Notes: MS 204 Chapter 4.2

Overview

- Paired data

An example

Is there a difference between reading and math SAT scores? Researchers are curious if students score better on the verbal or math sections of the SAT. They collect data on the average SAT score in each state.

```
library(tidyverse)
library(oilabs)
library(mosaic)
head(SAT)
```

```
##        state expend ratio salary frac verbal math  sat
## 1    Alabama  4.405  17.2 31.144    8    491  538 1029
## 2     Alaska  8.963  17.6 47.951   47    445  489  934
## 3    Arizona  4.778  19.3 32.175   27    448  496  944
## 4   Arkansas  4.459  17.1 28.934    6    482  523 1005
## 5 California  4.992  24.0 41.078   45    417  485  902
## 6   Colorado  5.443  18.4 34.571   29    462  518  980
```

```
SAT <- SAT %>%
  mutate(score.diff = verbal - math)
SAT %>%
  summarise(ave.diff = mean(score.diff), sd.diff = sd(score.diff))
```

```
##   ave.diff  sd.diff
## 1   -51.64 10.46034
```

Inference for paired data

parameter

point estimate

population

Inference

Central limit theorem for sample mean differences from paired data

Assumptions?

Hypothesis test

Do these data provide convincing evidence that there is a difference between the verbal and math sections of the SAT?

Confidence interval

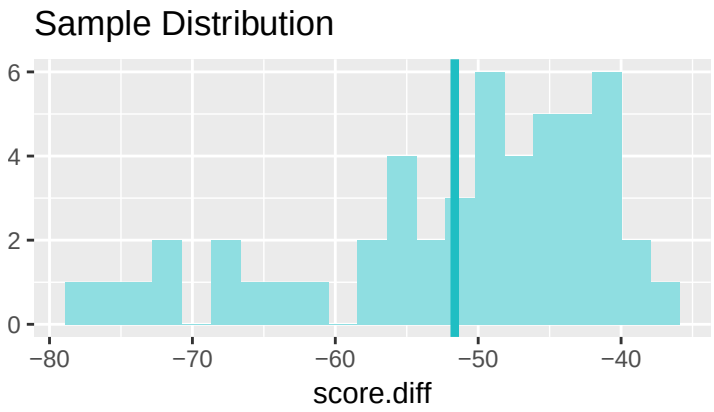
```
qt(0.025, df = 49)
```

```
## [1] -2.009575
```

Code

```
inference(score.diff, data = SAT, statistic = "mean",
          type = "ci", method = "theoretical")
```

```
## Single numerical variable
## n = 50, y-bar = -51.64, s = 10.4603
## 95% CI: (-54.6128 , -48.6672)
```




```
inference(score.diff, data = SAT, statistic = "mean",
          type = "ht", alternative = "twosided", method = "theoretical", null = 0)
```

```
## Single numerical variable
## n = 50, y-bar = -51.64, s = 10.4603
## H0: mu = 0
## HA: mu != 0
## t = -34.908, df = 49
## p_value = < 0.0001
```