IEOR165 Discussion
Week 12

Sheng Liu

University of California, Berkeley

Apr 15, 2016

Outline

Type I errors & Type II errors

Table: Error types in hypothesis testing

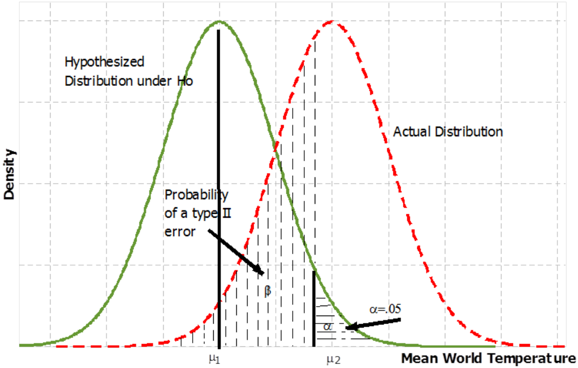
Decision	Null Hypothesis TRUE	Null Hypothesis FALSE
Reject	Type I	Correct
Fail to reject	Correct	Type II

- Type I error
  - Reject H_0 when it is true, also called false positive
  - $P(\text{Type I error}) = P(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$
- Type II error
  - Fail to reject H_0 when it is false, also called false negative
  - $P(\text{Type II error}) = P(\text{fail to reject } H_0 | H_0 \text{ is false}) = \beta$

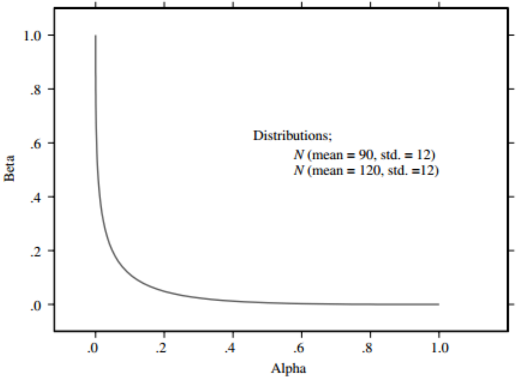
Tradeoff

- We want to perform a test that yields low errors: both type I and type II errors
- However, choosing a lower level of type I level will increase the chance of type II error. Intuitively, if it is hard to reject a true null hypothesis (low type I error), it is also very likely that a false null hypothesis is not rejected (high type II error)
- When fixing the significance level, are we restricting the probability of type I error or Type II error?
- Only way to reduce both type I and type II errors: **increase the sample size**

Probability of Type I and Type II errors



Tradeoff



Reference: James B. Ramsey, 1998
`http://www.econ.nyu.edu/user/ramseyj/textbook/chapter11.pdf`

Familywise Error Rate (FWER)

Let fix the significance level as $\alpha = 0.05$

- When doing a single hypothesis testing, the probability of not making a Type I error is

$$1 - \alpha = 1 - 0.05 = 0.95$$

- When doing k hypothesis testings at the same time, the probability of not making a type I error on all k tests is

$$(1 - \alpha)^k = 0.95^k$$

  - $0.95^4 = 0.814$, $0.95^6 = 0.735$
  - You need to improve on this

- The probability of making one or more Type I errors on the family of tests (familywise error rate), with the assumption of independence

$$FWER = 1 - (1 - \alpha)^k$$

Bonferroni Correction

We want to bound our familywise error rate by α_F (differentiate it with error rate per test α_T), the key question is, if we set $\alpha_F = 0.05$, what should be our α_T?

- Directly from the equation (also called Šidàk equation)

$$\alpha_T = 1 - (1 - \alpha_F)^{1/k}$$

- Bonferroni approximation (Taylor expansion)

$$\alpha_T \approx \frac{\alpha_F}{k}$$

- $\alpha_F = 0.05$ and $k = 4$:

$$\alpha_T = 1 - (1 - 0.05)^{1/4} = 0.0127$$

$$\alpha_T = \frac{\alpha_F}{k} = \frac{0.05}{4} = 0.0125$$

From Bonferroni to Holm-Bonferroni

- Compare p_i with α_F/k, which is equal to comparing $k \cdot p_i$ with α_F
- Bonferroni correction actually drops the independence assumption and thus is very general but conservative. Requiring each test at the significance level of α/k is too strict and may lead to higher type II errors
- A more complicated but more powerful test is Holm-Bonferroni test:
  1. Sort p-values from the smallest to the largest.
  2. Compare $p_{(1)}$ with α/k, if $p_{(1)} > \alpha/k$, **accept all null hypothesis.**
  3. Otherwise we continue to compare $p_{(i)}$ and $\alpha/(k-i+1)$: if $p_{(i)} > \alpha/(k-i+1)$, we stop; otherwise we **reject** $H_{(i)}$ and move on.
- Intuitively, you can see why this method is less conservative

Example

We want to test three hypothesis at the same time, their p-values are 0.0004, 0.0201 and 0.0612. Use both the Bonferroni correction and Holm-Bonferroni method.

Basics

We have k groups of data and each group j has n_j observations. Let $\overline{X}^j$ be the sample average in group j and $\overline{X}$ be the sample average of all observations.

- The total sum of squares

$$SS_T = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_i^j - \overline{X})^2$$

- The between group sum of squares (between groups variation)

$$SS_B = \sum_{j=1}^{k} n_j (\overline{X}^j - \overline{X})^2$$

- The within groups sum of squares (within groups variation)

$$SS_W = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_i^j - \overline{X}^j)^2$$

Basics

- We can show that

$$SS_T = SS_B + SS_W$$

- And similar equation applies to degrees of freedom

$$df(SS_T) = df(SS_B) + df(SS_W)$$

$$N - 1 = k - 1 + N - k$$

Now assume these groups have the same variance σ^2 (unknown), we have

$$E(SS_W/(N-k)) = \sigma^2$$

If group means are the same, we have

$$E(SS_B/(k-1)) = \sigma^2$$

ANOVA

Berkeley UNIVERSITY OF CALIFORNIA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad (\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2)$$

Under the null hypothesis, we have

$$\frac{SS_B/(k-1)}{SS_W/(N-k)} = \frac{MSG}{MSE} \sim F(k-1, N-k)$$

When the null hypothesis is true, this ratio should be close to one. In the contrary, if their means are different (there is a treatment effect), this ratio should be greater than 1:

$$p - value = P(F > \frac{SS_B/(k-1)}{SS_W/(N-k)})$$

We only do one-tailed test here, why?

Example

Low Calorie	Low Fat	Low Carbohydrate
8	2	5
9	4	7
4	3	
	3	

Calculate $SS_T, SS_B, SS_W, MSG, MSE$ and test whether the means of these three groups are the same or not (assume normal population and the same variance across groups)

References

ANOVA:

- http://www.lboro.ac.uk/media/wwwlboroacuk/content/mlsc/downloads/1.5_OnewayANOVA.pdf
- http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_HypothesisTesting-ANOVA/BS704_HypothesisTesting-Anova_print.html

Holm and Bonferroni:

- https://www.utdallas.edu/~herve/abdi-Holm2010-pretty.pdf