

Lecture 17: Paired Data and Difference of Two Means

Chapter 5.2, 5.1

1 / 17

Goals for Today

- ▶ Difference of means
- ▶ Note on Practical vs Statistical Significance
- ▶ Paired differences of means

2 / 17

6 Types of Questions

Here are the 6 broad types of questions about **population parameters** we'll be answering with statistical methods: confidence intervals and hypothesis tests

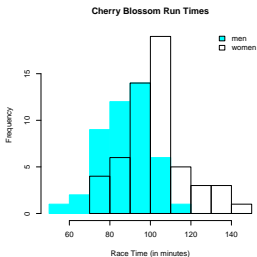
1. What is the mean value μ ?
2. Are the means μ_1 and μ_2 of two groups different?
3. What is the mean paired difference μ_{diff} ?
4. What is the proportion p of "successes"?
5. Are the proportions of "successes" p_1 and p_2 of two groups different?
6. Are the means μ_1, \dots, μ_k of k groups different?

Today we look at 3 and 2.

General Outline

Chapter 5.2: Are Two Means μ_1 & μ_2 Different?

We randomly sample 45 men (of 7192) and 55 women (of 9732) runners in the 2012 Cherry Blossom Run. Did men run faster than women?



	men	women
\bar{x}	87.65	102.13
s	12.5	15.2
n	45	55

5 / 17

Difference in Means

6 / 17

Normality of Sampling Distribution

Normality of Sampling Distribution

Confidence Interval

Hypothesis Test

Practical vs Statistical Significance

When rejecting H_0 , we call this a **statistically significant** result. But statistically significant results aren't always **practically significant**.

Say for **very** large n_M & n_F we observe $\bar{x}_M = 87.65$ and $\bar{x}_F = 87.651$ and reject H_0 .

The point estimate of the difference $\bar{x}_M - \bar{x}_F = 0.001$. Near negligible!

However, the 95% CI might be:

$$[0.0005, 0.0015]$$

11 / 17

Practical vs Statistical Significance

Moral of the story

- ▶ Hypothesis tests with "rejections of H_0 " focus almost entirely on **statistical significance**.
- ▶ Confidence intervals allow you to also focus on **practical significance**.

12 / 17

Hypothesis Test

13 / 17

Chapter 5.1: Paired Data

Two sets of observations are **paired** if each observation in one set has a special correspondence or connection with exactly one observation in the other data set.

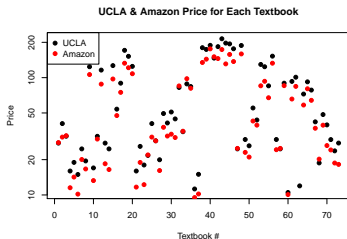
Examples:

- ▶ Cholesterol levels before and after some intervention for the same person
- ▶ Disease rates amongst pairs of twins
- ▶ In the text: price of the same textbook at the UCLA bookstore vs Amazon

14 / 17

Paired Differences

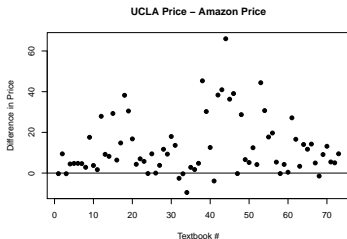
The methodology for paired data remains the same, except our **observations** are the difference in pairs. Example, for the UCLA Bookstore vs Amazon book price example in the text



15 / 17

Paired Differences

The methodology for paired data remains the same, except our **observations** are the difference in pairs. Example, for the UCLA Bookstore vs Amazon book price example in the text



16 / 17

Paired Differences

We have

- ▶ population parameter is μ_{diff} with point estimate \bar{x}_{diff}
- ▶ Check the conditions not on the original observations, but rather the [differences](#).
- ▶ If met, \bar{x}_{diff} has a normal sampling distribution
 - ▶ mean μ_{diff}
 - ▶ $SE_{diff} = \frac{\sigma_{diff}}{\sqrt{n_{diff}}} \approx \frac{s_{diff}}{\sqrt{n_{diff}}}$

17 / 17

Next Time

- ▶ t-test

18 / 17