Lecture 12: Multiple Testing

Module 3: part 5

Spring 2025

Logistics

- End of Module 3 today
- Lab will cover Hypothesis testing
- Module 3 assessment up today, due Mar 16

Hypothesis Testing Errors: Types and Consequences

	Null Not Rejected	Null Rejected
Null is True	Correct Decision (Probability $1-\alpha$)	Type I Error (Probability α)
Null is False	Type II Error (Probability β)	Correct Decision (Power $= 1-\beta$)

Type I Error (False Positive)

- Rejecting H_0 when it is actually true
- Probability controlled by significance level α
- Typically set to 0.05 (5%)
- More "costly" in most research contexts
- Example: Claiming a drug works when it doesn't

Type II Error (False Negative)

- Failing to reject H_0 when it is actually false
- Probability denoted as β
- Depends on:
  - True effect size
  - Sample size
  - Data variability
  - Significance level α

Hypothesis Testing for a Single Regression Coefficient

Hypothesis Formulation

$$H_0 : \beta_1 = 0 \quad \text{(Null Hypothesis)}$$
$$H_A : \beta_1 \neq 0 \quad \text{(Alternative Hypothesis)}$$

Test Statistic

$$t = \frac{\hat{\beta}_1}{\text{SE}(\hat{\beta}_1)} = \frac{\text{Point Estimate}}{\text{Standard Error}}$$

Interpretation of the test statistic:

- Measures how many standard errors the estimate is from zero
- Large absolute values suggest the coefficient is significantly different from zero

Decision rule:

- Reject H_0 if $|t| > t_{\alpha/2,n-p-1}$
- Equivalent to p-value $< \alpha$

Distribution under H_0:

- Follows a t-distribution with $n - p - 1$ degrees of freedom
- n: number of observations
- p: number of predictors

F-test: Hypothesis Testing in Multiple Regression

Suppose the data is generated by the following regression model:

$$Y_i = b_0 + b_1 X_{i,1} + b_2 X_{i,2} + b_3 X_{i,3} + b_4 X_{i,4} + \ldots + b_p X_{i,p} + \varepsilon_i$$

Hypothesis Formulation

- Null Hypothesis H_0: $b_1 = b_2 = b_3 = 0$
- Alternative Hypothesis H_1: At least one $b_j \neq 0$ for $j \in \{1, 2, 3\}$

F-test: Hypothesis Testing in Multiple Regression

Suppose the data is generated by the following regression model:

$$Y_i = b_0 + b_1 X_{i,1} + b_2 X_{i,2} + b_3 X_{i,3} + b_4 X_{i,4} + \ldots + b_p X_{i,p} + \varepsilon_i$$

Hypothesis Formulation

- Null Hypothesis H_0: $b_1 = b_2 = b_3 = 0$
- Alternative Hypothesis H_1: At least one $b_j \neq 0$ for $j \in \{1, 2, 3\}$

Model Comparison Approach:

- **Alternative Model** (Full Model):

$$Y_i = \tilde{b}_0 + \tilde{b}_1 x_{i,1} + \tilde{b}_2 x_{i,2} + \tilde{b}_3 x_{i,3} + \tilde{b}_4 x_{i,4} + \ldots + \tilde{b}_p x_{i,p} + \varepsilon_i$$

- **Null Model** (Restricted Model):

$$Y_i = \hat{b}_0 + 0 \cdot x_{i,1} + 0 \cdot x_{i,2} + 0 \cdot x_{i,3} + \hat{b}_4 x_{i,4} + \ldots + \hat{b}_p x_{i,p} + \varepsilon_i$$

F-test: Hypothesis Testing in Multiple Regression

Suppose the data is generated by the following regression model:

$$Y_i = b_0 + b_1 X_{i,1} + b_2 X_{i,2} + b_3 X_{i,3} + b_4 X_{i,4} + \ldots + b_p X_{i,p} + \varepsilon_i$$

Hypothesis Formulation

- Null Hypothesis H_0: $b_1 = b_2 = b_3 = 0$
- Alternative Hypothesis H_1: At least one $b_j \neq 0$ for $j \in \{1, 2, 3\}$

Model Comparison Approach:

- **Alternative Model** (Full Model):

$$Y_i = \tilde{b}_0 + \tilde{b}_1 x_{i,1} + \tilde{b}_2 x_{i,2} + \tilde{b}_3 x_{i,3} + \tilde{b}_4 x_{i,4} + \ldots + \tilde{b}_p x_{i,p} + \varepsilon_i$$

- **Null Model** (Restricted Model):

$$Y_i = \hat{b}_0 + 0 \cdot x_{i,1} + 0 \cdot x_{i,2} + 0 \cdot x_{i,3} + \hat{b}_4 x_{i,4} + \ldots + \hat{b}_p x_{i,p} + \varepsilon_i$$

F-statistic:

$$F = \frac{[RSS(Null) - RSS(Alt)]/(p_{\text{alt}} - p_{\text{null}})}{RSS(Alt)/(n - p_{\text{alt}} - 1)}$$

Multiple Hypothesis Testing: Motivation and Challenges

What are Multiple Hypothesis Tests?

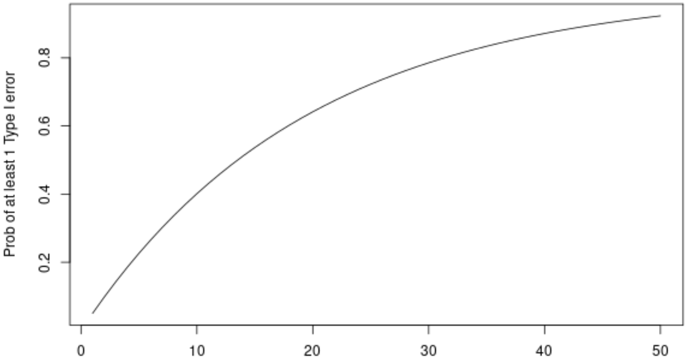
- Scenarios involving simultaneous testing of numerous hypotheses
- Prevalent in fields with large-scale data exploration

Multiple Hypothesis Testing: Motivation and Challenges

What are Multiple Hypothesis Tests?

- Scenarios involving simultaneous testing of numerous hypotheses
- Prevalent in fields with large-scale data exploration

Even when the Type I error is controlled for each specific test, false positives may happen frequently when testing many hypotheses



Multiple testing

- Instead of controlling Type I error for each individual test, we might try to control the Family-wise error rate
- **Family-wise error rate**: What is the probability that at least 1 false positive occurs?

$$P(\text{at least one false positive}) \leq \alpha$$

Bonferroni Procedure: A Conservative Approach

Basic Principle

When conducting m simultaneous hypothesis tests, adjust the significance level to:

$$\text{New Significance Level} = \frac{\alpha}{m}$$

Practical Example:

- Original significance level: $\alpha = 0.05$
- Number of simultaneous tests: $m = 1000$
- Bonferroni-corrected significance level: $0.05/1000 = 0.00005$

Bonferroni Procedure: A Conservative Approach

Basic Principle

When conducting m simultaneous hypothesis tests, adjust the significance level to:

$$\text{New Significance Level} = \frac{\alpha}{m}$$

Practical Example:

- Original significance level: $\alpha = 0.05$
- Number of simultaneous tests: $m = 1000$
- Bonferroni-corrected significance level: $0.05/1000 = 0.00005$

Equivalent P-value Adjustment:

- Original p-value: p_i
- Adjusted p-value: $\tilde{p}_i = m \times p_i$
- Reject null hypothesis if $\tilde{p}_i < \alpha$

Bonferroni Procedure: A Conservative Approach

Basic Principle

When conducting m simultaneous hypothesis tests, adjust the significance level to:

$$\text{New Significance Level} = \frac{\alpha}{m}$$

Practical Example:

- Original significance level: $\alpha = 0.05$
- Number of simultaneous tests: $m = 1000$
- Bonferroni-corrected significance level: $0.05/1000 = 0.00005$

Equivalent P-value Adjustment:

- Original p-value: p_i
- Adjusted p-value: $\tilde{p}_i = m \times p_i$
- Reject null hypothesis if $\tilde{p}_i < \alpha$

Key Characteristics:

- Extremely conservative method
- Guarantees control of Family-Wise Error Rate (FWER)
- Dramatically reduces the chance of Type I errors
- Comes at the cost of statistical power

Holm Procedure: Controlling Family-Wise Error Rate (FWER)

Purpose: Control the probability of making at least one Type I error when performing multiple hypothesis tests

Procedure Steps

Given m total hypotheses and significance level α:

1. Sort p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
2. Sequentially test hypotheses using adjusted significance levels:
   - First test: $p_{(1)} < \frac{\alpha}{m}$
   - Second test: $p_{(2)} < \frac{\alpha}{m-1}$
   - j-th test: $p_{(j)} < \frac{\alpha}{m-j+1}$

FWER: Balancing Type I Error Control and Statistical Power

Key Considerations

- **Strict Criterion:** Minimizing the probability of at least one Type I error
- **Trade-off:** Stringent control comes at the cost of reduced statistical power

Advantages of FWER Control

- Prevents false discoveries
- Provides strong Type I error protection

Limitations

- Overly conservative
- Reduces ability to detect true effects

FWER: Balancing Type I Error Control and Statistical Power

Key Considerations

- **Strict Criterion:** Minimizing the probability of at least one Type I error
- **Trade-off:** Stringent control comes at the cost of reduced statistical power

Advantages of FWER Control

- Prevents false discoveries
- Provides strong Type I error protection

Limitations

- Overly conservative
- Reduces ability to detect true effects

Comparative Methods:

- Holm procedure offers more power than Bonferroni
- Suggests a potential compromise between error control and statistical power

FWER: Balancing Type I Error Control and Statistical Power

Key Considerations

- **Strict Criterion:** Minimizing the probability of at least one Type I error
- **Trade-off:** Stringent control comes at the cost of reduced statistical power

Advantages of FWER Control

- Prevents false discoveries
- Provides strong Type I error protection

Limitations

- Overly conservative
- Reduces ability to detect true effects

Comparative Methods:

- Holm procedure offers more power than Bonferroni
- Suggests a potential compromise between error control and statistical power

Emerging Question: Can we develop methods that allow more Type I errors while maintaining reasonable control?

False Discovery Rate: Quantifying Multiple Testing Errors

Contingency Table of Hypothesis Testing

	Rejected	Not Rejected	Total
Null True	A	B	A + B
Null False	C	D	C + D
Total	A + C	B + D	m

Error Metrics

- **False Positive Rate:**

$$\frac{A}{A+B}$$

  Proportion of true nulls incorrectly rejected

- **False Discovery Rate:**

$$\frac{A}{A+C}$$

  Proportion of rejected hypotheses that are false

False Discovery Rate: Quantifying Multiple Testing Errors

Contingency Table of Hypothesis Testing

	Rejected	Not Rejected	Total
Null True	A	B	A + B
Null False	C	D	C + D
Total	A + C	B + D	m

Error Metrics

- **False Positive Rate:**

$$\frac{A}{A+B}$$

  Proportion of true nulls incorrectly rejected

- **False Discovery Rate:**

$$\frac{A}{A+C}$$

  Proportion of rejected hypotheses that are false

Goal: Control the expected proportion of false discoveries

$$E\left(\frac{A}{A+C}\right)$$

Benjamini-Hochberg Procedure: Controlling False Discovery Rate

Setup: Testing m hypotheses with significance level α

Procedure Steps

1. Sort p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
2. Sequentially test hypotheses:
    - Reject if $p_{(j)} \leq \frac{j}{m}\alpha$
    - Stop when $p_{(j)} > \frac{j}{m}\alpha$

Benjamini-Hochberg Procedure: Controlling False Discovery Rate

Setup: Testing m hypotheses with significance level α

Procedure Steps

1. Sort p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
2. Sequentially test hypotheses:
   - Reject if $p_{(j)} \leq \frac{j}{m}\alpha$
   - Stop when $p_{(j)} > \frac{j}{m}\alpha$

Alternative Formulation:

- Adjusted p-values: $\tilde{p}_{(j)} = p_{(j)}(m/j)$
- Reject for $j = 1, 2, \ldots, J$ where J is the largest index with $\tilde{p}_{(j)} < \alpha$

Benjamini-Hochberg Procedure: Controlling False Discovery Rate

Setup: Testing m hypotheses with significance level α

Procedure Steps

1. Sort p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
2. Sequentially test hypotheses:
   - Reject if $p_{(j)} \leq \frac{j}{m}\alpha$
   - Stop when $p_{(j)} > \frac{j}{m}\alpha$

Alternative Formulation:

- Adjusted p-values: $\tilde{p}_{(j)} = p_{(j)}(m/j)$
- Reject for $j = 1, 2, \ldots, J$ where J is the largest index with $\tilde{p}_{(j)} < \alpha$

Key Assumption:

- Assumes independence between hypothesis tests
- Works best when tests are not strongly correlated

Benjamini-Yekutieli Procedure: Robust FDR Control

Purpose: Control False Discovery Rate without assuming test independence

Procedure Details

- Sort p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- Rejection criterion:

$$p_{(j)} \leq \frac{j}{m \sum_{i=1}^{m} i^{-1}} \alpha$$

- Continue until first $p_{(j)}$ fails this condition

Benjamini-Yekutieli Procedure: Robust FDR Control

Purpose: Control False Discovery Rate without assuming test independence

Procedure Details

- Sort p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- Rejection criterion:

$$p_{(j)} \leq \frac{j}{m \sum_{i=1}^{m} i^{-1}} \alpha$$

- Continue until first $p_{(j)}$ fails this condition

Key Characteristics:

- Works with *dependent* and *independent* hypothesis tests
- More conservative than Benjamini-Hochberg procedure
- Accounts for potential correlations between tests

Benjamini-Yekutieli Procedure: Robust FDR Control

Purpose: Control False Discovery Rate without assuming test independence

Procedure Details

- Sort p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- Rejection criterion:

$$p_{(j)} \leq \frac{j}{m \sum_{i=1}^{m} i^{-1}} \alpha$$

- Continue until first $p_{(j)}$ fails this condition

Key Characteristics:

- Works with *dependent* and *independent* hypothesis tests
- More conservative than Benjamini-Hochberg procedure
- Accounts for potential correlations between tests

Computational Insight:

- Introduces a correction factor $\sum_{i=1}^{m} i^{-1}$
- Provides more robust error rate control across various testing scenarios

Summary

- Controlling the Type I error can come at the cost of lower power
- When testing many hypothesis, the number of false positives can grow
- Controlling family wise error rate ensures no false positives with high probability, but increases Type II error
- Controlling False Discovery Rate allow some (but not too many) false positives and does not lose as much power