Statistics with Recitation: TA Session

Danny Po-Hsien Kang (康柏賢)

November 25, 2025

Today's agenda

1. Hypothesis Testing: t-test
   - One-Sample t-test
   - Paired Sample t-test
   - Two Independent Sample t-test

Reminders

- Quiz 6 will be held on December 2.
- The R Quiz will focus on Week 7-13 materials in the TA Session.
- All of the problems will be multiple-choice questions.
  - 15 questions, 2 points each.
  - Similar to Quiz 3.

Today's Dataset

- Please download the three datasets from the OpenIntro website.
  - `china.csv`: Weekly time spent on child care in China.
  - `textbooks.csv`: Prices of textbooks for various courses at UCLA Bookstore and Amazon.com.
  - `ncbirths.csv`: A 2004 dataset from North Carolina on births, including mothers' habits and practices, with 1,000 cases.
- After that, import the data

```
china <- read.csv("data/china.csv")
book <- read.csv("data/textbooks.csv")
birth <- read.csv("data/ncbirths.csv")
```

One-Sample t-test: `t.test()`

- **When to use:** Test whether the population mean equals a target μ.
- **Syntax:**

```
t.test(vec, mu,
       alternative, conf.level)
```

- **Example 1:**

```
china <- china %>% filter(child_care >= 0)

china_male <- china %>% filter(gender == 1)
t.test(china_male$child_care,
       mu = 25,
       conf.level = 0.95)
```

- Default value for parameter `alternative` is `"two.sided"`
  - Use `"less"` or `"greater"` for one-sided tests.

One-Sample t-test: t.test()

- **Output:**

```
        One Sample t-test

data:  china_male$child_care
t = -9.2583, df = 357, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 25

95 percent confidence interval:
 13.17385 17.31777

sample estimates:
mean of x
15.24581
```

One-Sample t-test: `t.test()`

- **Example 2:**

```
t.test(china$child_care,
       mu = 25,
       alternative = "less",
       conf.level = 0.95)
```

One-Sample t-test: `t.test()`

- **Output:**

```
        One Sample t-test

data:  china$child_care
t = -0.73041, df = 1021, p-value = 0.2327

alternative hypothesis: true mean is less than 25

95 percent confidence interval:
     -Inf 25.79633

sample estimates:
mean of x
 24.36497
```

Paired Sample t-test: `t.test()`

- **When to use:**
  - Pre/post on the same subjects or two methods measured on the same items.
  - You want to test whether the mean difference equals μ (usually 0).
- **Syntax:**

```
t.test(vec_1, vec_2,
       paired = TRUE, conf.level)
```

- **Example 1:**

```
t.test(book$ucla_new,
       book$amaz_new,
       paired = TRUE,
       coef.level = 0.95)
```

Paired Sample t-test: `t.test()`

- **Output:**

```
        Paired t-test

data:  book$ucla_new and book$amaz_new
t = 7.6488, df = 72, p-value = 6.928e-11

alternative hypothesis: true mean difference is not equal to 0

95 percent confidence interval:
  9.435636 16.087652

sample estimates:
mean difference
       12.76164
```

Two Independent Sample t-test: `t.test()`

- **When to use:**
  - Comparing the means of two distinct groups (e.g., treatment vs control; smokers vs non-smokers).
- **Syntax:**

```
t.test(data$ValueVar ~ data$GroupVar,
       conf.level)
```

- **Example 1:**

```
t.test(birth$weight ~ birth$habit,
       conf.level = 0.95)
```

Two Independent Sample t-test: `t.test()`

- **Output:**

```
        Welch Two Sample t-test

data:  birth$weight by birth$habit
t = 2.359, df = 171.32, p-value = 0.01945

alternative hypothesis: true difference in means between
group nonsmoker and group smoker is not equal to 0

95 percent confidence interval:
 0.05151165 0.57957328

sample estimates:
mean in group nonsmoker    mean in group smoker
               7.144273                6.828730
```