

Multiple testing (in R)

Catalina Vallejos (The Alan Turing Institute and UCL)
Aaron Lun (Cancer Research UK - Cambridge Institute)

April 5th, 2017

Outline for the course

- ▶ Introduction to statistics
- ▶ Hypothesis testing (in R)
- ▶ Regression analysis (in R)
- ▶ Multiple testing (in R)

Hypothesis testing

Often, the aim of the analysis is not only **estimation**.

For example,

- ▶ *Is the treatment effective?*
- ▶ *Is a gene differentially expressed?*
- ▶ *Is there an association between genotype and phenotype?*



These questions can be translated as **hypothesis testing** problems

Hypothesis testing

No test is exact:

	Null hypothesis does not hold	Null hypothesis holds
Reject null hypothesis	Correct True positive	Wrong False positive
Do not reject null hypothesis	Wrong False negative	Correct True negative

Hypothesis testing

No test is exact:

	Null hypothesis does not hold	Null hypothesis holds
Reject null hypothesis	$1 - \beta$	α
Do not reject null hypothesis	β	$1 - \alpha$

Hypothesis testing

No test is exact:

	Null hypothesis does not hold	Null hypothesis holds
Reject null hypothesis	$1 - \beta$	α
Do not reject null hypothesis	β	$1 - \alpha$

What happens when we run more than one test?

Multiple testing

There are many examples where **multiple tests** are required:

Multiple testing

There are many examples where **multiple tests** are required:

For example, Genome Wide Association Studies (GWAS):

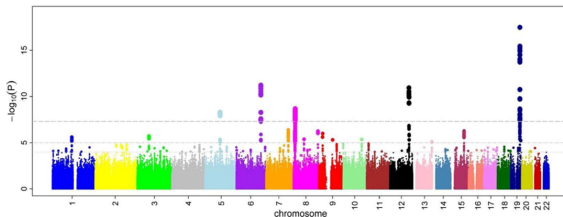
- ▶ Usually framed as a regression problem
- ▶ Outcome: disease (yes/no), biomarker value, height, etc.
- ▶ Covariates: SNPs in the genome (millions!)

Multiple testing

There are many examples where **multiple tests** are required:

For example, Genome Wide Association Studies (GWAS):

- ▶ Usually framed as a regression problem
- ▶ Outcome: disease (yes/no), biomarker value, height, etc.
- ▶ Covariates: SNPs in the genome (millions!)



Multiple testing

One approach to GWAS is to perform *one-at-the-time-tests*

Multiple testing

One approach to GWAS is to perform *one-at-the-time-tests*

Notation:

- ▶ Y_i : response variable (e.g. biomarker value) for person i
- ▶ $\text{SNP}_i^{(j)}$: observed value for SNP j , person i

Typically codified as $\text{SNP}_i^{(j)} = 0, 1, 2$ for e.g. AA, AT, TT

Multiple testing

One approach to GWAS is to perform *one-at-the-time-tests*

Notation:

- ▶ Y_i : response variable (e.g. biomarker value) for person i
- ▶ $\text{SNP}_i^{(j)}$: observed value for SNP j , person i

Typically codified as $\text{SNP}_i^{(j)} = 0, 1, 2$ for e.g. AA, AT, TT

This can be formulated using simple linear regressions:

$$Y_i = \alpha_j + \beta_j \text{SNP}_i^{(j)} + \epsilon_i, \quad j = 1, \dots, J$$

Multiple testing

One approach to GWAS is to perform *one-at-the-time-tests*

Notation:

- ▶ Y_i : response variable (e.g. biomarker value) for person i
- ▶ $\text{SNP}_i^{(j)}$: observed value for SNP j , person i

Typically codified as $\text{SNP}_i^{(j)} = 0, 1, 2$ for e.g. AA, AT, TT

This can be formulated using simple linear regressions:

$$Y_i = \alpha_j + \beta_j \text{SNP}_i^{(j)} + \epsilon_i, \quad j = 1, \dots, J$$

Is there a significant association between SNP j and the outcome?

Multiple testing

Is there a significant association between SNP j and the outcome?

This can be formulated as the following hypothesis test:

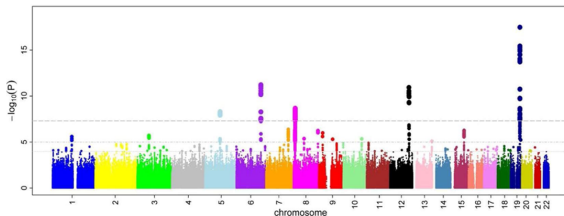
$$H_0 : \beta_j = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0$$

Multiple testing

Is there a significant association between SNP j and the outcome?

This can be formulated as the following hypothesis test:

$$H_0 : \beta_j = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0$$



There are as many tests as SNPs in the genome!

Multiple testing

No test is exact:

	Null hypothesis does not hold	Null hypothesis holds
Reject null hypothesis	$1 - \beta$	α
Do not reject null hypothesis	β	$1 - \alpha$

What happens to these errors when we run more multiple tests?

Multiple testing

What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

Multiple testing

What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

P(At least 1 FP in T tests)

Multiple testing

What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

$$\begin{aligned} & \text{P(At least 1 FP in } T \text{ tests)} \\ = & 1 - \text{P(No FPs in } T \text{ tests)} \end{aligned}$$

Multiple testing

What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

$$\begin{aligned} & \text{P(At least 1 FP in } T \text{ tests)} \\ = & 1 - \text{P(No FPs in } T \text{ tests)} \\ = & 1 - \text{P(No FP in test 1)} \times \cdots \times \text{P(No FP in test } T) \end{aligned}$$

Multiple testing

What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

$$\begin{aligned} & \text{P(At least 1 FP in } T \text{ tests)} \\ = & 1 - \text{P(No FPs in } T \text{ tests)} \\ = & 1 - \text{P(No FP in test 1)} \times \cdots \times \text{P(No FP in test } T) \\ = & 1 - (1 - \alpha) \times \cdots \times (1 - \alpha) \end{aligned}$$

Multiple testing

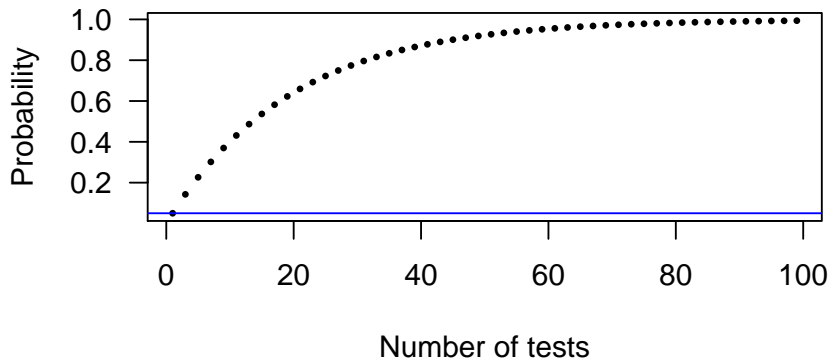
What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

$$\begin{aligned} & \text{P(At least 1 FP in } T \text{ tests)} \\ = & 1 - \text{P(No FPs in } T \text{ tests)} \\ = & 1 - \text{P(No FP in test 1)} \times \cdots \times \text{P(No FP in test } T) \\ = & 1 - (1 - \alpha) \times \cdots \times (1 - \alpha) \\ = & 1 - (1 - \alpha)^T \end{aligned}$$

Multiple testing

What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

$$P(\text{At least 1 FP in } T \text{ tests}) = 1 - (1 - \alpha)^T; \quad \alpha = 0.05$$



For $T = 100$ tests, the probability of at least 1 FP is almost 1!

Multiple testing

What is the probability of obtaining 1 or more false positives (i.e. Reject H_0 when H_0 is true) among T independent tests?

We need to correct for multiple testing to control error rates

Multiple testing: Bonferroni's correction

Multiple testing: Bonferroni's correction

To control the **Family-Wise Error Rate** (FWER)

$$\text{FWER} = P(\text{At least 1 FP in } T \text{ tests})$$

Idea: for T tests with significance level α , it can be shown that

$$\text{FWER} \leq T\alpha$$

Multiple testing: Bonferroni's correction

To control the **Family-Wise Error Rate** (FWER)

$$\text{FWER} = P(\text{At least 1 FP in } T \text{ tests})$$

Idea: for T tests with significance level α , it can be shown that

$$\text{FWER} \leq T\alpha$$

Bonferroni's solution is to use:

$$\alpha^* = \min\{\alpha/T, 1\}$$

Doing this for each test ensures that $\text{FWER} \leq \alpha$

Multiple testing: Bonferroni's correction

Bonferroni's correction:

$$\alpha^* = \min\{\alpha/T, 1\}$$

For example, for 1 million tests and $\alpha = 0.05$, this leads to

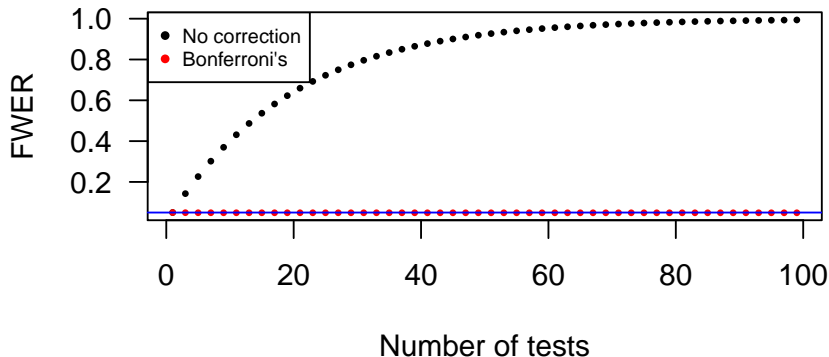
$$\alpha^* = 5 \times 10^{-8}$$

This is commonly used as a p-value threshold in GWAS

Multiple testing: Bonferroni's correction

Bonferroni's correction:

$$\alpha^* = \min\{\alpha/T, 1\}$$



Multiple testing: Bonferroni's correction

- ▶ It can be too conservative \Rightarrow too many false negatives
- ▶ Holm's method is less conservative while still controlling FWER

Multiple testing

Is FWER the right quantity to control?

- ▶ Yes, if we only care about false positives
- ▶ No, if we care about false negatives (at the cost of false positives)

Multiple testing

Is FWER the right quantity to control?

- ▶ Yes, if we only care about false positives
- ▶ No, if we care about false negatives (at the cost of false positives)

An alternative idea is to control the **False Discovery Rate** (FDR)

$$\text{FDR} = \frac{\text{FP}}{\text{Positives}}$$

Multiple testing

Is FWER the right quantity to control?

- ▶ Yes, if we only care about false positives
- ▶ No, if we care about false negatives (at the cost of false positives)

An alternative idea is to control the **False Discovery Rate** (FDR)

$$\text{FDR} = \frac{\text{FP}}{\text{Positives}} = \frac{\text{FP}}{\text{FP} + \text{TP}}$$

Multiple testing

Is FWER the right quantity to control?

- ▶ Yes, if we only care about false positives
- ▶ No, if we care about false negatives (at the cost of false positives)

An alternative idea is to control the **False Discovery Rate** (FDR)

$$\text{FDR} = \frac{\text{FP}}{\text{Positives}} = \frac{\text{FP}}{\text{FP} + \text{TP}}$$

Do not confuse with the False Positive Rate! $\text{FPR} = \frac{\text{FP}}{\text{True positives}}$

Multiple testing

Is FWER the right quantity to control?

- ▶ Yes, if we only care about false positives
- ▶ No, if we care about false negatives (at the cost of false positives)

An alternative idea is to control the **False Discovery Rate** (FDR)

$$\text{FDR} = \frac{\text{FP}}{\text{Positives}} = \frac{\text{FP}}{\text{FP} + \text{TP}}$$

Do not confuse with the False Positive Rate! $\text{FPR} = \frac{\text{FP}}{\text{True positives}}$

Benjamini and Hochberg's method controls the FDR

Questions + practical