SPH3101 Biostatistics for Public Health

Lecture 2

Hypothesis testing

(short lecture)

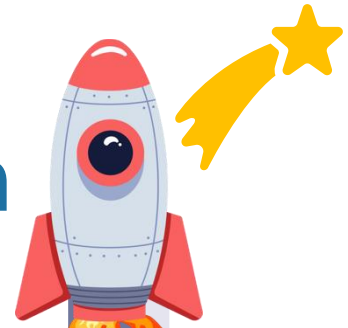
Kiesha Prem

Saw Swee Hock School of Public Health, National University of Singapore

TM-CM02

Biostatistics for Public Health

- quizzes
- 2 assignments



4

Logistic regression

Data prep, descriptive & inferential statistics, logistics regression and others if we have time

3

Linear regression

Data prep, descriptive & inferential statistics, linear regression

2

Hypothesis testing

Inferential statistics

1

Categorical data analysis

Data prep, descriptive & inferential statistics

Hypothesis testing

Steps for conducting significance tests:

1. **State the null hypothesis** (H_0)
2. **State the alternate hypothesis** (H_α)
3. **Calculate test statistic** (parameter of interest divided by standard error)
4. **Look up and interpret p-value:**
   - Remember that statistical significance is not equivalent to medical or biological significance!
   - Interpret a p-value in terms of the level of evidence (α) against the null hypothesis.

Hypothesis tests overview

NUS National University of Singapore | Saw Swee Hock School of Public Health

Purpose	Parametric	Non-parametric	Categorical data
Compare one group to a hypothetical value	One sample t-test	Signed Rank Test (Wilcoxon Test)	Sign Test (Binomial Test)
Compare two unpaired groups	Two sample t-test	Wilcoxon-Mann-Whitney Test	Chi Square Test Fisher's Exact Test
Compare two paired groups	Paired t-test	Signed Rank Test (Wilcoxon Test)	McNemar Test

Parametric tests

Hypothesis tests overview

Purpose	Parametric	Non-parametric	Categorical data
Compare one group to a hypothetical value	One sample t-test	Signed Rank Test (Wilcoxon Test)	Sign Test (Binomial Test)
Compare two unpaired groups	Two sample t-test	Wilcoxon-Mann Whitney Test	Chi-Square Test Fishers Exact Test
Compare two paired groups	Paired t-test	Signed Ra (Wilcoxo	mar Test

Let's focus on parametric tests for now

You only need to know 1 function:

`t.test()`

Students' t-tests: one sample

T-test

Null hypothesis

H_0: $\mu = \mu_0$

Alternative hypothesis

H_1: $\mu \neq \mu_0$

μ: true population mean

μ_0: hypothesised mean

What are some examples?

You will see more of this later

Students' t-tests: two independent sample (1)

NUS National University of Singapore | Saw Swee Hock School of Public Health

T-test

Null hypothesis

$H_0: \mu_1 = \mu_2$

Alternative hypothesis

$H_1: \mu_1 \neq \mu_2$

μ_1: mean of group 1

μ_2: mean of group 2

```
p1=log10(Platelet[DHFever=='Yes'])
p2=log10(Platelet[DHFever=='No'])
t.test(p1,p2)
```

```
## Welch Two Sample t-test
## data:  p1 and p2
## t = -7.4175, df = 559.01, p-value = 4.458e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2225354 -0.1293523
## sample estimates:
## mean of x mean of y
##  1.588267  1.764211
```

Students' t-tests: two independent sample (1)

T-test

Null hypothesis

$H_0: \mu_1 = \mu_2$

Alternative hypothesis

$H_1: \mu_1 \neq \mu_2$

μ_1: mean of group 1

μ_2: mean of group 2

```
p1=log10(Platelet[DHFever=='Yes'])
p2=log10(Platelet[DHFever=='No'])
t.test(p1,p2)
```

```
## Welch Two Sample t-test
## data:  p1 and p2
## t = -7.4175, df = 559.01, p-value = 4.458e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2225354 -0.1293523
## sample estimates:
## mean of x mean of y
##  1.588267  1.764211
```

What do these results mean?

Students' t-tests: two independent sample (2)

T-test

Null hypothesis

$H_0: \mu_1 = \mu_2$

Alternative hypothesis

$H_1: \mu_1 \neq \mu_2$

μ_1: mean of group 1

μ_2: mean of group 2

SAME
SAME
BUT
DIFFERENT

Same tests, different code – which do you prefer?

```
t.test(log10(Platelet)~DHFever)

##
##  Welch Two Sample t-test
##
## data:  log10(Platelet) by DHFever
## t = 7.4175, df = 559.01, p-value = 4.458e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1293523 0.2225354
## sample estimates:
## mean in group No mean in group Yes
##         1.764211          1.588267
```

Students' t-tests: paired sample

NUS National University of Singapore | Saw Swee Hock School of Public Health

T-test

Null hypothesis

$H_0: \mu_1 = \mu_2$

Alternative hypothesis

$H_1: \mu_1 \neq \mu_2$

μ_1: mean before

μ_2: mean after

If you have two measurements per individual and wish to test whether the means are the same, you can use a **paired t-test**.

The same function is used but with an additional `paired=TRUE` argument. Although there are no paired measurements in the dengue data that make sense to test, if there were platelet counts before and after treatment, the code would look like this:

```
t.test(plateletbefore,plateletafter,paired=TRUE)
```

Non-parametric tests

Hypothesis tests overview

Purpose	Parametric	Non-parametric	Categorical data
Compare one group to a hypothetical value	One sample t-test	Signed Rank Test (Wilcoxon Test)	Sign Test (Binomial Test)
Compare two unpaired groups	Two sample t-test	Wilcoxon-Mann-Whitney Test	Chi Square Test Fisher's Exact Test
Compare two paired groups	Paired t-test	Signed Rank Test (Wilcoxon Test)	McNemar Test

Non-parametric test: Wilcoxon-Mann-Whitney Test

T-test

Null hypothesis

$H_0: \mu_1 = \mu_2$

Alternative hypothesis

$H_1: \mu_1 \neq \mu_2$

μ_1: mean of group 1

μ_2: mean of group 2

```
p1=log10(Platelet[DHFever=='Yes'])
p2=log10(Platelet[DHFever=='No'])
```

```
wilcox.test(p1,p2)


##
##  Wilcoxon rank sum test with continuity correction
##
## data:  p1 and p2
## W = 49492, p-value = 1.15e-12
## alternative hypothesis: true location shift is not equal to 0
```

What do these results mean?

Non-parametric test: Wilcoxon-Mann-Whitney Test

NUS National University of Singapore | Saw Swee Hock School of Public Health

T-test

Null hypothesis

H_0: $f_1 = f_2$

Alternative hypothesis

H_1: $f_1 \neq f_2$

μ_1: distribution of group 1

μ_2: distribution of group 2

```
wilcox.test(log10(Platelet)~DHFever)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  log10(Platelet) by DHFever
## W = 93287, p-value = 1.15e-12
## alternative hypothesis: true location shift is not equal to 0
```

SAME
SAME
BUT
DIFFERENT

Non-parametric test: Wilcoxon-Mann-Whitney Test

NUS National University of Singapore | Saw Swee Hock School of Public Health

T-test

Null hypothesis

H_0: $f_1 = f_2$

Alternative hypothesis

H_1: $f_1 \neq f_2$

f_1: distribution of group 1

f_2: distribution of group 2

Even when you did not **log transform** your data

```
wilcox.test(Platelet~DHFever)
```

```
wilcox.test(log10(Platelet)~DHFever)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  log10(Platelet) by DHFever
## W = 93287, p-value = 1.15e-12
## alternative hypothesis: true location shift is not equal to 0
```

SAME
SAME
BUT
DIFFERENT

Non-parametric test: paired sample Signed Rank Test

T-test

Null hypothesis

H_0: $f_1 = f_2$

Alternative hypothesis

H_1: $f_1 \neq f_2$

f_1: distribution before

f_2: distribution after

If you have two measurements per individual and wish to test whether the distribution are the same, you can use a **paired t-test**.

The same function is used but with an additional `paired=TRUE` argument. Although there are no paired measurements in the dengue data that make sense to test, if there were platelet counts before and after treatment, the code would look like this:

Parametric

```
t.test(plateletbefore,plateletafter,paired=TRUE)
```

Non-parametric

```
wilcox.test(plateletbefore,plateletafter,paired=TRUE)
```

Tests for categorical data

Hypothesis tests overview

Purpose	Parametric	Non-parametric	Categorical data
Compare one group to a hypothetical value	One sample t-test	Signed Rank Test (Wilcoxon Test)	Sign Test (Binomial Test)
Compare two unpaired groups	Two sample t-test	Wilcoxon-Mann-Whitney Test	Chi Square Test Fisher's Exact Test
Compare two paired groups	Paired t-test	Signed Rank Test (Wilcoxon Test)	McNemar Test

Chi-square tests for categorical data


```
contable = table(DHFever,Sex)
print(contable)
```

```
##         Sex
## DHFever Female Male
##      No     149  374
##      Yes    109  164
```

```
chisq.test(contable)
```

What do these results mean?

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  contable
## X-squared = 10.195, df = 1, p-value = 0.001408
```

Fisher's Exact tests for categorical data


```
contable = table(DHFever,Sex)
print(contable)
```

```
##         Sex
## DHFever Female Male
##      No     149  374
##      Yes    109  164
```

```
fisher.test(contable)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  contable
## p-value = 0.001391
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.4357820 0.8259849
## sample estimates:
## odds ratio
##  0.5998013
```

What do these results mean?

How does it compare to the Chi-square test?

Non-parametric test: paired sample McNemar's Test


```
mcnemar.test(x, y = NULL, correct = TRUE)
```

The McNemar's test requires a contingency table with counts of paired observations falling into four categories:

1. agreement on both measurements,
2. disagreement on both measurements,
3. agreement on the first but not the second, and
4. agreement on the second but not the first.

Non-parametric test: paired sample McNemar's Test


```
mcnemar.test(x, y = NULL, correct = TRUE)
```

```
# Create a 2x2 matrix representing the contingency table
data <- matrix(c(30, 12, 20, 25), nrow = 2,
               dimnames = list("After Method B" = c("Passed", "Failed"),
                               "After Method A" = c("Passed", "Failed")))
```

```
                After Method A
After Method B Passed Failed
        Passed     30     20
        Failed     12     25
```

What do these results mean?

```
> mcnemar.test(data)

        McNemar's Chi-squared test with continuity correction

data:  data
McNemar's chi-squared = 1.5312, df = 1, p-value = 0.2159
```

Non-parametric test: paired sample McNemar's Test


```
mcnemar.test(x, y = NULL, correct = TRUE)
```

```
# Create a 2x2 matrix representing the contingency table
data <- matrix(c(30, 12, 20, 25), nrow = 2,
               dimnames = list("After Method B" = c("Passed", "Failed"),
                               "After Method A" = c("Passed", "Failed")))
```

```
                After Method A
After Method B Passed Failed
        Passed     30     20
        Failed     12     25
```

The McNemar's test helps assess whether there's a significant difference in outcomes between Method A and Method B.

```
> mcnemar.test(data)

        McNemar's Chi-squared test with continuity correction

data:  data
McNemar's chi-squared = 1.5312, df = 1, p-value = 0.2159
```

The **p-value** suggests that there's no significant difference between the two methods, as it is above the significance level of 0.05.

How do you know what to do?

1) Check for normality (Gaussian distribution)

- Plot density
- QQ plots
- Statistical Tests – Shapiro Test, Kolmogorov-Smirnov test

2) Check if there are EQUAL variances

- Test to compare variances – F Test

Hypothesis tests overview

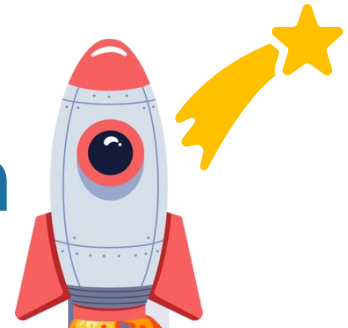
1. Check for normality

Purpose	Parametric	Non-parametric	Categorical data
Compare one group to a hypothetical value	One sample t-test	Signed Rank Test (Wilcoxon Test)	Sign Test (Binomial Test)
Compare two unpaired groups	Two sample t-test	Wilcoxon-Mann-Whitney Test	Chi Square Test Fisher's Exact Test
Compare two paired groups	Paired t-test	Signed Rank Test (Wilcoxon Test)	McNemar Test

TM-CM02

Biostatistics for Public Health

- quizzes
- 2 assignments



4

Logistic regression

Data prep, descriptive & inferential statistics, logistics regression and others if we have time

3

Linear regression

Data prep, descriptive & inferential statistics, linear regression

2

Hypothesis testing

Inferential statistics

1

Categorical data analysis

Data prep, descriptive & inferential statistics

Thank you

Appendices

Why GitHub?

One of the easiest ways to use GitHub is through GitHub Desktop without directly using Git commands:

1. **Sign up for a GitHub account:** https://github.com/

2. **Download GitHub Desktop:** Visit https://desktop.github.com/download/ and download the appropriate version for your operating system (Windows or macOS).

3. **Install the application:** Follow the installation instructions specific to your operating system.

4. **Sign in to GitHub:** Once installed, open the GitHub Desktop application. You will be prompted to sign in to your GitHub account (step 1).

5. In the next in-person practical session, we will review how to clone an existing repository, create a new repository, make changes to files, commit changes and collaborate with others.

GitHub account and GitHub Desktop

- ✓ GitHub is a platform for version control and **collaboration**.
- ✓ It allows multiple people to work on projects **simultaneously**.
- ✓ Helps in **tracking** changes, managing versions, and working with others.

Working on GitHub

Cloning a Repository

Steps to Clone:

- Go to File > Clone Repository in GitHub Desktop
- Choose a repository from GitHub or enter the repository URL
- Select the local path where you want to store the repository

Creating a New Repository

Steps to Create:

- Go to File > New Repository
- Name your repository and select a local path
- [Optional but recommended] Initialise with a README file

Making changes & committing

Making changes

- Edit files in your repository using your preferred text editor.
- GitHub Desktop automatically detects changes.

Committing changes

- Review changes in GitHub Desktop.
- Write a descriptive commit message.
- Click Commit to main (or your branch name).

Pushing Changes to GitHub

Steps to Push

- Click the Push origin button after committing.
- This uploads your local commits to the GitHub repository

Working with branches

Creating a Branch

- Go to Branch > New Branch.
- Name the branch and create it.

Switching Branches

- Use the branch dropdown to switch between branches.

Merging Branches

- Merge branches via Branch > Merge into in GitHub Desktop.

Branching allows for the development of new features without affecting the main codebase.

Collaborating with others

Creating pull requests

- Click Branch > Create Pull Request.
- Review and submit Pull Requests on GitHub in your web browser.

Reviewing changes

- Use GitHub's web interface to review and discuss changes.

Sync changes by pulling updates

- Click **Fetch** origin or **Pull** origin to get the latest changes from GitHub.
- Resolve conflicts if any appear during a pull or merge.

- ❑ Keep commits small and frequent
- ❑ Write descriptive commit messages
- ❑ Regularly push changes

Thank you