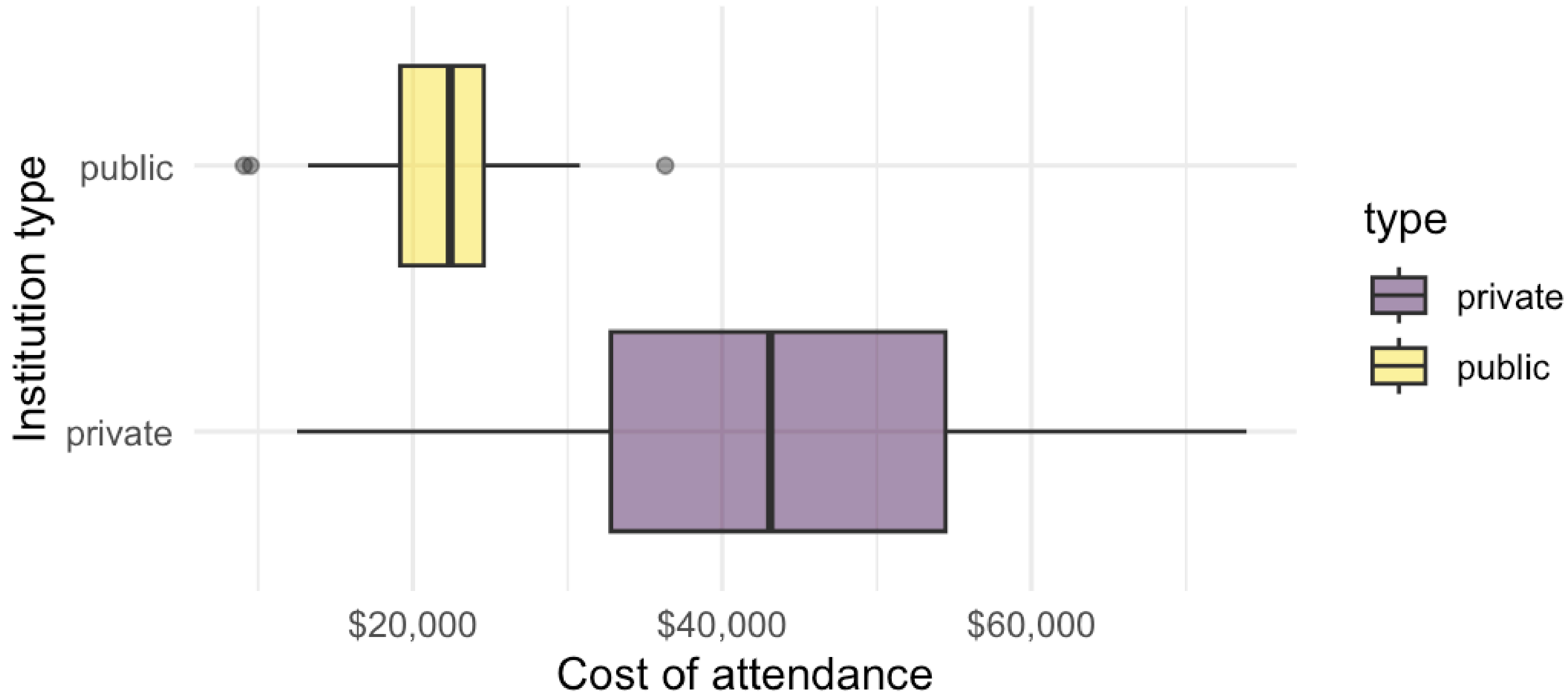
College score card

> The College Scorecard is designed to increase transparency, putting the power in the hands of the public — from those choosing colleges to those improving college quality — to see how well different schools are serving their students.

Select variables:

- `type` — public/private
- `cost` — total cost of attendance
- `grad_rate` — proportion of students graduating within six years
- `region` — region of the U.S.

How does cost of attendance differ by inst. type?



type	min	Q1	median	Q3	max	mean	sd	n	missing
private	12,510	32,796	43,113	54,427	73,892	43,616	14,872	118	9
public	9,075	19,182	22,415	24,580	36,319	21,610	4,984	55	5

The two-sample bootstrap

1. Draw a resample of size *m* with replacement from the first sample and a separate resample of size *n* from the second sample.
2. Compute a statistic that compares the two groups, such as the difference between the two sample means.
3. Repeat the above steps many times, say 10,000.
4. Construct the bootstrap distribution of the statistic. Inspect its spread, bias, and shape.

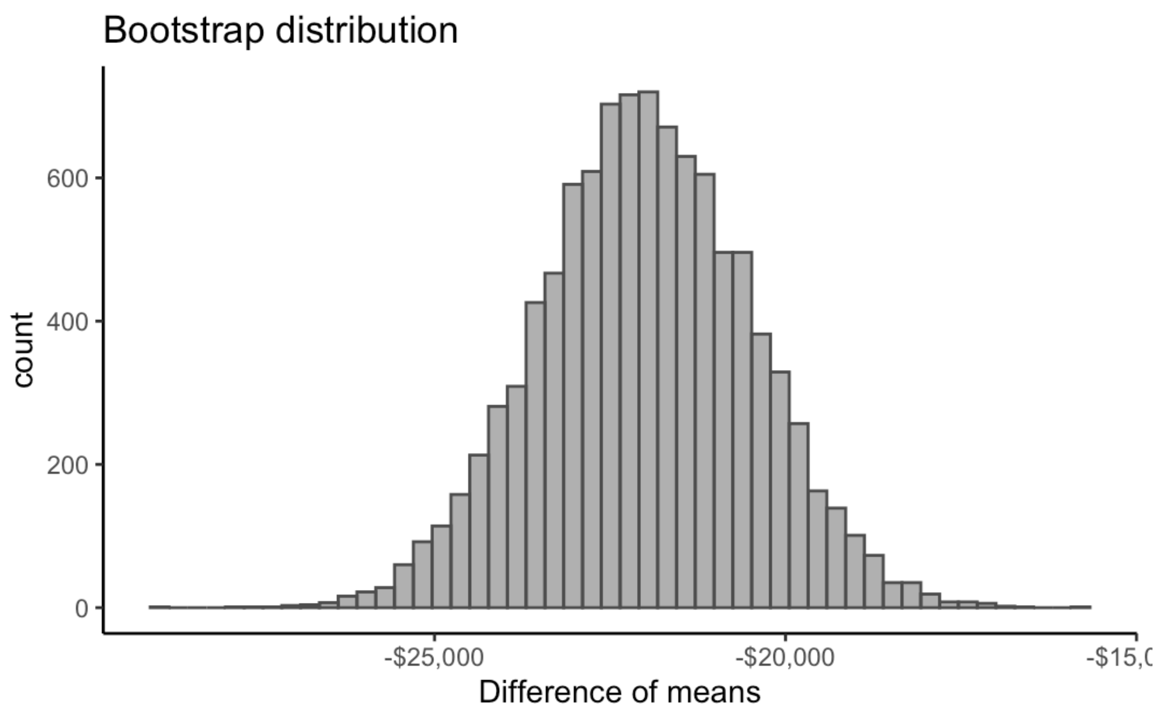
R implementation

```r
library(tidyverse) # for some data manipulation tools

N <- 10^4 # Number of bootstrap resamples

# Create a vector for each group
public <- college |>
  filter(type == "public") |>
  pull(cost) |> na.omit()
private <- college |>
  filter(type == "private") |>
  pull(cost) |>
  na.omit()
```

R implementation

```r
# A place to store the statistics
diff_mean <- numeric(N)

# Resample and calculate the statistic
for (i in 1:N) {
  boot_public <- sample(public, replace = TRUE)
  boot_private <- sample(private, replace = TRUE)
  diff_mean[i] <- mean(boot_public) - mean(boot_private)
}
```

Bootstrap distribution



Statistic	Value
Mean	−22,012.61
SD	1,510.15
Bias	−7.20

90% Percentile confidence interval

Calculation:

```
quantile(diff_mean, probs = c(0.05, 0.95))
```

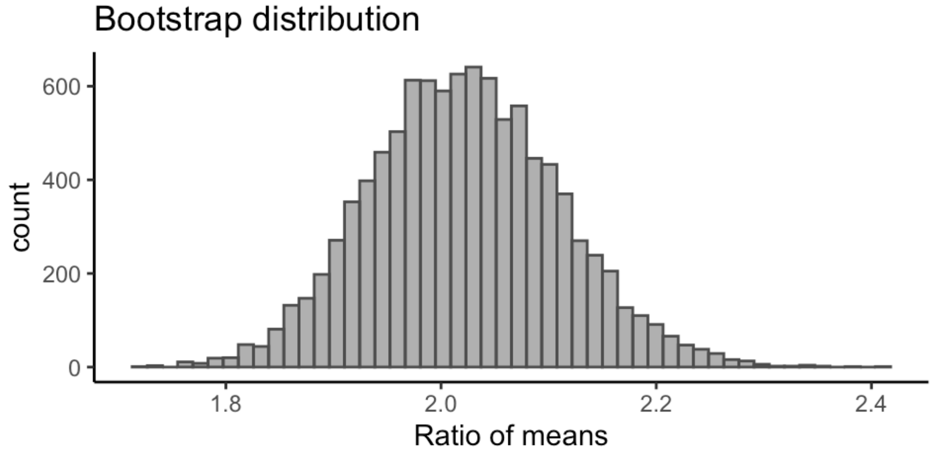
```
     5%     95% 
-24516 -19544 
```

Interpretation:

The bootstrap is flexible

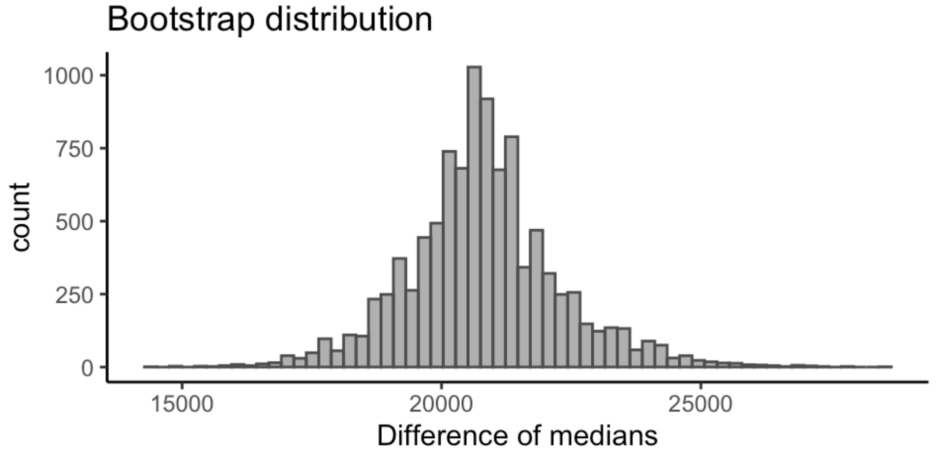
Ratio of means

```
N <- 10^4

ratio_means <- numeric(N)
for (i in 1:N) {
  boot_public <- sample(public, replace = TRUE)
  boot_private <- sample(private, replace = TRUE)
  ratio_means[i] <- mean(boot_private) / mean(boot_publi
}
```



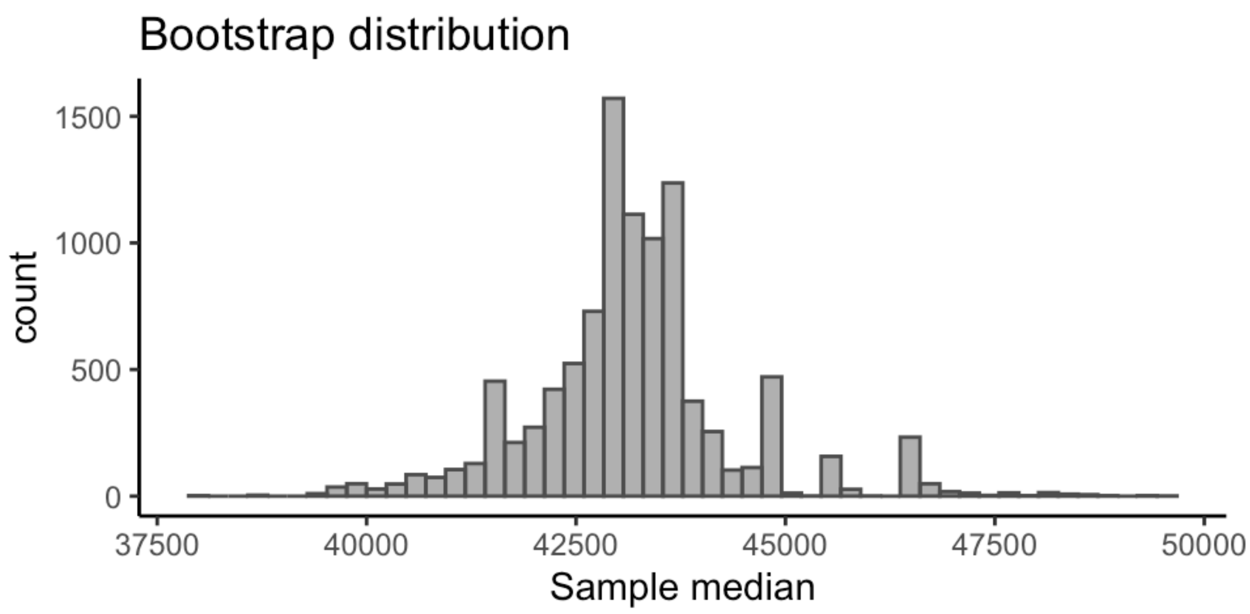
Difference of medians

```
N <- 10^4

diff_medians <- numeric(N)
for (i in 1:N) {
  boot_public <- sample(public, replace = TRUE)
  boot_private <- sample(private, replace = TRUE)
  diff_medians[i] <- median(boot_private) - median(boot_
}
```



Medians

```
N <- 10^4

medians <- numeric(N)
for (i in 1:N) {
  medians[i] <- median(sample(private, replace = TRUE))
}
```



Cautions

- Bootstrap often provides a poor approximation to the sampling distribution of medians and other quantiles
- When $|\text{Bias}/\text{Boot SE}| > 0.02$ there could be substantial impact on accuracy of confidence intervals, so reflect on setting and may need to consider other approaches to construct CIs
- Bootstrapping does not overcome the issue of small sample sizes

Matched pairs

Is it safe to look at social media while driving?

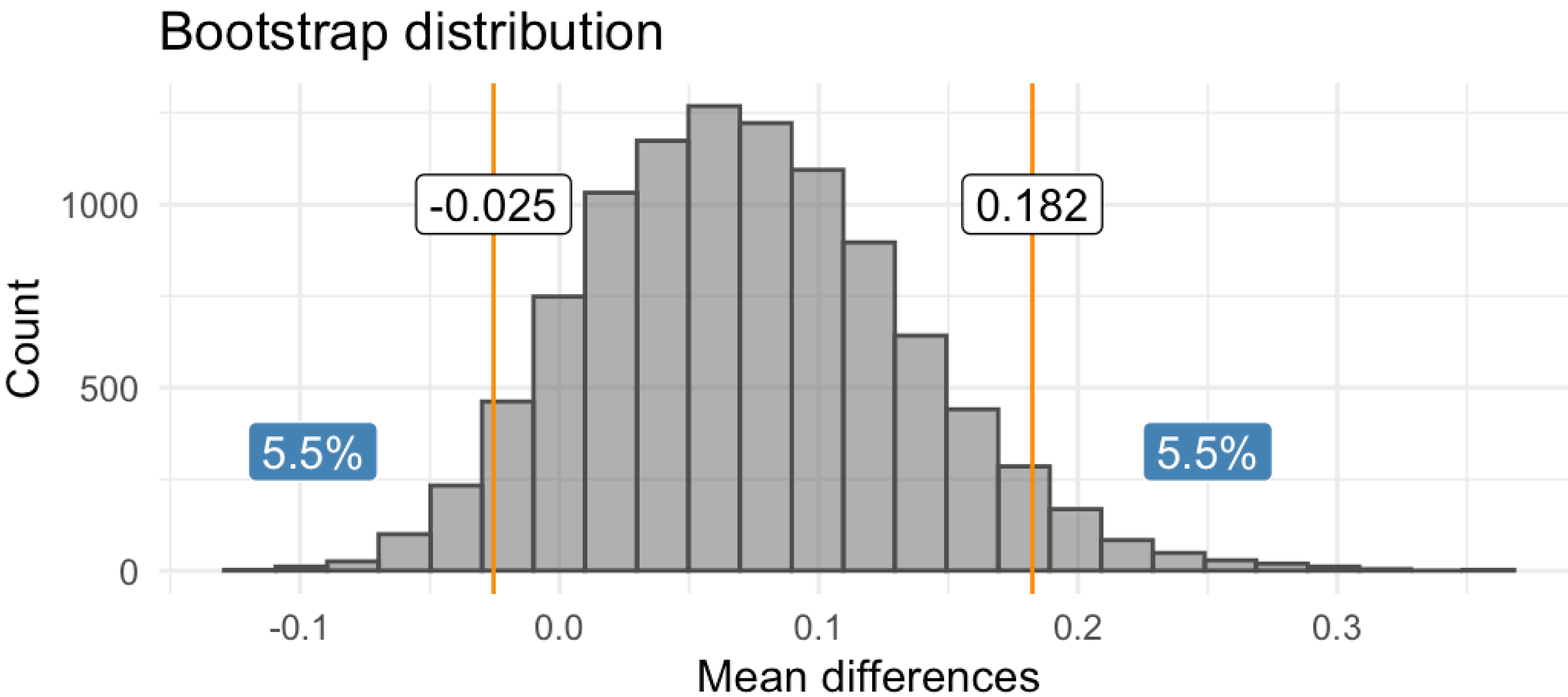
- Previous research on smart phone use while driving has primarily focused on phone calls and texting.
- Study looked at the effects of different smart phone tasks on car-following performance in a driving simulator.
- Drivers performed driving only baseline simulation
- Drivers performed other phone tasks: texting, reading Facebook posts, exchanging photos on Snapchat, viewing updates on Instagram
- Brake reaction times (in seconds) recorded

Matched pairs

Subject	Baseline	SnapChat	Diff
1	0.863	0.865	0.00164
2	0.847	0.783	-0.06333
3	0.836	0.808	-0.02776
4	0.655	1.010	0.35421
5	0.900	0.837	-0.06324
6	0.957	1.175	0.21849
7	0.780	0.817	0.03695
8	0.954	0.861	-0.09368
9	0.970	0.717	-0.25335
10	1.103	1.141	0.03873
11	0.925	0.583	-0.34167
12	0.833	0.883	0.04999
13	0.833	0.995	0.16119
14	0.773	0.837	0.06407
15	0.914	1.008	0.09476
16	0.858	1.137	0.27844
17	0.822	1.733	0.91115
18	0.963	0.883	-0.07945

- Two measurements on each case
- Goal is to estimate the true mean difference in reaction time
- Calculate differences, then it’s a one-sample bootstrap

Bootstrap percentile interval



We are 89% confident that the mean difference in reaction times is between -0.025 and 0.182 seconds.

implementation

```r
brake <- brake |> select(Subject, Baseline, SnapChat) |>
  mutate(Diff = SnapChat - Baseline)

N <- 1e4
n <- length(brake$Diff)
mean_diffs <- numeric(N)
set.seed(1112)

for(i in 1:N) {
  mean_diffs[i] <- mean(sample(brake$Diff, n, replace =
}

quants <- quantile(mean_diffs, probs = c(0.045, 0.955))
```