Statistical Methods for Population Health

Week 1: Introduction to Statistics

Ruoqing Zhu, Ph.D. <rqzhu@illinois.edu>

June 27, 2022

Department of Statistics
University Illinois Urbana-Champaign

Welcome!

- Welcome to the Statistics section!

Welcome!

- Welcome to the Statistics section!
- This is a sequence of three lectures (I call them guided TBL)

Welcome!

- Welcome to the Statistics section!
- This is a sequence of three lectures (I call them guided TBL)
- Core skills
    - Statistical principles
    - Result interpretation
    - Basic data analysis using R
    - Some modeling techniques

Weekly Schedule

- Week 1: R Introduction and Statistical Principles

Weekly Schedule

- Week 1: R Introduction and Statistical Principles

Weekly Schedule

- Week 1: R Introduction and Statistical Principles
- Week 2: Testing Mean Differences and Associations

Weekly Schedule

- Week 1: R Introduction and Statistical Principles
- Week 2: Testing Mean Differences and Associations
- Week 3: Statistical Models for Multivariate Analysis

Weekly Schedule

- Week 1: R Introduction and Statistical Principles
- Week 2: Testing Mean Differences and Associations
- Week 3: Statistical Models for Multivariate Analysis

Weekly Schedule

- Week 1: R Introduction and Statistical Principles
- Week 2: Testing Mean Differences and Associations
- Week 3: Statistical Models for Multivariate Analysis

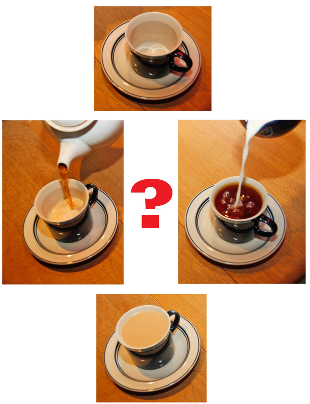
Weekly Schedule

- Week 1: R Introduction and Statistical Principles

The Lady Tasting Tea

The Lady Tasting Tea Problem

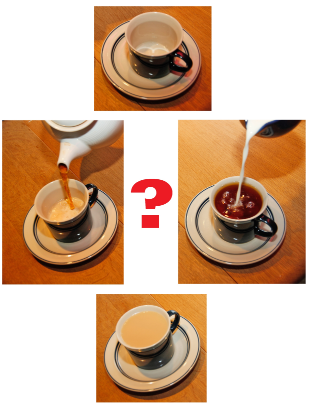
- In 1920s Cambridge, England, a Lady, named Muriel Bristol, claimed to be able to tell whether the tea or the milk was added first by the taste of it!



"The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century" (2001) by David Salsburg

The Lady Tasting Tea Problem

- In 1920s Cambridge, England, a Lady, named Muriel Bristol, claimed to be able to tell whether the tea or the milk was added first by the taste of it!
- A statistician Ronald Fisher what to test if thats true or not using probability principles



“The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century” (2001) by David Salsburg

The Essential Idea

- Suppose Lady Bristol does not have that ability, then she will be randomly guessing

The Essential Idea

- Suppose Lady Bristol does not have that ability, then she will be randomly guessing
- Let's prepare many cups of tea for her to identify, then we would expect her to identify, on average, half of them correctly.

The Essential Idea

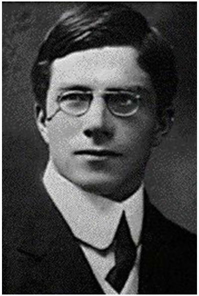
- Suppose Lady Bristol does not have that ability, then she will be randomly guessing
- Let's prepare many cups of tea for her to identify, then we would expect her to identify, on average, half of them correctly.
- However, if she can identify many of them correctly, then we may have to reject the assumption of random guessing

The Essential Idea

- Suppose Lady Bristol does not have that ability, then she will be randomly guessing
- Let's prepare many cups of tea for her to identify, then we would expect her to identify, on average, half of them correctly.
- However, if she can identify many of them correctly, then we may have to reject the assumption of random guessing
- The question is, how many is too many?

The Essential Idea

- Two important concepts:
    1. Experimental design
    2. Hypothesis testing



Sir Ronald A. Fisher
(1890 - 1962)

Fisher's Exact Test

- Fisher prepared 8 cups of tea, 4 with milk added first and 4 with tea added first.

Fisher's Exact Test

- Fisher prepared 8 cups of tea, 4 with milk added first and 4 with tea added first.
- Lady Bristol was asked to identify them
- There are totally $\frac{8!}{4!(8-4)!} = 70$ possible results. If she is randomly guessing, then each result has equal chance:

Fisher's Exact Test

- Fisher prepared 8 cups of tea, 4 with milk added first and 4 with tea added first.
- Lady Bristol was asked to identify them
- There are totally $\frac{8!}{4!(8-4)!} = 70$ possible results. If she is randomly guessing, then each result has equal chance:
  - The chance of identifying all 4 correctly is 1/70
  - The chance of 3 is 16/70
  - The chance of 2 is 36/70
  - The chance of 1 is 16/70
  - The chance of 0 is 1/70

Fisher's Exact Test

- Fisher prepared 8 cups of tea, 4 with milk added first and 4 with tea added first.
- Lady Bristol was asked to identify them
- There are totally $\frac{8!}{4!(8-4)!} = 70$ possible results. If she is randomly guessing, then each result has equal chance:
  - The chance of identifying all 4 correctly is 1/70
  - The chance of 3 is 16/70
  - The chance of 2 is 36/70
  - The chance of 1 is 16/70
  - The chance of 0 is 1/70
- What can be considered as "surprising" evidence given the assumption that she is randomly guessing?

Fisher's Exact Test

- Fisher used 0.05 as the cut-off:

  If the actual result falls into the top 5% of the most extreme cases, we consider this as a surprising evidence and claim that she is not randomly guessing.

Fisher's Exact Test

- Fisher used 0.05 as the cut-off:

  If the actual result falls into the top 5% of the most extreme cases, we consider this as a surprising evidence and claim that she is not randomly guessing.

- How many cups Lady Bristol identified correctly?

Recap

Some **key steps** in hypothesis testing:

1). Form Null and Alternative hypotheses:

Null H_0 : Random Guessing vs. Alt. H_1 : Not Random Guessing

Recap

Some **key steps** in hypothesis testing:

1). Form Null and Alternative hypotheses:

Null H_0 : Random Guessing   vs.   Alt. H_1 : Not Random Guessing

2). Perform an experiment and observe that the lady identified the 4 correctly.

Recap

Some **key steps** in hypothesis testing:

1). Form Null and Alternative hypotheses:

Null H_0 : Random Guessing   vs.   Alt. H_1 : Not Random Guessing

2). Perform an experiment and observe that the lady identified the 4 correctly.

3). If the Null hypothesis is correct, there is only 1.4% chance that one can guess 4 correctly

Recap

Some **key steps** in hypothesis testing:

1). Form Null and Alternative hypotheses:

Null H_0 : Random Guessing vs. Alt. H_1 : Not Random Guessing

2). Perform an experiment and observe that the lady identified the 4 correctly.

3). If the Null hypothesis is correct, there is only 1.4% chance that one can guess 4 correctly

4). This is a "small probability event" (smaller than a pre-determined significance level, $\alpha = 0.05$), so we will make a conclusion to reject the Null.

Recap

- If we reject the Null hypothesis, does it mean that Lady Bristol actually has the ability to identify them?

Correct or Wrong decision?

- We could still make a wrong decision. In fact, there are four situations:

	Accept H_0	Reject H_0
H_0 true	✓	Type I Error
H_0 false	Type II Error	✓

- Type I error: H_0 true but we reject it.
- Type II error: H_0 false but we accept it.

Correct or Wrong decision?

- Type I error can be controlled using the α level we choose.
- $1 -$ Type I error is called the confidence level

Correct or Wrong decision?

- Type I error can be controlled using the α level we choose.
- $1 -$ Type I error is called the confidence level
- Type II error is difficult to analyze because we don't know what the alternative may look like. For example, the lady may have 0.7 probability to identify a correct one, or 0.9, 0.51, etc. They all can have different Type II errors.
- $1 -$ Type II error is called the power.

Summary

- Statistics is a tool to analyze data and find patterns
- However, statistics cannot provide a definitive answer
- Definitive answers come from understanding the science

Homework

- Further reading (textbook): Sections 11.3.3 and 11.3.4
  - "Quantitative methods for health research: a practical interactive guide to epidemiology and statistics" by Nigel Bruce, Daniel Pope, Debbi Stanistreet. Hoboken, NJ:Wiley, 2018 2nd edition. Wiley Online Library [Download Link]
- Install RStudio and R