A Brief Introduction to Intersection-Union Tests

Jimmy Akira Doi
jadoi@unity.ncsu.edu
North Carolina State University
Department of Statistics

Slide 1

Introduction
Often, the quality of a product is determined by several parameters. The product is determined to be acceptable if each of the parameters meets certain standards.

Acceptance Sampling: Upholstery Fabric Example
ASTM Standards (American Society for Testing and Materials)
- Mean breaking strength (in pounds)
- Probability of passing flammability test
- Fabric deemed acceptable when each of these parameters satisfy certain criteria

Slide 2

Hypothesis Testing Framework
- Formulate the null hypothesis $H_0$ which states that one or more of the parameters do not meet their criteria.
- Formulate the alternative hypothesis $H_A$ which states that all of the parameters do meet their criteria.

Under this framework, note that rejection of the null hypothesis $H_0$ corresponds to the decision that the product is acceptable. Also, the probability of a Type I error will be the consumer’s risk.

Slide 3

Some Concerns
- Multiple Testing Procedure
  Is there a need for a multiplicity adjustment?
- Need to Control Consumer’s Risk

Slide 4
Outline

- Review of the Problem of Multiple Comparisons
- Definition of the Intersection-Union Test
- 2 Relevant Theorems
- Example with Simulation Results
- Bioequivalence Introduction

Multiple Comparisons

Example: Experiment with 5 treatments

- 10 possible pairwise comparisons or contrasts (0.05 level t-tests: $t_1, t_2, \ldots, t_{10}$)
- Combine all tests into $\Phi$, where $\Phi_i$ rejects if any of the $t_i$ rejects.
- Familywise error rate $\alpha^* \neq 0.05$

<table>
<thead>
<tr>
<th>$n$</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha^*$</td>
<td>14%</td>
<td>23%</td>
<td>40%</td>
<td>64%</td>
<td>90%</td>
</tr>
</tbody>
</table>

- Need multiplicity adjustment

“According to one survey, multiple comparison methods constitute the second most frequently applied group of statistical methods, second only to the F-test ... If they rank second to frequency of use, they rank perhaps first in frequency of abuse.”

Jason Hsu,
Multiple Comparisons: Theory and Methods

Intersection-Union Tests (IUT)

Given $X \sim f(x|\theta)$, suppose $H_0$ is expressed as a union of $k$ sets (the index set need not be finite):

$$H_0 : \theta \in \Theta_0 = \bigcup_{i=1}^{k} \Theta_i \text{ vs. } H_A : \theta \in \Theta_0^c = \bigcap_{i=1}^{k} \Theta_i^c \quad (1)$$

Suppose for each $i$, $R_i$ is the rejection region for a test of

$$H_{0i} : \theta \in \Theta_i \text{ vs. } H_{Ai} : \theta \in \Theta_i^c.$$

The rejection region for the IUT of $H_0$ vs. $H_A$ is $\bigcap_{i=1}^{k} R_i$.

In other words, the IUT rejects only if all of the tests reject.

Do we need a multiplicity adjustment?
**Theorem 1**

The following theorems are from: Berger, RL (1997), “Likelihood Ratio Tests and Intersection-Union Tests”

Theorem 1: If $R_i$ is a level-\(\alpha\) test of $H_0$ for $i = 1, \ldots, k$, then the IUT with rejection region $R = \bigcap_{i=1}^{k} R_i$ is a level-\(\alpha\) test of $H_0$ versus $H_A$ in (1).

(No adjustment necessary!)

Usefulness of IUT: Rely upon the simpler tests of the individual hypotheses and not worry about the joint multivariate dist'n.

---

**Slide 9**

---

**Proof of Theorem 1**

Theorem 1: If $R_i$ is a level-\(\alpha\) test of $H_0$, for $i = 1, \ldots, k$, then the IUT with rejection region $R = \bigcap_{i=1}^{k} R_i$ is a level-\(\alpha\) test of $H_0$ versus $H_A$ in (1).

Proof: Let $\theta \in \Theta_0 = \bigcup_{i=1}^{k} \Theta_i$. So, for some $i = 1, \ldots, k$ (say $i = i'$), $\theta \in \Theta_{i'}$. Thus,

$$P_{\theta}(\cap R_i) \leq P_{\theta}(R_{i'}) \leq \alpha$$

Since $\theta \in \Theta_0$ was arbitrarily chosen, the IUT is level-\(\alpha\) as

$$\sup_{\theta \in \Theta_0} P_{\theta}(\cap R_i) \leq \alpha$$

(No adjustment necessary)

---

**Theorem 2**

Theorem 2: For some $i = 1, \ldots, k$, suppose $R_i$ is a size-\(\alpha\) rejection region for testing $H_{0i}$ vs. $H_{Ai}$. For every $j = 1, \ldots, k, j \neq i$, suppose $R_j$ is a level-\(\alpha\) rejection region for testing $H_{0j}$ vs. $H_{Aj}$. Suppose there exists a sequence of parameter points $\theta_l$, $l = 1, 2, \ldots$, in $\Theta_i$ such that

$$\lim_{l \to \infty} P_{\theta_l}(R_i) = \alpha,$$

and, for every $j = i, \ldots, k, j \neq i$,

$$\lim_{l \to \infty} P_{\theta_l}(R_j) = 1.$$

Then the IUT with rejection region $R = \bigcap_{i=1}^{k} R_i$ is a size-\(\alpha\) test of $H_0$ vs. $H_A$.

---

**Slide 10**

---

**Proof of Theorem 2**

Proof: \(\lim_{l \to \infty} P_{\theta_l} \left( \bigcap_{m=1}^{k} R_m \right) \geq \lim_{l \to \infty} \sum_{m=1}^{k} P_{\theta_l}(R_m) - (k - 1)\)

\[= \alpha + (k - 1) - (k - 1) = \alpha.\]

---

**Slide 11**

---

**Slide 12**
Acceptance Sampling: Upholstery Fabric Example

- \( \theta_1 \), mean breaking strength (in pounds)
- \( \theta_2 \), probability of passing flammability test
- Fabric is acceptable when meeting the standards:
  \[ \theta_1 > 50 \text{ and } \theta_2 > 0.95 \]
- \( H_0: \{ \theta_1 \leq 50 \text{ or } \theta_2 \leq 0.95 \} \)
- \( H_A: \{ \theta_1 > 50 \text{ and } \theta_2 > 0.95 \} \)

\[ X_1, \ldots, X_n \text{ iid } N(\theta_1, \sigma^2), \text{ use LRT of } H_{01}: \theta_1 \leq 50. \]
\[ Y_1, \ldots, Y_m \text{ iid } \text{Bern}(\theta_2), \text{ use LRT of } H_{02}: \theta_2 \leq 0.95. \]

IUT rejection region is \( \{ (x, y) : \frac{\sum_{i=1}^n x_i}{\sqrt{n}} > t \text{ and } \sum_{i=1}^m y_i > b \} \).

MC Simulation: \( n = m = 58, t = 1.672, b = 57 \)
\( \Rightarrow \) LRTs are approx. size-\( \alpha \) (0.05) tests.

Table 1. Monte Carlo Estimates of \( \alpha \)

| Fixed \( \theta_2 \) and Increasing \( \theta_1 \) |
|------------------|-------------------|------------------|------------------|------------------|
| \( \theta_2 = .95 \) |
| \( \theta_1 = 50 \) |
| \( \theta_1 = 60 \) |
| \( \theta_1 = 75 \) |
| \( \theta_1 = 100 \) |
| \( \theta_1 = 50 \) |
| \( \theta_1 = 60 \) |
| \( \theta_1 = 80 \) |
| \( \theta_1 = 90 \) |

Table 2. Monte Carlo Estimates of \( \alpha \)

| Fixed \( \theta_1 \) and Increasing \( \theta_2 \) |
|------------------|-------------------|------------------|------------------|------------------|
| \( \theta_1 = .95 \) |
| \( \theta_2 = .95 \) |
| \( \theta_2 = .99 \) |
| \( \theta_2 = .999 \) |
| \( \theta_2 = .9999 \) |
| \( \theta_1 = 50 \) |
| \( \theta_1 = 60 \) |
| \( \theta_1 = 70 \) |
| \( \theta_1 = 80 \) |

Table 3. MC Estimates of Power

<table>
<thead>
<tr>
<th>( \theta_2 )</th>
<th>( \theta_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_2 = .95 )</td>
<td>( \theta_1 = 50 )</td>
</tr>
<tr>
<td>( \theta_1 = 50 )</td>
<td>( \theta_1 = 60 )</td>
</tr>
<tr>
<td>( \theta_1 = 70 )</td>
<td>( \theta_1 = 80 )</td>
</tr>
</tbody>
</table>

Uses of the IUT

- Acceptance Sampling
- Comparisons of Regression Lines
- Testing for Contingency Tables
- Bioequivalence
New Drug Development and Approval

Typically, the approval of a new drug product requires a process that is very time consuming and expensive.

- Phase I, II, III Clinical Trials
- Efficacy, Safety, Risk (side effects)

In the US, if a biopharmaceutical company successfully introduces a new drug into the market, the process typically requires 15 years and approximately $500 million.

Companies that produce generic drugs can bypass this process by showing the drug regulatory agency that their product is similar (therapeutically) to that of an approved reference drug. That is, they want to establish bioequivalence.

A Definition

(Bio)equivalence

“Two different drugs or formulations of the same drug are called bioequivalent if they are absorbed into the blood and become available at the drug action site at about the same rate and concentration.”

Berger and Hsu (1996)

In order for the generic drug to be approved in the US, the Food and Drug Administration (FDA, www.fda.gov) requires a test of the following:

Hypothesis Testing Framework for Bioequivalence (FDA)

Let
\[ \mu_R = \text{Mean blood concentration of a reference drug} \]
\[ \mu_T = \text{Mean blood concentration of a generic drug} \]

\[ H_0 : \mu_T - \mu_R \geq \delta \quad \text{or} \quad \mu_T - \mu_R \leq -\delta \]
\[ H_A : -\delta < \mu_T - \mu_R < \delta \]

where \( \delta > 0 \) is a tolerance limit specified by the drug regulatory agency.

\[ IUT: \text{Combine two, one-sided, size-} \alpha \text{ tests to obtain an overall size-} \alpha \text{ test. This test is often referred to as the } TOST. \]

TOST

- Most widely used IUT
- Much wider recognition of IUTs due to this application
- But, TOST suffers from a lack of power
Summary

- *IUT* avoids the need for multiplicity adjustments.
- 2 relevant theorems:
  
  **Theorem 1:** IUT is level-α when all others are level-α.
  
  **Theorem 2:** Under certain conditions, IUT is size α.
- No need to postulate a multivariate model – just combine the simpler individual tests.
- Strongest application found through bioequivalence studies.