HYPOTHESIS TESTING OF BIO-EQUIVALENCE

BY MIN MA

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

Arthur Cohen and Harold B. Sackrowitz

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2013

ABSTRACT OF THE DISSERTATION

Hypothesis Testing of Bio-equivalence

by MIN MA

Dissertation Director: Arthur Cohen and Harold B. Sackrowitz

This thesis considers two statistics problems in bio-equivalence. The first problem concerns 2-stage testing for bio-equivalence of parameters emanating from two populations. The first stage of the procedure can be thought of as a pilot sample which is used to determine the feasibility of taking an additional sample that would lead to inferring bio-equivalence. If a second sample is taken, the combined sample can then be used for inference purposes regarding bio-equivalence. Many models, including normal, Poisson, binomial, matched pairs, testing means and variances or both simultaneously are considered.

The second part of the thesis is concerned with multiple testing of bio-equivalence. Here an actual data set concerned with analyzing different types of iron content with different instruments is studied for various bio-equivalent outcomes among pairs. To perform the statistical analysis, two standard statistical methods are used along with a new method. Both equivalence testing and simultaneous interval estimates are offered.

Acknowledgements

Through my doctorial pursuit, I have been so blessed to have the support from my advisor, Professor Arthur Cohen, who gave me the great opportunity to discover the joys and frustrations behind learning and research. He has always been the nicest and most patient guide to lead me through difficulties. This thesis would not have been possible without his illuminating instruction and enthusiastic encouragement.

My sincerest gratitude also goes to my co-advisor, Professor Harold B. Sackrowitz, who has always there to listen and inspired me many times throughout my dissertation research with his broad knowledge and wonderful intuitions. It is impossible to over express how thankful I am to have them as my advisers.

I am deeply grateful to Professor Kesar Singh for his encouragement and practical advice. He spent hours with me exploring my options to realize massive computations and simulations in the most efficient manner. I am indebted to him for his expertise and inspiring discussions. May he rest in peace.

I would like to thank my thesis committee, Professor Douglas Jones and Professor Cuihui Zhang for their direction, dedication, and invaluable advice. I would also like to express my deep appreciation to our Graduate Director, Professor John Kolassa for sharing his knowledge, support and help with me.

Finally, and most importantly, I would like to thank my husband Mingyuan for his persistent support, encouragement, patience and unwavering love. I thank my parents, Suyin and Shulan, for raising me, loving me, teaching me, and always believing me. To them I dedicate this thesis.

Dedication

To my family.

Table of Contents

List of Tables

List of Figures

Chapter 1
Two-Stage Bio-equivalence Test

1.1 Introduction

A typical two population problem in bio-equivalence considers a hypothesis testing problem about a parameter θ, where θ represents a difference or ratio of parameters, one from each population. The hypothesis testing problem considered has a null hypothesis $H : \theta \leq \epsilon_1$ or $\theta \geq \epsilon_2$ vs an alternative hypothesis $K : \epsilon_1 < \theta < \epsilon_2$. A typical interpretation is that a new treatment yields outcomes that are close to the outcomes of a standard treatment and this closeness is reflected in the alternative hypothesis.

Experimental, economical and ethical constraints often are such that it makes sense to take pilot samples from each population before embarking on a full scale sample from each population. If the pilot sample looks promising in favor of bio-equivalence, then a second sample from each population can be taken to confirm bio-equivalence or to decide against it. Data from the pilot sample and the second sample would be used to carry out the inference procedure, i.e. the hypothesis test.

We will consider a variety of two population models where the distributions for each population are a one or two parameter exponential family. Samples of size n_1 will be taken from each population. Based on this pilot sample, a two sample test statistic will be used to decide whether to continue the study or stop. If the study continues, samples of size n_2 will be taken from each population and the bio-equivalence hypothesis will be tested. The test at the end of stage 2 will be uniformly most powerful (UMP) or uniformly most powerful unbiased (UMPU) conditioned on the set determined at the end of stage 1, which indicated that a stage 2 sample should be taken.

In addition to deriving the procedure for various two population models, we indicate procedures for single population models where the bio-equivalence hypothesis is still expressed in terms of a parameter θ as before. In addition guidelines for sample sizes at each stage will be offered.

The following models will be studied in subsequent sections: Section 1.2.1: Normal distribution with known variance; Section 1.2.2: Matched pairs of normal distributions with unknown variances; Section 1.2.3: Two independent normal distributions with common unknown variance; Section 1.3.1: Binomial distribution; Section 1.3.2: Two independent binomial distributions; Section 1.3.3: Poisson distribution; Section 1.4: Two independent normal distributions with unknown means and variances. Bio-equivalence is defined in terms of both differences in means and ratios of variances.

Two key references for this work are as follows: (1) Wellek (2002). Most distributional models are studied although some formulations of bio-equivalence are different and the sample designs are different. (2) Cohen and Sackrowitz (1996) studied two-stage inference procedures where the first stage yields a pilot sample.

1.2 Bio-equivalence test under Normal distribution

1.2.1 Normal distribution with known variance

Consider the following statistical model: Let X_i, $i = 1, \cdots, n_1$ be a random sample from $\mathcal{N}(\theta, \sigma^2)$, with σ^2 known and without loss of generality, $\sigma^2 = 1$. The bio-equivalence problem is to test

$$H : |\theta| \geq \epsilon \text{ vs } K : |\theta| < \epsilon$$

Let $\bar{X}_1 = \sum_{i=1}^{n_1} X_i/n_1$. Consider the event

$$R = \{(x_1, \cdots, x_{n_1}) : \sqrt{n_1}|\bar{X}_1| < C_{\alpha_1;\sqrt{n_1}\epsilon}\} \tag{1.2.1}$$

where $C_{\alpha_1;\sqrt{n_1}\epsilon} = \sqrt{\chi^2_{1,\alpha_1}(n_1\epsilon^2)}$ and $\chi^2_{1,\alpha_1}(n_1\epsilon^2)$ denotes the $\alpha_1 th$ quantile of a χ^2-distribution with $df = 1$ and non-centrality parameter $n_1\epsilon^2$.

When R occurs, take a second sample of size $n_2 = n - n_1$. Let

$$\mathbf{X} = (X_1, \ldots, X_{n_1}, X_{n_1+1}, \ldots, X_n)$$

then the joint conditional density of $\mathbf{X}$ given R is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|R;\theta) &= \prod_{i=1}^{n} f_{X_i}(x_i;\theta)/P_\theta(R) \qquad\qquad \mathbf{x} \in R \\ &= \frac{(1/2\pi)^{n/2}\exp(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2)\exp(-\frac{n}{2}(\bar{x}-\theta)^2)}{P_\theta(R)} \end{aligned} \tag{1.2.2}$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$. From (1.2.2), we see that given a second sample is taken, $\bar{X}$ is a sufficient statistic for θ. And the conditional density of $\bar{X}$ can be written as

$$\begin{aligned} f_{\bar{X}}(\bar{x}|R;\theta) &= f(\bar{x};\theta)P(R|\bar{x})/P_\theta(R) \qquad \bar{x} \in R \\ &= K(\theta)h(\bar{x})\exp(n\bar{x}\theta) \end{aligned} \tag{1.2.3}$$

where

$$K(\theta) = \sqrt{\frac{n}{2\pi}}\exp(-\frac{n}{2}\theta^2)/P_\theta(R)$$

$$h(\bar{x}) = \exp(-\frac{n}{2}\bar{x}^2)P(R|\bar{x})$$

$$P_\theta(R) = \Phi(C_{\alpha_1;\sqrt{n_1}\epsilon} - \sqrt{n_1}\theta) + \Phi(C_{\alpha_1;\sqrt{n_1}\epsilon} + \sqrt{n_1}\theta) - 1 \tag{1.2.4}$$

$$P(R|\bar{x}) = \Phi\left(\frac{C_{\alpha_1;\sqrt{n_1}\epsilon} - \sqrt{n_1}\bar{x}}{\sqrt{n_2/n}}\right) + \Phi\left(\frac{C_{\alpha_1;\sqrt{n_1}\epsilon} + \sqrt{n_1}\bar{x}}{\sqrt{n_2/n}}\right) - 1 \tag{1.2.5}$$

See Appendix 3.1 for detailed derivation of (1.2.4) and (1.2.5). Notice that

$$f_{\bar{X}}(\bar{x}|R;\epsilon) = f_{\bar{X}}(-\bar{x}|R;-\epsilon) \qquad \bar{x} \in R$$

which implies that the rejection region for the second stage test will be a symmetric interval with respect to 0. Furthermore, from (1.2.3), the conditional distribution of $\bar{X}$ belongs to a single parameter exponential family. By virtue of A1.5 Theorem and A1.7 Corollary of Wellek (2002), a UMP test is given as

$$\phi_{\bar{X}}(\bar{x}) = \begin{cases} 1 & |n\bar{x}| < C \\ 0 & \text{otherwise} \end{cases} \tag{1.2.6}$$

where C is determined by the following equation

$$\int_{|n\bar{x}|<C} f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}|R;\epsilon)d\bar{x} = \alpha_2 \tag{1.2.7}$$

And the conditional power of this test at $\theta = \theta_0$, $|\theta_0| < \epsilon$ is

$$\beta(\theta_0) = \int_{|n\bar{x}|<C} f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}|R;\theta_{\mathbf{0}})d\bar{x} \tag{1.2.8}$$

1.2.2 Matched pairs of normal distributions with unknown variances

Given n_1 mutually independent pairs $(X_1, Y_1), \ldots, (X_{n_1}, Y_{n_1})$ of random variables, consider the case that the distribution of the individual differences

$$D_i = X_i - Y_i$$

are i.i.d. normal distribution with parameter $\mu = E(D_i)$ and $\sigma_D^2 = Var(D_i)$. Then the equivalence problem referring to these differences reads

$$H : |\mu| \geq \epsilon \text{ vs } K : |\mu| < \epsilon$$

Let

$$\bar{D}_1 = \sum_{i=1}^{n_1} D_i/n_1 \quad S_1^2 = \sum_{i=1}^{n_1}(D_i - \bar{D}_1)^2/n_1$$

Then a classic α_1-level two one-sided tests (TOST) procedure proposed in Schuirmann (1987), can be performed, which has a rejection region as follows

$$R = \{(d_1, \cdots, d_{n_1}) : |\bar{D}_1| < \epsilon - t_{\alpha_1, n_1-1}\frac{S_1}{\sqrt{n_1 - 1}}\} \tag{1.2.9}$$

where t_{α_1,n_1-1} is the upper $\alpha_1 th$ quantile of a t distribution with $n_1 - 1$ degrees of freedom.

Or alternatively let $T_1 = \sum_{i=1}^{n_1} D_i^2$, then $S_1^2 = \frac{T_1}{n_1} - \bar{D}_1^2$, then the above rejection region (1.2.9) can be rewritten as a function of $(\bar{D}_1, T_1)$ for later convenience.

$$R = \left\{ (d_1, \cdots, d_{n_1}) : |\bar{D}_1| < \epsilon - t_{\alpha_1, n_1 - 1} \sqrt{\frac{T_1/n_1 - \bar{D}_1^2}{n_1 - 1}} \right\} \tag{1.2.10}$$

If the observed values of $\bar{D}_1$ and T_1 fall out of this rejection region, we stop and declare non-equivalence. Otherwise we proceed to take a second sample of size $n_2 = n - n_1$ consisting of pairs (X_i, Y_i), $i = n_1 + 1, \cdots, n$. Let

$$\begin{aligned} \bar{D}_2 = & \sum_{i=n_1+1}^{n} D_i/n_2, \quad T_2 = \sum_{i=n_1+1}^{n} D_i^2 \\ \bar{D} = & \sum_{i=1}^{n} D_i/n, \qquad T = \sum_{i=1}^{n} D_i^2 \end{aligned}$$

Then $(\bar{D}_1, T_1)$ and $(\bar{D}_2, T_2)$ are mutually independent. Also

$$\bar{D}_2 = \frac{n\bar{D} - n_1\bar{D}_1}{n_2}$$

and

$$T_2 = T - T_1$$

It is easy to see that the joint density of $(\bar{D}_1, T_1, \bar{D}, T)$ is

$$f(\bar{D}_1, T_1, \bar{D}, T) = K(\mu, \sigma_D^2) h(\bar{D}_1, T_1, \bar{D}, T) \exp\left(\frac{n\mu\bar{D}}{\sigma_D^2} - \frac{T}{2\sigma_D^2} \right) \quad (\bar{D}_1, T_1) \in R$$

where

$$\begin{aligned} h(\bar{D}_1, T_1, \bar{D}, T) = & (T_1 - n_1\bar{D}_1^2)^{\frac{n_1-3}{2}} \times \left(T - T_1 - \frac{(n\bar{D} - n_1\bar{D}_1)^2}{n_2} \right)^{\frac{n_2-3}{2}} \times \\ & I_{(0,T_1)}(n_1\bar{D}_1^2) \times I_{(0,T-T_1)} \left(\frac{(n\bar{D} - n_1\bar{D}_1)^2}{n_2} \right) \end{aligned}$$

.

Thus the conditional density of $(\bar{D}, T)$ given rejection occurs is

$$\begin{aligned} f(\bar{D}, T|R) = & \frac{f(\bar{D}, T, R)}{Pr(R)} \\ = & \frac{\int_R f(\bar{D}_1, T_1, \bar{D}, T) d\bar{D}_1 dT_1}{Pr(R)} \\ = & K'(\mu, \sigma_D^2) h'(\bar{D}, T) \exp\left(\frac{n\mu\bar{D}}{\sigma_D^2} - \frac{T}{2\sigma_D^2} \right) \end{aligned}$$

where $h^{'}(\bar{D},T)=\int_R h(\bar{D}_1,T_1,\bar{D},T)d\bar{D}_1dT_1$.

From the above joint density, it is easy to tell that $\bar{D}$ and T are sufficient statistics for μ and σ_D^2. Therefore here adopting the same idea of the classic two one-sided t tests, we propose to perform two one-sided tests analogously to decide the rejection region based on all n observations given rejection based on the first stage sample. The two one-sided hypotheses are as follows,

$$H_1:\mu\geq\epsilon \text{ vs } K_1:\mu<\epsilon \tag{1.2.11}$$

$$H_2:\mu\leq-\epsilon \text{ vs } K_2:\mu>-\epsilon \tag{1.2.12}$$

For (1.2.11), let

$$Z_1=\bar{D}\quad Z_2=T-2n\epsilon\bar{D}\quad \nu_1=\frac{n_1(\mu-\epsilon)}{\sigma_D^2}\quad \nu_2=-\frac{1}{2\sigma_D^2}$$

Then the conditional density of (Z_1,Z_2) given rejection occurs is

$$K^{''}(\nu_1,\nu_2)h^{'}(z_1,z_2+2n\epsilon z_1)\exp(\nu_1z_1+\nu_2z_2)$$

From Lehmann (1986), a UMPU test of size α_2 is obtained by conditioning on $Z_2=z_2$ and rejecting when $Z_1<U(z_2)$ where $U(z_2)$ is determined from

$$\frac{\int_{-\infty}^{U(z_2)}h^{'}(u,z_2+2n\epsilon u)du}{\int_{-\infty}^{\infty}h^{'}(u,z_2+2n\epsilon u)du}=\alpha_2 \tag{1.2.13}$$

Similarly, for hypothesis (1.2.12), let

$$Z_1=\bar{D}\quad Z_2=T+2n\epsilon\bar{D}\quad \nu_1=\frac{n_1(\mu+\epsilon)}{\sigma_D^2}\quad \nu_2=-\frac{1}{2\sigma_D^2}$$

Then the conditional density of (Z_1,Z_2) given rejection occurs is

$$K^{''}(\nu_1,\nu_2)h^{'}(z_1,z_2-2n\epsilon z_1)\exp(\nu_1z_1+\nu_2z_2)$$

and a UMPU test of size α_2 is obtained by conditioning on $Z_2=z_2$ and rejecting when $Z_1>L(z_2)$ where $L(z_2)$ is determined from

$$\frac{\int_{L(z_2)}^{\infty}h^{'}(u,z_2-2n\epsilon u)du}{\int_{-\infty}^{\infty}h^{'}(u,z_2-2n\epsilon u)du}=\alpha_2 \tag{1.2.14}$$

Therefore, after rejecting non-equivalence on the first stage based on n_1 observations of differences between (X_i, Y_i), additional n_2 differences are observed. Then the value of $z_2 = t - 2n\epsilon\bar{d}$ and $z_2 = t + 2n\epsilon\bar{d}$ are calculated respectively. From (1.2.13) and (1.2.14) a rejection interval $(L(t + 2n\epsilon\bar{d}), U(t - 2n\epsilon\bar{d}))$ is obtained. If the observed value of Z_1 or $\bar{D}$ falls into this interval, bio-equivalence between X and Y is established.

1.2.3 Two independent normal distributions with common unknown variance

Now consider the following parallel design model

$$X_i \sim \mathcal{N}(\mu_1, \sigma^2) \quad i = 1, \cdots, n_{11}$$

$$Y_j \sim \mathcal{N}(\mu_2, \sigma^2) \quad j = 1, \cdots, n_{12}$$

where X_i are Y_j are independent. A bio-equivalence hypothesis is formulated as

$$H : |\mu_1 - \mu_2| \geq \epsilon \quad \text{vs} \quad K : |\mu_1 - \mu_2| < \epsilon$$

Let

$$\bar{X}_1 = \sum_{i=1}^{n_{11}} X_i/n_{11}, \qquad \bar{Y}_1 = \sum_{j=1}^{n_{12}} Y_j/n_{12}$$

$$S_1^2 = \frac{\sum_{i=1}^{n_{11}}(X_i - \bar{X}_1)^2 + \sum_{j=1}^{n_{12}}(Y_j - \bar{Y}_1)^2}{n_{11} + n_{12} - 2}$$

.

Similar to Section 1.2.2, the classic two one-sided tests (TOST) is performed with rejection region

$$R = \left\{(x_i, y_j) : |\bar{X}_1 - \bar{Y}_1| < \epsilon - t_{\alpha_1, n_{11}+n_{12}-2} S_1 \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}\right\} \qquad (1.2.15)$$

Again let

$$T_1 = \sum_{i=1}^{n_{11}} X_i^2 + \sum_{j=1}^{n_{12}} Y_j^2$$

then $S_1^2 = \frac{T_1 - n_{11}\bar{X_1^2} - n_{12}\bar{Y_1^2}}{n_{11}+n_{12}-2}$. Rewrite the rejection region (1.2.15) as a function of $(\bar{X}_1, \bar{Y}_1, T_1)$ for later convenience. That is (1.2.15) gives rise to

$$R = \left\{ |\bar{X}_1 - \bar{Y}_1| < \epsilon - t_{\alpha_1, n_{11}+n_{12}-2} \sqrt{\frac{T_1 - n_{11}\bar{X_1^2} - n_{12}\bar{Y_1^2}}{n_{11}+n_{12}-2} \left(\frac{1}{n_{11}} + \frac{1}{n_{12}}\right)} \right\} \tag{1.2.16}$$

Based on $n_{11} + n_{12}$ observations, if the test result is not significant, we stop and declare non-equivalence. When rejection occurs, we proceed to take an additional sample which consists of X_i of size $n_{21} = n_1 - n_{11}$ and Y_j of size $n_{22} = n_2 - n_{12}$. Let

$$\bar{X}_2 = \sum_{i=n_{11}+1}^{n_1} X_i/n_{21}, \quad \bar{Y}_2 = \sum_{j=n_{12}+1}^{n_2} Y_j/n_{22}, \quad T_2 = \sum_{i=n_{11}+1}^{n_1} X_i^2 + \sum_{j=n_{12}+1}^{n_2} Y_j^2,$$

$$\bar{X} = \sum_{i=1}^{n_1} X_i/n_1, \quad \bar{Y} = \sum_{j=1}^{n_2} Y_j/n_2, \quad T = \sum_{i=1}^{n_1} X_i^2 + \sum_{j=1}^{n_2} Y_j^2.$$

Then $(\bar{X}_1, \bar{Y}_1, T_1)$ and $(\bar{X}_2, \bar{Y}_2, T_2)$ are mutually independent. Also

$$\bar{X}_2 = \frac{n_1\bar{X} - n_{11}\bar{X}_1}{n_{21}}, \qquad \bar{Y}_2 = \frac{n_2\bar{Y} - n_{12}\bar{Y}_1}{n_{22}}, \qquad T_2 = T - T_1.$$

It is easy to see that the joint density of $(\bar{X}_1, \bar{Y}_1, T_1, \bar{X}, \bar{Y}, T)$ is

$$\begin{aligned} f(\bar{X}_1, \bar{Y}_1, T_1, \bar{X}, \bar{Y}, T) &= K(\mu_1, \mu_2, \sigma^2) \times h(\bar{X}_1, \bar{Y}_1, T_1, \bar{X}, \bar{Y}, T) \\ &\quad \times \exp\left(\frac{n_1\mu_1\bar{X} + n_2\mu_2\bar{Y}}{\sigma^2} - \frac{T}{2\sigma^2}\right) \end{aligned}$$

where

$$\begin{aligned} h(\bar{X}_1, \bar{Y}_1, T_1, \bar{X}, \bar{Y}, T) &= (T_1 - n_{11}\bar{X}_1^2 - n_{12}\bar{Y}_1^2)^{\frac{n_{11}+n_{12}-4}{2}} \times I_{(0,T_1)}(n_{11}\bar{X}_1^2 + n_{12}\bar{Y}_1^2) \times \\ &\quad \left(T - T_1 - \frac{(n_1\bar{X} - n_{11}\bar{X}_1)^2}{n_{21}} - \frac{(n_2\bar{Y} - n_{12}\bar{Y}_1)^2}{n_{22}}\right)^{\frac{n_{21}+n_{22}-4}{2}} \times \\ &\quad I_{(0,T-T_1)}\left(\frac{(n_1\bar{X} - n_{11}\bar{X}_1)^2}{n_{21}} + \frac{(n_2\bar{Y} - n_{12}\bar{Y}_1)^2}{n_{22}}\right) \end{aligned}$$

Thus the conditional joint density of $(\bar{X}, \bar{Y}, T)$ given that the rejection occurs in the first stage is

$$\begin{aligned} f(\bar{X}, \bar{Y}, T|R) &= \frac{f(\bar{X}, \bar{Y}, T, R)}{Pr(R)} \\ &= \frac{\int_R f(\bar{X}_1, \bar{Y}_1, T_1, \bar{X}, \bar{Y}, T) d\bar{X}_1 d\bar{Y}_1 dT_1}{Pr(R)} \\ &= K^{'}(\mu_1, \mu_2, \sigma^2) h^{'}(\bar{X}, \bar{Y}, T) \exp\left(\frac{n_1\mu_1\bar{X} + n_2\mu_2\bar{Y}}{\sigma^2} - \frac{T}{2\sigma^2}\right) \end{aligned}$$

where $h^{'}(\bar{X}, \bar{Y}, T) = \int_R h(\bar{X}_1, \bar{Y}_1, T_1, \bar{X}, \bar{Y}, T) d\bar{X}_1 d\bar{Y}_1 dT_1$.

Here we adopt the same technique used in the last section. Two one-sided tests for the following hypotheses are conducted to find the lower and upper bound which form a rejection interval.

$$H_1 : \mu_1 - \mu_2 \geq \epsilon \quad \text{vs} \quad K_1 : \mu_1 - \mu_2 < \epsilon \tag{1.2.17}$$

$$H_2 : \mu_1 - \mu_2 \leq -\epsilon \quad \text{vs} \quad K_2 : \mu_1 - \mu_2 > -\epsilon \tag{1.2.18}$$

First consider hypothesis (1.2.17). Let

$$Z_1 = n_1\bar{X}, \quad Z_2 = n_1\bar{X} + n_2\bar{Y}, \quad Z_3 = T - 2n_1\epsilon\bar{X}.$$

$$\nu_1 = \frac{\mu_1 - \mu_2 - \epsilon}{\sigma^2}, \quad \nu_2 = \frac{\mu_2}{\sigma^2}, \quad \nu_3 = -\frac{1}{2\sigma^2}.$$

Then the conditional joint density of (Z_1, Z_2, Z_3) given the rejection occurs in the first stage is

$$K^{''}(\nu_1, \nu_2, \nu_3) h^{'}\left(\frac{z_1}{n_1}, \frac{z_2 - z_1}{n_2}, z_3 + 2\epsilon z_1\right) \exp(\nu_1 z_1 + \nu_2 z_2 + \nu_3 z_3)$$

A UMPU test of size α_2 is obtained by conditioning on $Z_2 = z_2$, $Z_3 = z_3$ and rejecting when $Z_1 < U(z_2, z_3)$, where $U(z_2, z_3)$ is determined from

$$\frac{\int_{-\infty}^{U(z_2,z_3)} h^{'}\left(\frac{u}{n_1}, \frac{z_2-u}{n_2}, z_3 + 2\epsilon u\right) du}{\int_{-\infty}^{\infty} h^{'}\left(\frac{u}{n_1}, \frac{z_2-u}{n_2}, z_3 + 2\epsilon u\right) du} = \alpha_2 \tag{1.2.19}$$

Similarly, for hypothesis (1.2.18), let

$$Z_1 = n_1\bar{X}, \quad Z_2 = n_1\bar{X} + n_2\bar{Y}, \quad Z_3 = T + 2n_1\epsilon\bar{X}.$$

$$\nu_1 = \frac{\mu_1 - \mu_2 + \epsilon}{\sigma^2}, \quad \nu_2 = \frac{\mu_2}{\sigma^2}, \quad \nu_3 = -\frac{1}{2\sigma^2}.$$

Then the conditional joint density of (Z_1, Z_2, Z_3) given rejection occurs is

$$K''(\nu_1, \nu_2, \nu_3)h'\left(\frac{z_1}{n_1}, \frac{z_2 - z_1}{n_2}, z_3 - 2\epsilon z_1\right)\exp(\nu_1 z_1 + \nu_2 z_2 + \nu_3 z_3)$$

and a UMPU test of size α_2 is obtained by conditioning on $Z_2 = z_2$, $Z_3 = z_3$ and rejecting when $Z_1 > L(z_2, z_3)$, where $L(z_2, z_3)$ is determined from

$$\frac{\int_{L(z_2,z_3)}^{\infty} h'\left(\frac{u}{n_1}, \frac{z_2-u}{n_2}, z_3 - 2\epsilon u\right)du}{\int_{-\infty}^{\infty} h'\left(\frac{u}{n_1}, \frac{z_2-u}{n_2}, z_3 - 2\epsilon u\right)du} = \alpha_2 \tag{1.2.20}$$

Therefore, after rejecting non-equivalence in the first stage based on $n_{11} + n_{12}$ observations of $(X_i, Y_j, i = 1, \ldots, n_{11}, j = 1, \ldots, n_{12})$, an additional $n_{21} + n_{22}$ sample is collected. Then the values are $z_2 = n_1\bar{x} + n_2\bar{y}$ and $z_3 = t - 2n_1\epsilon\bar{x}$ or $z_2 = n_1\bar{x} + n_2\bar{y}$ and $z_3 = t + 2n_1\epsilon\bar{x}$. From (1.2.19) and (1.2.20), a rejection interval $(L(n_1\bar{x} + n_2\bar{y}, t + 2n_1\epsilon\bar{x}), U(n_1\bar{x} + n_2\bar{y}, t - 2n_1\epsilon\bar{x}))$ is obtained. If the observed value of Z_1 or $n_1\bar{X}$ falls into this interval, bio-equivalence between X and Y is concluded.

1.2.4 Sample size decision

A. Normal distribution with known variance

To decide the pilot sample size given certain significance level α_1 and power β_1 at a certain alternative $\theta = \theta_0$, $|\theta_0| < \epsilon$, consider the rejection region expressed as (1.2.1). The required pilot sample size n_1 can be solved from the following equation

$$\beta_1 = \Phi(C_{\alpha_1;\sqrt{n_1}\epsilon} - \sqrt{n_1}\theta_0) + \Phi(C_{\alpha_1;\sqrt{n_1}\epsilon} + \sqrt{n_1}\theta_0) - 1 \tag{1.2.21}$$

where $C_{\alpha_1;\sqrt{n_1}\epsilon} = \sqrt{\chi^2_{1,\alpha_1}(n_1\epsilon^2)}$.

To decide the additional sample size n_2 once R occurs, given certain significance level α_2 and power β_2 at certain alternative $\theta = \theta_0$, we need to find a pair of (C, n) that satisfies the following equations,

$$\int_{|n\bar{x}|<C} f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}|R;\epsilon)d\bar{x} = \alpha_2 \tag{1.2.22}$$

$$\int_{|n\bar{x}|<C} f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}|R;\theta_{\mathbf{0}})d\bar{x} = \beta_2 \tag{1.2.23}$$

To solve these two equations computationally, we can use iteration entailing the following steps:

- *Step* 1. Choose some initial estimate of n. A good starting point could be the value solved from (1.2.21) when using (α_2, β_2) instead of (α_1, β_1).
- *Step* 2. Solve C from (1.2.23)
- *Step* 3. Update n by solving (1.2.22)
- *Step* 4. Repeat *Step* 2 and *Step* 3 until the change of n's value is small enough. (we set as no larger than 10^{-4}.)

Then $n_2 = n - n_1$ is the required additional sample size.

In Table 1.1, we provide some computational results using the iteration procedure provided above. Notice that in Table 1.1 the last two columns, where $total\ expected\ size$ is defined as

$$n_1 + Pr(\text{rejection in stage one}) \times n_2$$

and the last column lists the sample size required if using a single stage test to achieve the same power as a two-stage test while controlling the Type I error. To compare these two sample sizes, we can tell that as long as the stage one test serves as a quick preliminary test to decide if it is worthy of further validation and does not require a high

Table 1.1: Pilot and additional sample size when $\epsilon = \log(10/8)$ and $\theta = 0$

α_1	β_1	α_2	β_2	pilot size	additional size	total expected size	one stage test sample size
0.2	0.5	0.05	0.7	43	139	113	145
			0.8	43	167	127	172
			0.9	43	212	149	218
	0.6		0.7	56	133	136	145
			0.8	56	161	153	172
			0.9	56	207	181	218
	0.5	0.1	0.7	43	102	94	108
			0.8	43	126	106	132
			0.9	43	167	127	172
	0.6		0.7	56	97	115	108
			0.8	56	121	129	132
			0.9	56	161	153	172
	0.5	0.15	0.7	43	81	84	86
			0.8	43	102	94	108
			0.9	43	139	113	145
	0.6		0.7	56	75	101	86
			0.8	56	97	115	108
			0.9	56	134	137	145
	0.5	0.2	0.7	43	65	76	71
			0.8	43	85	86	91
			0.9	43	119	103	125
	0.6		0.7	56	60	92	71
			0.8	56	80	104	91
			0.9	56	114	125	125
0.3	0.5	0.05	0.7	24	141	95	145
			0.8	24	169	109	172
			0.9	24	214	131	218
	0.6		0.7	36	137	119	145
			0.8	36	165	135	172
			0.9	36	210	162	218
	0.5	0.1	0.7	24	105	77	108
			0.8	24	129	89	132
			0.9	24	169	109	172
	0.6		0.7	36	101	97	108
			0.8	36	125	111	132
			0.9	36	165	135	172
	0.5	0.15	0.7	24	83	66	86
			0.8	24	105	77	108
			0.9	24	142	95	145
	0.6		0.7	36	79	84	86
			0.8	36	101	97	108
			0.9	36	137	119	145
	0.5	0.2	0.7	24	67	58	71
			0.8	24	88	68	91
			0.9	24	121	85	125
	0.6		0.7	36	63	74	71
			0.8	36	83	86	91
			0.9	36	117	107	125

power or low error, the two-stage test is more efficient in a way that it is as powerful as the one-stage test while using a smaller sample to establish the bio-equivalence.

B. Paired observations with unknown variance

In the first stage, since $\bar{D_1} \sim \mathcal{N}(\mu, \frac{\sigma_D^2}{n_1})$, based on rejection region (1.2.9), the power function can be expressed as

$$\begin{aligned}\beta(\mu, \sigma_D^2, \alpha_1) &= \int_0^\infty \left(\Phi\left(\frac{\epsilon + \mu - t_{\alpha_1, n_1 - 1}\sqrt{\frac{S_1^2}{n_1 - 1}}}{\sqrt{\sigma_D^2/n_1}} \right) \right. \\ &\qquad \left. + \Phi\left(\frac{\epsilon - \mu - t_{\alpha_1, n_1 - 1}\sqrt{\frac{S_1^2}{n_1 - 1}}}{\sqrt{\sigma_D^2/n_1}} \right) - 1 \right) f_{S_1^2}(s_1^2) ds_1^2\end{aligned}$$

For any given value of the nuisance parameter σ_D^2, the above power function is strictly decreasing in $|\mu|$. Therefore, the maximum of the power function occurs at $\mu = 0$. Also the power function is increasing in n_1, so for any given (α_1, β_1), a minimal n_1 can be found such that at $\mu = 0$ the power function above is no less than β_1.

In the second stage, since we proposed to perform two one-sided tests, to achieve certain power β_2 given level α_2 at alternative parameter, say $\mu = 0$, we could find a minimal sample size for each test separately. This can be realized because each test has a monotone power function. Then adopt the maximum of these two sample sizes and use it as the desired sample size of the second stage. When looking for each sample size computationally, use the same iteration by starting with an initial estimate of n, which can be acquired by using the formula in the first stage. Then repeatedly find the rejection interval expression and find a sample size under that interval until the value of n is stable.

1.3 Generalization to one-parameter exponential family

1.3.1 Binomial distribution

Now consider the model where X_i, $i = 1, \cdots, n_1$ are Bernoulli variables with parameter p. The equivalence problem referring to these observations reads

$$H : 0 < p \leq p_1 \text{ or } p_2 \leq p < 1 \text{ vs } K : p_1 < p < p_2$$

Let $Y_1 = \sum_{i=1}^{n_1} X_i$, and R be the event of rejection based on n_1 observations. From Wellek (2002), R is determined by the following rule:

$$\phi_{Y_1}(y_1) = \begin{cases} 1 & C^1_{\alpha_1}(n_1; p_1, p_2) < y_1 < C^2_{\alpha_1}(n_1; p_1, p_2) \\ \gamma^1_{\alpha_1}(n_1; p_1, p_2) & y_1 = C^1_{\alpha_1}(n_1; p_1, p_2) \\ \gamma^2_{\alpha_1}(n_1; p_1, p_2) & y_1 = C^2_{\alpha_1}(n_1; p_1, p_2) \\ 0 & y_1 < C^1_{\alpha_1}(n_1; p_1, p_2) \text{ or } y_1 > C^2_{\alpha_1}(n_1; p_1, p_2) \end{cases} \tag{1.3.1}$$

where constants $C^\nu_{\alpha_1}(n_1; p_1, p_2)$, $\gamma^\nu_{\alpha_1}(n_1; p_1, p_2)$, $\nu = 1, 2$, are determined by solving

$$\begin{aligned} \sum_{t=C_1+1}^{C_2-1} B(t; n_1, p_1) + \sum_{\nu=1}^{2} \gamma_\nu B(C_\nu; n_1, p_1) &= \alpha_1 \\ \sum_{t=C_1+1}^{C_2-1} B(t; n_1, p_2) + \sum_{\nu=1}^{2} \gamma_\nu B(C_\nu; n_1, p_2) &= \alpha_1 \end{aligned} \tag{1.3.2}$$

where $B(x; n, p)$ is the binomial density function, $0 \leq C_1 < C_2 \leq n_1$, and $0 \leq \gamma_1, \gamma_2 < 1$.

And the power of this test at $p = p_0$, $p_1 < p_0 < p_2$ is

$$\beta_1(p_0) = \sum_{t=C_1+1}^{C_2-1} B(t; n_1, p_0) + \sum_{\nu=1}^{2} \gamma_\nu B(C_\nu; n_1, p_0) \tag{1.3.3}$$

When R occurs, take a second sample of size $n_2 = n - n_1$. Let $Y = \sum_{i=1}^{n} X_i$ then given R, the conditional density of Y can be written as follows,

when $C_1 = C_2$

$$f_Y(y|R; p) = \begin{cases} 0 & y < C_1 \text{ or } y > C_1 + n_2 \\ B(y - C_1; n_2, p) & \text{otherwise} \end{cases}$$

when $C_2 \leq C_1 + n_2$

$$f_Y(y|R;p) = \begin{cases} 0 & y < C_1 \text{ or } y > C_2 + n_2 \\ \frac{\gamma_1 B(C_1;n_1,p)B(y-C_1;n_2,p) + \sum_{t=C_1+1}^{y} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_1 \leq y < C_2 \\ \frac{\sum_{i=1}^{2} \gamma_i B(C_i;n_1,p)B(y-C_i;n_2,p) + \sum_{t=C_1+1}^{C_2-1} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_2 \leq y \leq C_1 + n_2 \\ \frac{\gamma_2 B(C_2;n_1,p)B(y-C_2;n_2,p) + \sum_{t=y-n_2}^{C_2-1} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_1 + n_2 < y \leq C_2 + n_2 \end{cases}$$

when $C_2 = C_1 + n_2 + 1$

$$f_Y(y|R;p) = \begin{cases} 0 & y < C_1 \text{ or } y > C_2 + n_2 \\ \frac{\gamma_1 B(C_1;n_1,p)B(y-C_1;n_2,p) + \sum_{t=C_1+1}^{y} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_1 \leq y \leq C_1 + n_2 \\ \frac{\gamma_2 B(C_2;n_1,p)B(y-C_2;n_2,p) + \sum_{t=y-n_2}^{C_2-1} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_2 \leq y \leq C_2 + n_2 \end{cases}$$

when $C_2 \geq C_1 + n_2 + 2$

$$f_Y(y|R;p) = \begin{cases} 0 & y < C_1 \text{ or } y > C_2 + n_2 \\ \frac{\gamma_1 B(C_1;n_1,p)B(y-C_1;n_2,p) + \sum_{t=C_1+1}^{y} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_1 \leq y \leq C_1 + n_2 \\ \frac{\sum_{t=y-n_2}^{y} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_1 + n_2 < y < C_2 \\ \frac{\gamma_2 B(C_2;n_1,p)B(y-C_2;n_2,p) + \sum_{t=y-n_2}^{C_2-1} B(t;n_1,p)B(y-t;n_2,p)}{P(R;p)} & C_2 \leq y \leq C_2 + n_2 \end{cases}$$

From the above formulas, we can see that

$$f_Y(y|R;p) = K(p)h(y)\exp(y\log(\frac{p}{1-p})) \tag{1.3.4}$$

still belongs to the single parameter exponential family, thus a UMP test exists and can be given as

$$\phi_Y(y) = \begin{cases} 1 & \tilde{C}_1 < y < \tilde{C}_2 \\ \tilde{\gamma_1} & y = \tilde{C}_1 \\ \tilde{\gamma_2} & y = \tilde{C}_2 \\ 0 & y < \tilde{C}_1 \text{ or } y > \tilde{C}_2 \end{cases} \tag{1.3.5}$$

where the constants $\tilde{C}_\nu$, $\tilde{\gamma_\nu}$, $\nu = 1, 2$, are determined by solving

$$\begin{aligned}\tilde{\gamma_1} f(\tilde{C_1}|R; p_1) + \sum_{y=\tilde{C_1}+1}^{\tilde{C_2}-1} f(y|R; p_1) + \tilde{\gamma_2} f(\tilde{C_2}|R; p_1) &= \alpha_2 \\ \tilde{\gamma_1} f(\tilde{C_1}|R; p_2) + \sum_{y=\tilde{C_1}+1}^{\tilde{C_2}-1} f(y|R; p_2) + \tilde{\gamma_2} f(\tilde{C_2}|R; p_2) &= \alpha_2\end{aligned}$$

And the power of this test at $p = p_0,\ p_1 < p_0 < p_2$ is

$$\beta(p_0) = \tilde{\gamma_1} f(\tilde{C_1}|R; p_0) + \sum_{y=\tilde{C_1}+1}^{\tilde{C_2}-1} f(y|R; p_0) + \tilde{\gamma_2} f(\tilde{C_2}|R; p_0) \tag{1.3.6}$$

When it comes to sample size decisions, we notice that unlike the normal distribution scenario, the power function (1.3.3) is not continuous with respect to the sample size n_1. In other words, given a certain significance level α_1 and power β_1 at alternative parameter $p = p_0$, we may not achieve β_1 exactly by adjusting the sample size. Instead, we seek the smallest n_1 such that at level α_1 the power at $p = p_0$ is no less than β_1.

In the second stage when deciding the additional sample size, from (1.3.6), the power function is also discrete with respect to n, which implies we need to adopt the idea used earlier to look for the minimal n that yields a certain power. Notice that both power function (1.3.3) and (1.3.6) are strictly increasing in n_1 and n respectively, which guarantees that a unique solutions exist.

1.3.2 Two independent binomial distributions

To compare two independent binomial distributions, consider the setting where X_i, $i = 1, \ldots, n_{11}$, are Bernoulli variables with parameter p_1 and Y_j, $j = 1, \ldots, n_{12}$, are Bernoulli variables with parameter p_2. The equivalence problem referring to these observations can be expressed as follows

$$H : |p_1 - p_2| \geq \epsilon \quad \text{vs} \quad K : |p_1 - p_2| < \epsilon$$

However, instead of testing the above hypothesis directly, consider another measurement: the odds ratio. Let

$$\gamma_1 = \log(\text{odds ratio}) = \log \frac{p_1(1 - p_2)}{p_2(1 - p_1)}$$

then test the following hypothesis

$$H^{'}: |\gamma_1| \geq \gamma_0 \quad \text{vs} \quad K^{'}: |\gamma_1| < \gamma_0$$

Intuitively speaking, when p_1 and p_2 are close to each other, the odds ratio would be close to 1 and $|\gamma_1|$ would be within a small value, say γ_0. See Appendix 3.2 for detailed derivation of the relationship between ϵ and γ_0 Now let

$$Z_{11} = \sum_{i=1}^{n_{11}} X_i, \quad Z_{12} = \sum_{j=1}^{n_{12}} Y_j, \quad T_1 = Z_{11} + Z_{12} \tag{1.3.7}$$

and R be the rejection region based on the pilot sample. From Lehmann (1986), R can be determined by the following rule:

$$\phi_{T_1}(z_{11}) = \begin{cases} 1 & C_1(t_1) < z_{11} < C_2(t_1) \\ \nu_1(t_1) & z_{11} = C_1(t_1) \\ \nu_2(t_1) & z_{11} = C_2(t_1) \\ 0 & \text{otherwise} \end{cases} \tag{1.3.8}$$

Since the conditional density of Z_{11} given T_1 is

$$Pr(Z_{11} = z_{11}|T_1 = t_1) = C_{t_1}(\gamma_1)\binom{n_{11}}{z_{11}}\binom{n_{12}}{t_1 - z_{11}}\exp(z_{11}\gamma_1)$$

where

$$C_{t_1}(\gamma_1) = \left(\sum_{x^{'}=0}^{\max(t_1,n_{11})}\binom{n_{11}}{x^{'}}\binom{n_{12}}{t_1 - x^{'}}\exp(x^{'}\gamma_1)\right)^{-1}$$

the constants $C_i(t_1)$, $\nu_i(t_1)$, $i = 1, 2$, are determined by solving

$$E_{\gamma_1=-\gamma_0}\{\phi_{T_1}(Z_{11})|T_1 = t_1\} = \alpha_1 = E_{\gamma_1=\gamma_0}\{\phi_{T_1}(Z_{11})|T_1 = t_1\} \tag{1.3.9}$$

Based on $n_{11} + n_{12}$ observations, we first calculate the value of T_1 defined by (1.3.7). Then a rejection decision rule is obtained. If the observed value of Z_{11} indicates the significance of the alternative hypothesis, we proceed to collect an additional sample. Otherwise we stop and declare non-equivalence.

When rejection occurs, take a second sample which consists of X_i's of size $n_{21} = n_1 - n_{11}$ and Y_j's of size $n_{22} = n_2 - n_{12}$. Let

$$Z_{21} = \sum_{i=n_{11}+1}^{n_1} X_i, \quad Z_{22} = \sum_{j=n_{12}+1}^{n_2} Y_j, \quad T_2 = Z_{21} + Z_{22},$$

$$Z_1 = \sum_{i=1}^{n_1} X_i, \quad Z_2 = \sum_{j=1}^{n_2} Y_j, \quad T = Z_1 + Z_2. \tag{1.3.10}$$

Then

$$Z_1 = Z_{11} + Z_{21}, \quad Z_2 = Z_{12} + Z_{22}, \quad T = T_1 + T_2$$

Since (Z_{11}, Z_{12}) and (Z_{21}, Z_{22}) are mutually independent, it is easy to see the joint density of $(Z_{11}, Z_{12}, Z_1, Z_2)$ is

$$\begin{aligned} & Pr(Z_{11} = z_{11}, Z_{12} = z_{12}, Z_1 = z_1, Z_2 = z_2) \qquad (Z_{11}, Z_{12}) \in R \\ = \; & Pr(Z_{11} = z_{11}, Z_{12} = z_{12}, Z_{21} = z_1 - z_{11}, Z_{22} = z_2 - z_{12}) \\ = \; & \binom{n_{11}}{z_{11}} \binom{n_{12}}{z_{12}} \binom{n_1 - n_{11}}{z_1 - z_{11}} \binom{n_2 - n_{12}}{z_2 - z_{12}} p_1^{z_1} (1-p_1)^{n_1 - z_1} p_2^{z_2} (1-p_2)^{n_2 - z_2} \end{aligned}$$

Thus the joint density of (Z_{11}, T_1, Z_1, T) is

$$\begin{aligned} & Pr(Z_{11} = z_{11}, T_1 = t_1, Z_1 = z_1, T = t) \\ = \; & \binom{n_{11}}{z_{11}} \binom{n_{12}}{t_1 - z_{11}} \binom{n_1 - n_{11}}{z_1 - z_{11}} \binom{n_2 - n_{12}}{(t - z_1) - (t_1 - z_{11})} \\ & (1-p_1)^{n_1} (1-p_2)^{n_2} \exp(z_1\gamma_1 + t\gamma_2) \end{aligned} \tag{1.3.11}$$

where

$$\gamma_1 = \log\left(\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right) \quad \gamma_2 = \log\left(\frac{p_2}{1-p_2}\right)$$

Then the conditional density of (Z_1, T) given rejection occurs is

$$\begin{aligned} Pr(Z_1 = z_1, T = t|R) &= \frac{(1-p_1)^{n_1}(1-p_2)^{n_2} h(z_1, t) \exp(z_1\gamma_1 + t\gamma_2)}{Pr(R)} \\ &= K(\gamma_1, \gamma_2) h(z_1, t) \exp(z_1\gamma_1 + t\gamma_2) \qquad (Z_1, T) \in R \end{aligned}$$

where

$$h(z_1,t) = \sum_{t_1=0}^{\max(t,n_{11}+n_{12})} \left(\sum_{z_{11}=C_1(t_1)+1}^{C_2(t_1)-1} \binom{n_{11}}{z_{11}} \binom{n_{12}}{t_1-z_{11}} \binom{n_1-n_{11}}{z_1-z_{11}} \right.$$
$$\binom{n_2-n_{12}}{(t-z_1)-(t_1-z_{11})} + \sum_{i=1}^{2} \nu_i(t_1) \binom{n_{11}}{C_i(t_1)} \binom{n_{12}}{t_1-C_i(t_1)}$$
$$\left. \binom{n_1-n_{11}}{z_1-C_i(t_1)} \binom{n_2-n_{12}}{(t-z_1)-(t_1-C_i(t_1))} \right)$$

Then the conditional density of Z_1 given T and given rejection occurs is

$$Pr(Z_1 = z_1 | T = t) = C_t(\gamma_1) h(z_1,t) \exp(z_1\gamma_1) \qquad (Z_1,T) \in R$$

where

$$C_t(\gamma_1) = \left(\sum_{z_1'=0}^{\max(t,n_1)} h(z_1',t) \exp(z_1'\gamma_1) \right)^{-1}$$

Again from Lehmann (1986), a UMPU test is determined by the following rule:

$$\phi_T(z_1) = \begin{cases} 1 & \tilde{C}_1(t) < z_1 < \tilde{C}_2(t) \\ \tilde{\nu}_1(t) & z_1 = \tilde{C}_1(t) \\ \tilde{\nu}_2(t) & z_1 = \tilde{C}_2(t) \\ 0 & \text{otherwise} \end{cases} \tag{1.3.12}$$

And the constants $\tilde{C}_i(t)$, $\tilde{\nu}_i(t)$, $i = 1,2$, are determined by solving

$$E_{\gamma_1=-\gamma_0}\{\phi_T(Z_1)|T=t\} = \alpha_2 = E_{\gamma_1=\gamma_0}\{\phi_T(Z_1)|T=t\} \qquad (Z_1,T) \in R$$

After we reject non-equivalence in the first step based on $n_{11}+n_{12}$ observations of $(X_i, Y_j; i = 1,\cdots,n_{11}, j = 1,\cdots,n_{12})$, we observe additional $n_{21}+n_{22}$ data points $(X_i, Y_j; i = 1,\cdots,n_{21}, j = 1,\cdots,n_{22})$. The value of T defined by (1.3.10) is calculated, and then a rejection rule is obtained. If the observed value of Z_1 indicates that the null hypothesis is rejected, we declare an equivalence between X and Y.

1.3.3 Poisson distribution

When it comes to the bio-equivalence test under a Poisson distribution setting, first let X_i, $i = 1, \cdots, n_{11}$ follow a Poisson distribution with parameter μ, and Y_j, $j = 1, \cdots, n_{12}$ follow another Poisson distribution with parameter ν. Here X_i and Y_j are independent. Then let $\rho = \frac{\mu}{\nu}$ and formulate the equivalence problem based on above observations as follows

$$H : \rho \geq \rho_2 \quad \text{or} \quad \rho \leq \rho_1 \quad \text{vs} \quad K : \rho_1 < \rho < \rho_2$$

As Lehmann (1986) argues, the ratio $\rho = \mu/\nu$ is a reasonable measurement of the difference between two Poisson populations. Analogous to the Binomial setting, let

$$Z_{11} = \sum_{i=1}^{n_{11}} X_i, \quad Z_{12} = \sum_{j=1}^{n_{12}} Y_j, \quad T_1 = Z_{11} + Z_{12}.$$

Then $Z_{11} \sim P(n_{11}\mu)$ and $Z_{12} \sim P(n_{12}\nu)$. Their joint density is

$$\begin{aligned} & Pr(Z_{11} = z_{11}, Z_{12} = z_{12}) \\ = \quad & K(\mu, \nu) h(z_{11}, z_{12}) \exp(z_{11} \log \rho + (z_{11} + z_{12}) \log \nu) \end{aligned}$$

where

$$\begin{aligned} K(\mu, \nu) &= \exp(-n_{11}\mu - n_{12}\nu) \\ h(z_{11}, z_{12}) &= \frac{n_{11}^{z_{11}} n_{12}^{z_{12}}}{z_{11}! z_{12}!} \end{aligned}$$

Let $\lambda = \log \rho$, $\eta = \log \nu$, $\lambda_i = \log \rho_i$, $i = 1, 2$, equivalently we will test

$$H^{'} : \lambda \geq \lambda_2 \quad \text{or} \quad \lambda \leq \lambda_1 \quad \text{vs} \quad K^{'} : \lambda_1 < \lambda < \lambda_2$$

Notice that the conditional density of Z_{11} given T_1 is

$$Pr(Z_{11} = z_{11} | T_1 = t_1) = C_{t_1}(\lambda) h(z_{11}, t_1 - z_{11}) \exp(z_{11}\lambda) \qquad z_{11} \leq t_1$$

where

$$C_{t_1}(\lambda) = \left(\sum_{x^{'}=0}^{\max(t_1, n_{11})} h(x^{'}, t_1 - x^{'}) \exp(x^{'}\lambda) \right)^{-1}$$

Again a UMPU test is decided by the following rule:

$$\phi_{T_1}(z_{11}) = \begin{cases} 1 & C_1(t_1) < z_{11} < C_2(t_1) \\ \omega_1(t_1) & z_{11} = C_1(t_1) \\ \omega_2(t_1) & z_{11} = C_2(t_1) \\ 0 & \text{otherwise} \end{cases} \tag{1.3.13}$$

and the constants $C_i(t_1)$, $\omega_i(t_1)$, $i = 1, 2$, are determined by solving

$$E_{\lambda=\lambda_1}\{\phi_{T_1}(Z_{11})|T_1 = t_1\} = \alpha_1 = E_{\lambda=\lambda_2}\{\phi_{T_1}(Z_{11})|T_1 = t_1\}$$

After rejection occurs in the first stage, an additional sample is taken consisting of X_i's of size $n_{21} = n_1 - n_{11}$ and Y_j's of size $n_{22} = n_2 - n_{12}$. Adopting the same notation from the binomial setting, the joint density of $(Z_{11}, Z_{12}, Z_{21}, Z_{22})$ is

$$\begin{aligned} &Pr(Z_{11} = z_{11}, Z_{12} = z_{12}, Z_{21} = z_{21}, Z_{22} = z_{22}) \qquad (Z_{11}, Z_{12}) \in R \\ =&K_2(\mu, \nu)h_2(z_{11}, z_{12}, z_{21}, z_{22}) \exp((z_{11} + z_{21})\lambda + (z_{11} + z_{12} + z_{21} + z_{22})\eta) \end{aligned}$$

where

$$\begin{aligned} K_2(\mu, \nu) &= \exp(-n_1\mu - n_2\nu) \\ h_2(z_{11}, z_{12}, z_{21}, z_{22}) &= \frac{n_{11}^{z_{11}} n_{12}^{z_{12}} n_{21}^{z_{21}} n_{22}^{z_{22}}}{z_{11}! z_{12}! z_{21}! z_{22}!} \end{aligned}$$

Therefore the joint density of (Z_{11}, T_1, Z_1, T) is

$$\begin{aligned} &Pr(Z_{11} = z_{11}, T_1 = t_1, Z_1 = z_1, T = t) \\ = \quad &K_2(\mu, \nu)h_2'(z_{11}, t_1, z_1, t) \exp(z_1\lambda + t\eta) \end{aligned}$$

where

$$h_2'(z_{11}, t_1, z_1, t) = h_2(z_{11}, t_1 - z_{11}, z_1 - z_{11}, (t - z_1) - (t_1 - z_{11}))$$

Now the conditional joint density of (Z_1, T) given rejection, which can be written as follows

$$\begin{aligned} Pr(Z_1 = z_1, T = t|R) &= \frac{\sum\limits_R Pr(Z_{11} = z_{11}, T_1 = t_1, Z_1 = z_1, T = t)}{Pr(R)} \\ &= K_2'(\lambda, \eta)h_2''(z_1, t) \exp(z_1\lambda + t\eta) \end{aligned}$$

where

$$h_2^{''}(z_1,t) = \sum_{t_1=0}^{\max(t,n_{11}+n_{12})} \left(\sum_{z_{11}=C_1(t_1)+1}^{C_2(t_1)-1} h_2^{'}(z_{11},t_1,z_1,t) + \sum_{i=1}^{2} \omega_i(t_1) h_2^{'}(C_i(t_1),t_1,z_1,t) \right)$$

And

$$Pr((Z_1 = z_1|T=t)|R) = C_t(\lambda) h_2^{''}(z_1,t)\exp(z_1\lambda)$$

where

$$C_t(\lambda) = \left(\sum_{z_1^{'}=0}^{\max(t,n_1)} h_2^{''}(z_1^{'},t)\exp(z_1^{'}\lambda) \right)^{-1}$$

A UMPU test is as follows,

$$\phi_T(z_1) = \begin{cases} 1 & \tilde{C}_1(t) < z_1 < \tilde{C}_2(t) \\ \tilde{\omega}_1(t) & z_1 = \tilde{C}_1(t) \\ \tilde{\omega}_2(t) & z_1 = \tilde{C}_2(t) \\ 0 & \text{otherwise} \end{cases} \tag{1.3.14}$$

and the constants $\tilde{C}_i(t)$, $\tilde{\omega}_i(t)$, $i = 1, 2$, are determined by solving

$$E_{\lambda=\lambda_1}\{(\phi_T(Z_1)|T=t)|R\} = \alpha_2 = E_{\lambda=\lambda_2}\{(\phi_T(Z_1)|T=t)|R\}$$

Therefore, if we could reject the null hypothesis in both stages based on the pilot sample and the additional sample, an equivalence between X and Y is established.

1.4 Two independent normal distributions with unknown means and variances

Now we move on to a more complicated yet important scenario: bio-equivalence of two normal distributions when both mean and variance are unknown. Consider the following model

$$\begin{aligned} X_i &\sim \mathcal{N}(\mu_x, \sigma_x^2) \qquad i = 1, \ldots, m_1 \\ Y_j &\sim \mathcal{N}(\mu_y, \sigma_y^2) \qquad j = 1, \ldots, n_1 \end{aligned} \tag{1.4.1}$$

where X's and Y's are independent.

1.4.1 Background

Both Anderson and Hauck (1990) and Liu and Chow (1992) have pointed out that, under the above distributional assumption, bio-equivalence should not be defined only in terms of the means of two distributions. Rather, the variances should also be compared. Two normal distributions with similar or even identical means, can not be claimed as equivalent if there is quite a difference between their variances. As a matter of fact, the concept of population bio-equivalence introduced by Anderson and Hauck (1990), has gained significant consideration and popularity in pertinent literature. However, in most of the current literature, including Sheiner (1992), Schall and Luus (1993), Schall (1995), Chow, et al. (2003), as well as in the guidance for industry (FDA/CDER, 2001), discussions and related testing methods are focused on a so-called aggregated criteria of population bio-equivalence. Instead of performing tests on separate hypotheses regarding means and variances, the test hypothesis is set up as one single real-valued function of all four parameters(two means and two variances) and compared to a given upper bound. Wellek (2000) lists several unappealing drawbacks of this prevalent aggregated criteria for the assessment of bio-equivalence. Therefore, in this paper, we

adopt the disaggregated hypotheses proposed by Berger and Hsu (1996). To demonstrate bio-equivalence between X and Y, we need to test

$$\begin{aligned} & & \mu_x - \mu_y \le \epsilon_1 \quad \text{or} \quad \mu_x - \mu_y \ge \epsilon_2 \\ H: & & \text{or} \\ & & \sigma_x^2/\sigma_y^2 \le \rho_1 \quad \text{or} \quad \sigma_x^2/\sigma_y^2 \ge \rho_2 \\ \text{vs} & & \\ K: & \quad \epsilon_1 < \mu_x - \mu_y < \epsilon_2 \quad \text{and} \quad \rho_1 < \sigma_x^2/\sigma_y^2 < \rho_2 \end{aligned} \tag{1.4.2}$$

Notice that in Berger and Hsu (1996), some procedures testing both means and variances are discussed, and mistakes regarding type I error control are also mentioned. However, no further discussion or improvement are made. Wellek (2000) also proposes a testing procedure based on disaggregated hypotheses. Our proposal differs in two aspects. On one hand, compared to Wellek (2000), we use an unscaled hypothesis for means instead of a scaled one. On the other hand, our test applies to two independent populations rather than the cross-over design in Wellek (2000).

1.4.2 Two-step testing on bio-equivalence

Once a pilot sample X_i of size m_1 and Y_j of size n_1 is collected from (1.4.1), in order to test hypothesis (1.4.2), first consider the following hypotheses:

$$H^\rho : \sigma_x^2/\sigma_y^2 \le \rho_1 \quad \text{or} \quad \sigma_x^2/\sigma_y^2 \ge \rho_2 \qquad \text{vs} \qquad K^\rho : \rho_1 < \sigma_x^2/\sigma_y^2 < \rho_2 \tag{1.4.3}$$

Wellek (2002) introduced a UMPI level-$\alpha_{1\rho}$ test with respect to the above hypotheses (1.4.3), which has the rejection region as follows

$$\left\{ C^{(1)}_{\alpha_{1\rho};df_1,df_2}(\rho_1,\rho_2) < Q_1 < C^{(2)}_{\alpha_{1\rho};df_1,df_2}(\rho_1,\rho_2) \right\}$$

where $Q_1 = S_{1x}^2/S_{1y}^2$, $df_1 = m_1 - 1$ and $df_2 = n_1 - 1$. The critical constants $C^{(k)}_{\alpha_{1\rho};df_1,df_2}(\rho_1,\rho_2)$, $k = 1, 2$, are determined by the following equations

$$F_{df_1,df_2}(C^{(2)}/\rho_1) - F_{df_1,df_2}(C^{(1)}/\rho_1) = \alpha_{1\rho} = F_{df_1,df_2}(C^{(2)}/\rho_2) - F_{df_1,df_2}(C^{(1)}/\rho_2)$$

where $F_{df_1,df_2}(\cdot)$ denotes the cumulative density function of the standard central F-distribution with (df_1, df_2) degrees of freedom.

If the homogeneity test turns out insignificantly, we stop and declare non-equivalence of the two distributions. Otherwise we proceed to a classic two one-sided tests (TOST) to verify the equivalence of two means, i.e.

$$H^{\delta} : \mu_x - \mu_y \leq \epsilon_1 \quad \text{or} \quad \mu_x - \mu_y \geq \epsilon_2 \quad \text{vs} \quad K^{\delta} : \epsilon_1 < \mu_x - \mu_y < \epsilon_2 \tag{1.4.4}$$

TOST is based on the following rejection rule:

$$\{|D_1| < \epsilon - t_{\alpha_{1\delta},m_1+n_1-2} S_{d_1}\}$$

where $D_1 = \bar{X}_1 - \bar{Y}_1$, $S_{d_1}^2 = (\frac{1}{m_1} + \frac{1}{n_1}) \frac{(m_1-1)S_{1x}^2 + (n_1-1)S_{1y}^2}{m_1+n_1-2}$, and $-\epsilon_1 = \epsilon_2 = \epsilon$.

If the test result is not significant, then we stop and declare non-equivalence of the two underlying distributions.

In order to decide the value of $\alpha_{1\rho}$ and $\alpha_{1\delta}$, we first evaluate the overall rejection probability of the proposed testing procedure at the first step. Denote

$$\rho = \sigma_x^2/\sigma_y^2, \quad \delta = \mu_x - \mu_y$$

then

$$D_1 \sim \mathcal{N}\left(\delta, \left(\frac{\rho}{m_1} + \frac{1}{n_1}\right)\sigma_y^2\right), \quad S_{1x}^2 = Q_1 \times S_{1y}^2$$

and the power function is expressed as follows,

$$\begin{aligned}
&\beta\left(\delta, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta}\right) \\
=&Pr\left(\left\{C^{(1)}_{\alpha_{1\rho};df_1,df_2}(\rho_1,\rho_2) < Q_1 < C^{(2)}_{\alpha_{1\rho};df_1,df_2}(\rho_1,\rho_2)\right\}\right. \\
&\left.\cap\{|D_1| < \epsilon - t_{\alpha_{1\delta},m_1+n_1-2} S_{d_1}\}\right) \\
=&\int_0^{\infty} \int_{C^{(1)}_{\alpha_{1\rho}}}^{C^{(2)}_{\alpha_{1\rho}}} Pr\left(|D_1| < \epsilon - t_{\alpha_{1\delta},m_1+n_1-2} S_{d_1} | Q_1, S_{1y}^2\right) f_{\rho,\sigma_y^2}\left(q_1, S_{1y}^2\right) dq_1 dS_{1y}^2 \\
=&\int_0^{\infty} \int_{C^{(1)}_{\alpha_{1\rho}}}^{C^{(2)}_{\alpha_{1\rho}}} \left(\Phi\left(\frac{\epsilon + \delta - t_{\alpha_{1\delta}} S_{d_1}}{\sqrt{(\frac{\rho}{m_1} + \frac{1}{n_1})\sigma_y^2}}\right) + \Phi\left(\frac{\epsilon - \delta - t_{\alpha_{1\delta}} S_{d_1}}{\sqrt{(\frac{\rho}{m_1} + \frac{1}{n_1})\sigma_y^2}}\right) - 1\right) \times \\
&I\left(\epsilon > t_{\alpha_{1\delta}} S_{d_1}\right) f_{\rho,\sigma_y^2}\left(q_1, S_{1y}^2\right) dq_1 dS_{1y}^2
\end{aligned}$$

where

$\alpha_{1\rho}$ ∼nominal level of test for variances.

$\alpha_{1\delta}$ ∼nominal level of test for means.

$$S_{d_1}^2 = \left(\frac{1}{m_1} + \frac{1}{n_1}\right) \frac{((m_1-1)Q_1 + (n_1-1))S_{1y}^2}{m_1+n_1-2}$$

$$f_{\rho,\sigma_y^2}(q, S_{1y}^2) = \frac{(\frac{m_1-1}{2})^{\frac{m_1-1}{2}}(\frac{n_1-1}{2})^{\frac{n_1-1}{2}}\rho^{-\frac{m_1-1}{2}}(\sigma_y^2)^{-\frac{m_1+n_1-2}{2}}}{\Gamma(\frac{m_1-1}{2})\Gamma(\frac{n_1-1}{2})} q^{\frac{m_1-3}{2}}(S_{1y}^2)^{\frac{m_1+n_1-4}{2}}$$
$$\times \exp\left\{-\left[\frac{(m_1-1)q}{\rho} + (n_1-1)\right]\frac{S_{1y}^2}{2\sigma_y^2}\right\}$$

First notice that the power function $\beta(\delta, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta})$ is strictly decreasing in $|\delta|$, because $f_{\rho,\sigma_y^2}(q, S_{1y}^2)$ is independent of δ and the first multiplier of the integrand in the power function is strictly decreasing in $|\delta|$. Then we performed intensive simulations to study the monotonicity of the power function w.r.t. ρ. The results indicated that $\beta(\delta, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta})$ increases then decreases as ρ increases.

To control type I error of the proposed two-step test, consider the null parameter space, which consists of three subsets, including $H^\delta \cap H^\rho$, $H^\delta \cap K^\rho$ and $K^\delta \cap H^\rho$. Or rewrite them as two subsets:

$$H^I = \{(\delta, \rho) : \rho \geq \rho_2 \text{ or } \rho \leq \rho_1\} \quad \text{and} \quad H^{II} = \{(\delta, \rho) : |\delta| \geq \epsilon,\ \rho_1 < \rho < \rho_2\}$$

Based on the monotonicity property of δ, and simulation results for ρ, for any given σ_y^2 (which is a nuisance parameter here) and a significance level α_1, we suggest looking for a pair of $\alpha_{1\rho}$ and $\alpha_{1\delta}$ which maximizes the power function

$$\beta(0, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta}) \qquad \rho_1 < \rho < \rho_2$$

such that

$$\max_{H^I} \beta(\delta, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta}) = \max_{\rho \geq \rho_2 \text{ or } \rho \leq \rho_1} \beta(0, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta}) \leq \alpha_1$$

$$\max_{H^{II}} \beta(\delta, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta}) = \max_{\rho_1 < \rho < \rho_2} \beta(\epsilon, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta}) \leq \alpha_1$$

Once $\alpha_{1\rho}$ and $\alpha_{1\delta}$ are chosen, rejection region can be constructed and decision can be made based upon observed statistics.

1.4.3 Two-step testing under two-stage setting: First Stage

From the discussion above, we notice that the value of $\alpha_{1\rho}$ and $\alpha_{1\delta}$ vary with respect to different values of σ_y^2, which is usually an unknown parameter and makes application difficult. However, the good news is, in practice, very often one of the two scenarios occurs in the first stage of bio-equivalence testing with two-stage setting.

1. A pilot sample is already collected and required further analyses, i.e.
   - What's the power given a certain level in the first stage;
   - Whether an additional sample is needed based on the test result in the first stage;
   - What are the overall Type I error and power.
2. No sample has been collected at all. Instead, a desirable Type I error and power for pilot study are specified, and the pilot sample size is asked as very first instruction of the whole test.

As a matter of fact, our proposed two-step test is feasible in both cases, though a little different when performing it. The first case is easier, in which either of the two samples can be treated as from population "Y" and its sample variance s_y^2 is calculated and can be used as an estimator of σ_y^2. The second case is harder but still achievable, considering that in a lot of scenarios, bio-equivalence test is called upon to compare a new formulation with a "well-known" or "widely-used" formulation, which can be treated as population "Y" in our test. In other words, either "Y" is studied enough to assume that its variance is known or at least we already have ample observations from "Y" in hand to calculate its sample variance s_y^2 as a reliable estimator of σ_y^2. Simulation

Table 1.2: Sample size decision table when $\sigma_y^2 = 0.2$

α_1	ρ_0	sample size	power	$\alpha_{1\rho}$	$\alpha_{1\delta}$	α_1	ρ_0	sample size	power	$\alpha_{1\rho}$	$\alpha_{1\delta}$
0.2	2	10	0.29	0.24	0.52	0.3	2	10	0.43	0.35	0.56
		13	0.33	0.23	0.46			13	0.48	0.34	0.51
		16	0.37	0.23	0.41			16	0.52	0.34	0.44
		19	0.43	0.24	0.35			19	0.57	0.33	0.43
		22	0.47	0.23	0.32			22	0.62	0.32	0.41
		25	0.51	0.21	0.32			25	0.66	0.32	0.37
		28	0.56	0.22	0.29			28	0.7	0.31	0.37
		31	0.6	0.22	0.27			31	0.74	0.31	0.36
		34	0.64	0.21	0.26			34	0.78	0.3	0.35
		37	0.68	0.21	0.25			37	0.8	0.31	0.33
		40	0.72	0.21	0.24			40	0.83	0.3	0.33
	3	10	0.34	0.29	0.32		3	10	0.49	0.37	0.4
		13	0.4	0.27	0.28			13	0.55	0.37	0.36
		16	0.45	0.26	0.25			16	0.61	0.33	0.34
		19	0.51	0.26	0.23			19	0.67	0.3	0.33
		22	0.57	0.24	0.22			22	0.72	0.33	0.31
		25	0.62	0.23	0.21			25	0.77	0.32	0.31
		28	0.68	0.22	0.21			28	0.8	0.26	0.31
		31	0.71	0.17	0.21			31	0.84	0.31	0.3
		34	0.76	0.2	0.2			34	0.87	0.28	0.3
		37	0.79	0.21	0.2			37	0.89	0.29	0.3
		40	0.82	0.18	0.2			40	0.91	0.28	0.3
	4	10	0.35	0.26	0.29		4	10	0.5	0.39	0.36
		13	0.41	0.28	0.25			13	0.56	0.37	0.33
		16	0.47	0.25	0.23			16	0.63	0.32	0.32
		19	0.52	0.18	0.22			19	0.69	0.34	0.31
		22	0.59	0.23	0.21			22	0.74	0.32	0.3
		25	0.64	0.12	0.21			25	0.78	0.32	0.3
		28	0.69	0.21	0.2			28	0.82	0.31	0.3
		31	0.73	0.21	0.2			31	0.85	0.31	0.3
		34	0.77	0.2	0.2			34	0.87	0.25	0.3
		37	0.81	0.2	0.2			37	0.89	0.3	0.3
		40	0.83	0.2	0.2			40	0.91	0.25	0.3

Table 1.3: Sample size decision table when $\sigma_y^2 = 0.5$

α_1	ρ_0	sample size	power	$\alpha_{1\rho}$	$\alpha_{1\delta}$	α_1	ρ_0	sample size	power	$\alpha_{1\rho}$	$\alpha_{1\delta}$
0.2	2	10	0.25	0.34	0.47	0.3	2	10	0.37	0.48	0.51
		13	0.27	0.31	0.44			13	0.39	0.5	0.46
		16	0.29	0.36	0.37			16	0.42	0.45	0.44
		19	0.3	0.38	0.33			19	0.44	0.42	0.42
		22	0.33	0.37	0.31			22	0.47	0.44	0.39
		25	0.35	0.36	0.29			25	0.49	0.4	0.38
		28	0.37	0.34	0.28			28	0.52	0.42	0.36
		31	0.39	0.31	0.27			31	0.55	0.41	0.35
		34	0.41	0.3	0.26			34	0.57	0.41	0.34
		37	0.44	0.3	0.25			37	0.6	0.36	0.34
		40	0.46	0.29	0.24			40	0.63	0.38	0.33
	3	10	0.25	0.47	0.37		3	10	0.38	0.46	0.45
		13	0.27	0.41	0.34			13	0.4	0.46	0.41
		16	0.29	0.24	0.33			16	0.43	0.4	0.39
		19	0.31	0.27	0.3			19	0.45	0.38	0.37
		22	0.33	0.18	0.29			22	0.48	0.45	0.35
		25	0.35	0.36	0.26			25	0.51	0.43	0.34
		28	0.38	0.33	0.25			28	0.54	0.43	0.33
		31	0.4	0.33	0.24			31	0.56	0.41	0.32
		34	0.42	0.12	0.24			34	0.59	0.29	0.32
		37	0.45	0.14	0.23			37	0.61	0.15	0.32
		40	0.47	0.29	0.22			40	0.64	0.36	0.31
	4	10	0.25	0.21	0.39		4	10	0.38	0.46	0.43
		13	0.27	0.14	0.36			13	0.4	0.28	0.41
		16	0.29	0.26	0.31			16	0.43	0.31	0.38
		19	0.31	0.23	0.29			19	0.45	0.33	0.36
		22	0.33	0.1	0.28			22	0.48	0.21	0.35
		25	0.35	0.16	0.26			25	0.51	0.16	0.34
		28	0.38	0.13	0.25			28	0.54	0.19	0.33
		31	0.4	0.15	0.24			31	0.56	0.38	0.32
		34	0.42	0.29	0.23			34	0.58	0.37	0.31
		37	0.44	0.29	0.22			37	0.61	0.3	0.31
		40	0.48	0.23	0.22			40	0.64	0.36	0.31

Table 1.4: Sample size decision table when $\sigma_y^2 = 1$

α_1	ρ_0	sample size	power	$\alpha_{1\rho}$	$\alpha_{1\delta}$	α_1	ρ_0	sample size	power	$\alpha_{1\rho}$	$\alpha_{1\delta}$
0.2	2	10	0.22	0.41	0.49	0.3	2	10	0.34	0.65	0.51
		13	0.24	0.45	0.44			13	0.35	0.59	0.49
		16	0.24	0.51	0.4			16	0.36	0.61	0.46
		19	0.25	0.25	0.45			19	0.38	0.54	0.45
		22	0.26	0.49	0.36			22	0.38	0.43	0.45
		25	0.27	0.27	0.38			25	0.4	0.42	0.43
		28	0.28	0.37	0.34			28	0.41	0.52	0.4
		31	0.29	0.21	0.36			31	0.42	0.35	0.41
		34	0.3	0.39	0.31			34	0.44	0.49	0.38
		37	0.31	0.23	0.32			37	0.45	0.47	0.37
		40	0.32	0.38	0.29			40	0.46	0.51	0.36
	3	10	0.23	0.52	0.43		3	10	0.34	0.43	0.51
		13	0.23	0.54	0.4			13	0.35	0.47	0.47
		16	0.24	0.23	0.4			16	0.36	0.41	0.45
		19	0.25	0.15	0.39			19	0.37	0.28	0.44
		22	0.26	0.27	0.35			22	0.39	0.51	0.41
		25	0.27	0.46	0.33			25	0.4	0.35	0.4
		28	0.28	0.3	0.32			28	0.41	0.3	0.39
		31	0.29	0.2	0.31			31	0.42	0.25	0.38
		34	0.3	0.19	0.3			34	0.44	0.27	0.37
		37	0.31	0.18	0.29			37	0.45	0.41	0.36
		40	0.32	0.3	0.28			40	0.46	0.1	0.36
	4	10	0.23	0.56	0.42		4	10	0.34	0.61	0.48
		13	0.23	0.35	0.4			13	0.35	0.22	0.48
		16	0.24	0.1	0.4			16	0.36	0.1	0.48
		19	0.25	0.11	0.37			19	0.37	0.12	0.44
		22	0.26	0.1	0.35			22	0.39	0.11	0.42
		25	0.27	0.22	0.33			25	0.4	0.15	0.4
		28	0.26	0.41	0.31			28	0.4	0.47	0.38
		31	0.28	0.37	0.3			31	0.42	0.53	0.37
		34	0.29	0.28	0.29			34	0.43	0.48	0.36
		37	0.29	0.39	0.28			37	0.45	0.17	0.36
		40	0.3	0.33	0.27			40	0.46	0.26	0.35

shows that even with small sample size (around 10 for each population), the rejection region derived from given σ_y^2 and from calculated s_y^2 are almost overlapping.

In Table 1.2 - Table 1.4, we provide numerical study results as a reference for a sample size decision. For computational simplicity, we take an equal size of sample from both populations, i.e. $m_1 = n_1$. Also we assume that $\epsilon = \log(10/8)$, $1/\rho_1 = \rho_2 = \rho_0$, and σ_y^2 is known.

In practice, once σ_y^2 is acquired or estimated, our computational technique can be adopted to produce similar tables as provided in order to calculate the optimal values of $\alpha_{1\rho}$ and $\alpha_{1\delta}$ corresponding to different sample size and power. On the other hand, if a speedy preliminary test in the first stage is preferred and the control of Type I error is not that crucial, our computation results can serve as benchmarks for sample size decisions, considering the strictly decreasing property of the power function with respect to σ_y^2.

1.4.4 Two-step testing under two-stage setting: Second Stage

Once rejection occurs in the first stage, a larger sample is called upon for further confirmation, which consists of X_i's of size $m_2 = m - m_1$ and Y_j's of size $n_2 = n - n_1$. Consider the conditional joint distribution given rejection occurs in the first stage,

$$f(\bar{X}, \bar{Y}, S_x^2, S_y^2 | R) = \frac{f(\bar{X}, \bar{Y}, S_x^2, S_y^2, R)}{Pr(R)} = \frac{f(\bar{X}, \bar{Y}, S_x^2, S_y^2, R)}{\beta(\delta, \rho, \sigma_y^2; \alpha_{1\rho}, \alpha_{1\delta})} \ (\bar{X}, \bar{Y}, S_x^2, S_y^2) \in R \tag{1.4.5}$$

where

$$R = \left\{ C^{(1)}_{\alpha_{1\rho}; df_1, df_2}(\rho_1, \rho_2) < Q_1 < C^{(2)}_{\alpha_{1\rho}; df_1, df_2}(\rho_1, \rho_2) \right\} \cap \{ |D_1| < \epsilon - t_{\alpha_{1\delta}, m_1+n_1-2} S_{d_1} \} \tag{1.4.6}$$

and

$$f(\bar{X}, \bar{Y}, S_x^2, S_y^2, R) = \int_R f\left(\bar{X}, \bar{Y}, S_x^2, S_y^2, \bar{X}_1, \bar{Y}_1, S_{1x}^2, S_{1y}^2\right) d\bar{x_1} d\bar{y_1} dS_{1x}^2 dS_{1y}^2$$

Since X_i's and Y_j's are independent, we only need to derive the joint distribution of $\left(\bar{X}, S_x^2, \bar{X}_1, S_{1x}^2\right)$ from the joint distribution of $\left(\bar{X}_1, S_{1x}^2, \bar{X}_2, S_{2x}^2\right)$ and the analogous result for the Y part would follow. Also it is easy to derive the following relationships:

$$
\begin{aligned}
m_1\bar{X}_1 + m_2\bar{X}_2 &= m\bar{X} \\
n_1\bar{Y}_1 + n_2\bar{Y}_2 &= n\bar{Y} \\
(m_1-1)S_{1x}^2 + (m_2-1)S_{2x}^2 &= (m-1)S_x^2 - m_1(\bar{X}-\bar{X}_1)^2 - m_2(\bar{X}-\bar{X}_2)^2 \\
(n_1-1)S_{1y}^2 + (n_2-1)S_{2y}^2 &= (n-1)S_y^2 - n_1(\bar{Y}-\bar{Y}_1)^2 - n_2(\bar{Y}-\bar{Y}_2)^2
\end{aligned}
$$

From the following distributions and linear transformations

$$
\begin{aligned}
&\bar{X}_1 \sim \mathcal{N}\left(\mu_x, \frac{\sigma_x^2}{m_1}\right) \qquad \bar{X}_2 \sim \mathcal{N}\left(\mu_x, \frac{\sigma_x^2}{m_2}\right) \\
&S_{1x}^2 \sim \frac{\sigma_x^2}{(m_1-1)}\chi_{m_1-1}^2 \qquad S_{2x}^2 \sim \frac{\sigma_x^2}{(m_2-1)}\chi_{m_2-1}^2 \\
&\bar{X}_1 = \bar{X}_1 \\
&\bar{X}_2 = \frac{m}{m_2}\bar{X} - \frac{m_1}{m_2}\bar{X}_1 \\
&S_{1x}^2 = S_{1x}^2 \\
&S_{2x}^2 = \frac{m-1}{m_2-1}S_x^2 - \frac{m_1-1}{m_2-1}S_{1x}^2 - \frac{m_1 m}{m_2(m_2-1)}(\bar{X}-\bar{X}_1)^2
\end{aligned}
$$

The joint distribution of $\left(\bar{X}, S_x^2, \bar{X}_1, S_{1x}^2\right)$ can be written as follows,

$$
\begin{aligned}
f(\bar{X}, S_x^2, \bar{X}_1, S_{1x}^2) =& C(m_1, m_2, m)K(\sigma_x^2; m)H(\bar{X}, S_x^2, \bar{X}_1, S_{1x}^2; m_1, m_2, m) \\
&\times \exp\left\{-\frac{m}{2\sigma_x^2}(\bar{X}-\mu_x)^2 - \frac{m-1}{2\sigma_x^2}S_x^2\right\} \qquad (1.4.7)
\end{aligned}
$$

where

$$
\begin{aligned}
&K(\sigma_x^2; m) = \left(\frac{1}{2\sigma_x^2}\right)^{m/2} \\
&C(m_1, m_2, m) = \frac{m(m-1)\sqrt{m_1 m_2}}{\pi\Gamma(\frac{m_1-1}{2})\Gamma(\frac{m_2-1}{2})m_2(m_2-1)}(m_1-1)^{\frac{m_1-1}{2}}(m_2-1)^{\frac{m_2-1}{2}} \\
&H(\bar{X}, S_x^2, \bar{X}_1, S_{1x}^2; m_1, m_2, m) \\
=&(S_{1x}^2)^{\frac{m_1-3}{2}}\left(\frac{m-1}{m_2-1}S_x^2 - \frac{m_1-1}{m_2-1}S_{1x}^2 - \frac{m_1 m}{m_2(m_2-1)}(\bar{X}-\bar{X}_1)^2\right)^{\frac{m_2-3}{2}}
\end{aligned}
$$

Therefore, the joint density of $\left(\bar{X}, \bar{Y}, S_x^2, S_y^2, \bar{X}_1, \bar{Y}_1, S_{1x}^2, S_{1y}^2\right)$ is

$$\begin{aligned}
& f(\bar{X}, \bar{Y}, S_x^2, S_y^2, \bar{X}_1, \bar{Y}_1, S_{1x}^2, S_{1y}^2) \\
= \; & C(m_1, m_2, m)C(n_1, n_2, n)K(\sigma_x^2; m)K(\sigma_y^2; n) \\
& \times H(\bar{X}, S_x^2, \bar{X}_1, S_{1x}^2; m_1, m_2, m)H(\bar{Y}, S_y^2, \bar{Y}_1, S_{1y}^2; n_1, n_2, n) \\
& \times \exp\left\{-\frac{m}{2\sigma_x^2}(\bar{X} - \mu_x)^2 - \frac{n}{2\sigma_y^2}(\bar{Y} - \mu_y)^2\right\} \exp\left\{-\frac{m-1}{2\sigma_x^2}S_x^2 - \frac{n-1}{2\sigma_y^2}S_y^2\right\}
\end{aligned}$$

Notice that the rejection region R (1.4.6), from the first stage, is expressed in terms of $(D_1, Q_1, S_{d_1}^2)$. Using the following transformations,

$$S_{1x}^2 = Q_1 S_{1y}^2 \qquad \bar{X}_1 = D_1 + \bar{Y}_1$$

now get

$$\begin{aligned}
& f(\bar{X}, \bar{Y}, S_x^2, S_y^2, D_1, \bar{Y}_1, Q_1, S_{1y}^2) \\
= \; & C(m_1, m_2, m)C(n_1, n_2, n)K(\sigma_x^2; m)K(\sigma_y^2; n) \\
& \times H\left(\bar{X}, S_x^2, (D_1 + \bar{Y}_1), Q_1 S_{1y}^2; m_1, m_2, m\right) H\left(\bar{Y}, S_y^2, \bar{Y}_1, S_{1y}^2; n_1, n_2, n\right) S_{1y}^2 \\
& \times \exp\left\{-\frac{m}{2\sigma_x^2}(\bar{X} - \mu_x)^2 - \frac{n}{2\sigma_y^2}(\bar{Y} - \mu_y)^2\right\} \exp\left\{-\frac{m-1}{2\sigma_x^2}S_x^2 - \frac{n-1}{2\sigma_y^2}S_y^2\right\}
\end{aligned}$$

Therefore

$$\begin{aligned}
& f(\bar{X}, \bar{Y}, S_x^2, S_y^2 | R) \\
= & \int_R f(\bar{X}, \bar{Y}, S_x^2, S_y^2, D_1, \bar{Y}_1, Q_1, S_{1y}^2) d\bar{d_1} d\bar{y_1} dq_1 dS_{1y}^2 \\
= & \int_{-\infty}^{\infty} \int_0^{\infty} \int_{C_{\alpha_{1\rho}}^{(1)}}^{C_{\alpha_{1\rho}}^{(2)}} \int_{d_1^{\min}}^{d_1^{\max}} f(\bar{X}, \bar{Y}, S_x^2, S_y^2, D_1, \bar{Y}_1, Q_1, S_{1y}^2) d\bar{d_1} dq_1 dS_{1y}^2 d\bar{y_1} \\
= & C(m_1, m_2, m)C(n_1, n_2, n)K(\sigma_x^2; m)K(\sigma_y^2; n)\tilde{H}(\bar{X}, \bar{Y}, S_x^2, S_y^2) \\
& \times \exp\left\{-\frac{m}{2\sigma_x^2}(\bar{X} - \mu_x)^2 - \frac{n}{2\sigma_y^2}(\bar{Y} - \mu_y)^2\right\} \exp\left\{-\frac{m-1}{2\sigma_x^2}S_x^2 - \frac{n-1}{2\sigma_y^2}S_y^2\right\}
\end{aligned}$$

where

$$
\begin{aligned}
d_1^{\min} =& t_{\alpha_{1\delta},m_1+n_1-2}S_{d_1} - \epsilon \\
d_1^{\max} =& \epsilon - t_{\alpha_{1\delta},m_1+n_1-2}S_{d_1} \\
S_{d_1}^2 =& \left(\frac{1}{m_1}+\frac{1}{n_1}\right)\frac{(m_1-1)Q_1S_{1y}^2+(n_1-1)S_{1y}^2}{m_1+n_1-2} \\
\tilde{H}(\bar{X},\bar{Y},S_x^2,S_y^2) =& \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{C_{\alpha_{1\rho}}^{(1)}}^{C_{\alpha_{1\rho}}^{(2)}}\int_{d_1^{\min}}^{d_1^{\max}} S_{1y}^2 H(\bar{X},S_x^2,(D_1+\bar{Y_1}),Q_1S_{1y}^2;m_1,m_2,m) \\
& \times H(\bar{Y},S_y^2,\bar{Y_1},S_{1y}^2;n_1,n_2,n)d\bar{d_1}dq_1dS_{1y}^2d\bar{y_1}
\end{aligned}
$$

As we did in the first stage, now let

$$D = \bar{X}-\bar{Y}, \qquad Q = S_x^2/S_y^2$$

Then

$$
\begin{aligned}
& f(D,Q,\bar{Y},S_y^2|R) = \frac{f(D,Q,\bar{Y},S_y^2,R)}{\beta(\delta,\rho,\sigma_y^2;\alpha_{1\rho},\alpha_{1\delta})} \\
=& \frac{C(m_1,m_2,m)C(n_1,n_2,n)K(\rho\sigma_y^2;m)K(\sigma_y^2;n)}{\beta(\delta,\rho,\sigma_y^2;\alpha_{1\rho},\alpha_{1\delta})}\tilde{H}\left((D+\bar{Y}),\bar{Y},QS_y^2,S_y^2\right)S_y^2 \\
& \times \exp\left\{-\frac{m}{2\rho\sigma_y^2}((D+\bar{Y})-(\mu_y+\delta))^2-\frac{n}{2\sigma_y^2}(\bar{Y}-\mu_y)^2\right\} \\
& \times \exp\left\{-\frac{S_y^2}{2\sigma_y^2}\left((m-1)\frac{Q}{\rho}+(n-1)\right)\right\} \qquad (1.4.8)
\end{aligned}
$$

Unfortunately, from the above joint density (1.4.8), it is difficult to derive an analytical expression for a marginal density of either D or Q or their joint density given rejection in the first stage. Alternatively, we are able to examine their density curves based on intensive simulation. Numerical results indicates that analogous to the unconditional density of D_1 in the first stage, which follows a normal distribution,

$$f(D;\delta|R) = f(-D;-\delta|R)$$

Also when $m=n$, similar to Q_1,

$$f(Q;\rho|R) = f(\frac{1}{Q};\frac{1}{\rho}|R)$$

Inspired by the similar distribution shapes, we propose a test procedure in the second stage. After an additional sample is collected, a test based on observed Q to check homogeneity of variances again is carried out and the rejection region is of the form $\left\{C_{2\rho}^{(1)} < Q < C_{2\rho}^{(2)}\right\}$, where constants $C_{2\rho}^{(1)}$ and $C_{2\rho}^{(2)}$ depend on a given level of $\alpha_{2\rho}$. If the test result is not significant, we stop and declare non-equivalence. Otherwise, a test of equal means based on observed D is performed with a rejection region $\{|D| < C_d\}$, where the constant C_d depends on a given level of $\alpha_{2\delta}$. And a final declaration of bio-equivalence is made if the test is significant.

Next, in order to control the overall Type I error α, we use the same logic when controlling α_1 in the first stage. Two subsets need to be considered:

$$H^I = \{(\delta, \rho) : \rho \geq \rho_2 \quad \text{or} \quad \rho \leq \rho_1\}$$

and

$$H^{II} = \{(\delta, \rho) : |\delta| \geq \epsilon, \rho_1 < \rho < \rho_2\}$$

Particularly, three parameter sets of the boundary of the null parameter space:

$$\{\delta = \epsilon, \rho = 1\}, \quad \{\delta = 0, \rho = \rho_1\}, \quad \{\delta = 0, \rho = \rho_2\} \tag{1.4.9}$$

are relevant.

To control the overall Type I error at level α, we need to find a pair of $(\alpha_{2\rho}, \alpha_{2\delta})$ such that under each of these three-parameter spaces above, the rejection probability does not exceed α_2. Further notice that a final rejection occurs when we reject the null hypothesis in both stages, and the rejection probability in the first stage was already controlled at level α_1. Since

$$\begin{aligned}
\alpha &= Pr_{H_0}\{\text{rejection in both stages}\} \\
&= Pr_{H_0}\{\text{rejection in the first stage}\} \times \\
&\quad\ Pr_{H_0}\{\text{rejection in the second stage given rejection in the first stage}\} \\
&= \alpha_1 \times \alpha_2
\end{aligned}$$

for given α and α_1, α_2 should be set as no more than α/α_1.

The acceptable pair of $(\alpha_{2\rho}, \alpha_{2\delta})$ must exist because of the continuity of each distribution as well as the rejection region, which results in a continuous power function. However, this pair may not be unique. Analogous to how we calculate in the first stage, we seek a pair that maximizes the power of the test while controlling the Type I error. In mathematical terms, given α_1 and α, we look for $\alpha_{2\rho}$ and $\alpha_{2\delta}$ to maximize the power of the test given by

$$Pr_{\{\delta=0,\rho=1\}}\{C_{2\rho}^{(1)} < Q < C_{2\rho}^{(2)} \quad \text{and} \quad |D| < C_d|R\}$$

such that

$$\max_{\substack{\{\delta=\epsilon,\rho=1\} \\ \{\delta=0,\rho=\rho_i;i=1,2\}}} Pr\{C_{2\rho}^{(1)} < Q < C_{2\rho}^{(2)} \quad \text{and} \quad |D| < C_d|R\} \leq \alpha/\alpha_1$$

To realize the proposed procedure above, we perform the following steps to find the desired $\alpha_{2\rho}$ and $\alpha_{2\delta}$ along with the corresponding $C_{2\rho}^{(1)}$, $C_{2\rho}^{(2)}$, and C_d. Again for simplicity, we assume equal sample sizes in both stages and $1/\rho_1 = \rho_2 = \rho_0$. As a result, the rejection region for testing variances is simplified to $\{1/C_q < Q < C_q\}$. Also here we assume that σ_y^2 is known. Notice that this assumption is reasonable since in the second stage we already have observed data from the first stage. Using either the sample variance as an estimator or a pre-knowledge value of σ_y^2 works fine in our procedure. To be consistent, we suggest using the same value of σ_y^2 for both stages. In addition, we need one more parameter μ_y to simulate the joint conditional density of D and Q given rejection. Based on our numerical studies, the sample mean of population Y from the first stage can be used as a good estimator. As a matter of fact, the value of C_q and C_d stay stable while the value of μ_y varies, which is consistent with the fact that the test of the difference between two means would not depend on either of these two means themselves. We next describe the steps of the procedure:

- *Step* 1. For any given $\alpha_{2\rho}$, simulate data from the conditional distribution of

$$Q; \rho_0|R \quad \text{and} \quad Q; 1/\rho_0|R$$

and find a constant based on empirical density curves such that

$$F^Q_{emp}(C_q;\rho_0) - F^Q_{emp}(1/C_q;\rho_0) = \alpha_{2\rho} = F^Q_{emp}(C_q;1/\rho_0) - F^Q_{emp}(1/C_q;1/\rho_0) \tag{1.4.10}$$

Then the corresponding rejection region for Q at level $\alpha_{2\rho}$ is

$$\{1/C_q < Q < C_q\}$$

Note: There are two equations in (1.4.10), which yield two values of C_q. These two values in theory should be identical. Observing the different values due to numerical error, we use the average of them as the final constant.

- *Step* 2. Analogously, for any given $\alpha_{2\delta}$, simulate data from the conditional distribution of

$$D;\epsilon|R \quad \text{and} \quad D;-\epsilon|R$$

and find a constant based on empirical density curves such that

$$F^D_{emp}(C_d;\epsilon) - F^D_{emp}(-C_d;\epsilon) = \alpha_{2\delta} = F^D_{emp}(C_d;-\epsilon) - F^D_{emp}(-C_d;-\epsilon) \tag{1.4.11}$$

Then the corresponding rejection region for D at level $\alpha_{2\delta}$ is

$$\{|D| < C_d\}$$

Note: For a similar reason as *Step* 1, we adopt the average of two C_d's as desired constant.

- *Step* 3. Under the three parameter settings of the null hypothesis boundary (1.4.9), apply the rejection rule decided from the previous two steps to the empirical conditional joint density of D and Q given rejection in the first stage . If any of three yields a rejection probability exceeding α/α_1, this pair of $(\alpha_{2\rho}, \alpha_{2\delta})$ should be discarded. Since the rejection probability is increasing with respect to

either of these two nominal errors, we can repeat either *Step* 1 or 2 or both by decreasing their values gradually. Once under all of three settings, the specified type I error is controlled, the power of this test is calculated analogously from the simulated conditional joint density of D and Q under $\{\delta = 0, \rho = 1\}$.

- *Step* 4. Repeat *Step* 1 through *Step* 3 on the 2×2 grid consisting of two perpendicular intervals (0.01, 0.99) by step = 0.01 to get all candidate pairs of $(\alpha_{2\rho}, \alpha_{2\delta})$. Then the best choice would be the one with the largest power.

Next we provide some numerical study results to demonstrate our test procedure in the second stage. First, Table 1.5 shows that the value of C_q and C_d are not affected a lot by the change of the value of μ_y. There are some variations of the rejection region with respect different values of μ_y. It is likely caused by the numerical error considering all our results are based on simulations and empirical densities. For preciseness, if some prior knowledge of μ_y is available or $\bar{Y}$ can be calculated as an estimator, we suggest using that value as a starting point of the simulation study for the second stage.

Table 1.6 - Table 1.11 provide different rejection values under certain parameter settings. Notice that even with identical total sample size n, the power of entire test is increasing with respect to the sample size of first stage n_1. This result is reasonable since the final rejection is the result that the alternative hypothesis is significant in both stages. If significance is supported in the first stage by a larger sample, it is more likely that it will occur again in the second stage, which reflects a larger rejection probability, a.k.a. power of the test.

Table 1.5: Rejection region decision with difference value of μ_y

σ_y^2	α	α_1	ρ_0	μ_y	power	$\alpha_{2\rho}$	$\alpha_{2\delta}$	C_q	$1/C_q$	C_d
0.5	0.1	0.2	2	-2.7	0.24	0.52	0.51	1.796	0.557	0.175
				10	0.243	0.55	0.51	1.822	0.549	0.176
				3.5	0.242	0.53	0.51	1.812	0.552	0.176
				-13.2	0.241	0.51	0.51	1.796	0.557	0.175
			3	-2.7	0.251	0.82	0.5	3.179	0.315	0.173
				10	0.246	0.83	0.5	3.21	0.311	0.168
				3.5	0.249	0.82	0.5	3.168	0.316	0.171
				-13.2	0.249	0.82	0.5	3.176	0.315	0.173
			4	-2.7	0.245	0.91	0.5	4.691	0.213	0.167
				10	0.249	0.9	0.5	4.642	0.215	0.172
				3.5	0.249	0.9	0.5	4.675	0.214	0.171
				-13.2	0.25	0.9	0.5	4.63	0.216	0.172
		0.3	2	-2.7	0.299	0.45	0.34	1.742	0.574	0.132
				10	0.299	0.45	0.34	1.745	0.573	0.131
				3.5	0.295	0.45	0.34	1.743	0.574	0.128
				-13.2	0.298	0.38	0.35	1.67	0.599	0.133
			3	-2.7	0.3	0.07	0.34	1.865	0.536	0.126
				10	0.308	0.07	0.34	1.875	0.533	0.13
				3.5	0.301	0.07	0.34	1.876	0.533	0.127
				-13.2	0.296	0.45	0.33	2.603	0.384	0.121
			4	-2.7	0.306	0.68	0.33	3.858	0.259	0.125
				10	0.297	0.69	0.33	3.865	0.259	0.12
				3.5	0.303	0.68	0.33	3.837	0.261	0.125
				-13.2	0.303	0.68	0.33	3.84	0.26	0.125

Table 1.6: Rejection region when $\sigma_y^2 = 0.2$, $\rho_0 = 2$ and $\alpha = 0.1$

α_1	n_1	n	power	C_q	$1/C_q$	C_d	α_1	n_1	n	power	C_q	$1/C_q$	C_d
0.2	10	30	0.245	1.658	0.603	0.193	0.3	10	30	0.294	1.484	0.674	0.16
		60	0.287	1.823	0.549	0.201			60	0.399	1.643	0.609	0.173
		90	0.293	1.875	0.533	0.206			90	0.424	1.73	0.578	0.18
		120	0.294	1.907	0.524	0.212			120	0.428	1.772	0.564	0.187
	15	30	0.291	1.533	0.652	0.169		15	30	0.326	1.403	0.713	0.146
		60	0.353	1.733	0.577	0.19			60	0.471	1.606	0.623	0.161
		90	0.362	1.824	0.548	0.198			90	0.503	1.678	0.596	0.175
		120	0.363	1.86	0.538	0.206			120	0.508	1.737	0.576	0.182
	20	60	0.42	1.687	0.593	0.176		20	60	0.536	1.551	0.645	0.155
		80	0.43	1.757	0.569	0.187			80	0.573	1.629	0.614	0.166
		100	0.433	1.811	0.552	0.194			100	0.584	1.667	0.6	0.174
		120	0.434	1.84	0.543	0.198			120	0.59	1.717	0.582	0.178
	25	60	0.489	1.635	0.612	0.167		25	60	0.597	1.532	0.653	0.151
		80	0.504	1.71	0.585	0.181			80	0.641	1.623	0.616	0.158
		100	0.509	1.776	0.563	0.189			100	0.655	1.672	0.598	0.168
		120	0.511	1.811	0.552	0.195			120	0.66	1.7	0.588	0.177
	30	60	0.554	1.571	0.637	0.156		30	60	0.643	1.506	0.664	0.143
		90	0.584	1.699	0.589	0.178			90	0.709	1.612	0.62	0.161
		120	0.589	1.776	0.563	0.19			120	0.721	1.68	0.595	0.172
		150	0.589	1.819	0.55	0.195			150	0.723	1.734	0.577	0.179

Table 1.7: Rejection region when $\sigma_y^2 = 0.2$, $\rho_0 = 3$ and $\alpha = 0.1$

α_1	n_1	n	power	C_q	$1/C_q$	C_d	α_1	n_1	n	power	C_q	$1/C_q$	C_d
0.2	10	30	0.302	2.459	0.407	0.159	0.3	10	30	0.373	2.212	0.452	0.123
		60	0.334	2.695	0.371	0.19			60	0.47	2.464	0.406	0.162
		90	0.338	2.813	0.355	0.201			90	0.483	2.581	0.387	0.175
		120	0.338	2.869	0.349	0.208			120	0.486	2.65	0.377	0.183
	15	30	0.369	2.303	0.434	0.132		15	30	0.425	2.146	0.466	0.102
		60	0.429	2.643	0.378	0.177			60	0.571	2.375	0.421	0.147
		90	0.434	2.783	0.359	0.192			90	0.593	2.524	0.396	0.169
		120	0.435	2.852	0.351	0.201			120	0.596	2.591	0.386	0.178
	20	60	0.526	2.519	0.397	0.163		20	60	0.655	2.316	0.432	0.142
		80	0.535	2.638	0.379	0.178			80	0.681	2.414	0.414	0.155
		100	0.537	2.712	0.369	0.187			100	0.69	2.503	0.4	0.167
		120	0.537	2.757	0.363	0.193			120	0.692	2.562	0.39	0.174
	25	60	0.612	2.426	0.412	0.157		25	60	0.725	2.265	0.442	0.135
		80	0.622	2.549	0.392	0.173			80	0.757	2.402	0.416	0.152
		100	0.624	2.636	0.379	0.181			100	0.767	2.466	0.405	0.163
		120	0.625	2.696	0.371	0.189			120	0.77	2.528	0.396	0.172
	30	60	0.683	2.419	0.413	0.146		30	60	0.769	2.245	0.445	0.13
		90	0.702	2.621	0.382	0.171			90	0.816	2.421	0.413	0.155
		120	0.704	2.738	0.365	0.184			120	0.823	2.526	0.396	0.169
		150	0.704	2.796	0.358	0.193			150	0.824	2.596	0.385	0.177

Table 1.8: Rejection region when $\sigma_y^2 = 0.5$, $\rho_0 = 2$ and $\alpha = 0.1$

α_1	n_1	n	power	C_q	$1/C_q$	C_d	α_1	n_1	n	power	C_q	$1/C_q$	C_d
0.2	10	30	0.179	1.95	0.513	0.17	0.3	10	30	0.186	1.792	0.558	0.119
		60	0.224	1.936	0.516	0.189			60	0.278	1.746	0.573	0.144
		90	0.242	1.943	0.515	0.2			90	0.324	1.765	0.567	0.157
		120	0.249	1.949	0.513	0.206			120	0.348	1.798	0.556	0.169
	15	30	0.185	1.798	0.556	0.139		15	30	0.195	1.787	0.559	0.098
		60	0.243	1.799	0.556	0.175			60	0.298	1.739	0.575	0.13
		90	0.264	1.872	0.534	0.187			90	0.36	1.741	0.574	0.151
		120	0.275	1.886	0.53	0.199			120	0.385	1.784	0.561	0.162
	20	60	0.263	1.773	0.564	0.161		20	60	0.316	1.71	0.585	0.118
		80	0.286	1.827	0.547	0.177			80	0.369	1.719	0.582	0.135
		100	0.297	1.848	0.541	0.184			100	0.403	1.756	0.57	0.146
		120	0.302	1.88	0.532	0.194			120	0.422	1.754	0.57	0.154
	25	60	0.289	1.741	0.575	0.146		25	60	0.34	1.643	0.609	0.111
		80	0.315	1.791	0.558	0.163			80	0.406	1.704	0.587	0.131
		100	0.329	1.817	0.55	0.175			100	0.437	1.685	0.594	0.143
		120	0.339	1.841	0.543	0.185			120	0.459	1.744	0.573	0.15
	30	60	0.305	1.734	0.577	0.128		30	60	0.343	1.658	0.603	0.097
		90	0.352	1.786	0.56	0.159			90	0.446	1.68	0.595	0.125
		120	0.37	1.827	0.547	0.175			120	0.497	1.724	0.58	0.143
		150	0.376	1.861	0.537	0.185			150	0.518	1.756	0.569	0.155

Table 1.9: Rejection region when $\sigma_y^2 = 0.5$, $\rho_0 = 3$ and $\alpha = 0.1$

α_1	n_1	n	power	C_q	$1/C_q$	C_d	α_1	n_1	n	power	C_q	$1/C_q$	C_d
0.2	10	30	0.181	2.357	0.424	0.164	0.3	10	30	0.196	2.313	0.432	0.109
		60	0.226	2.795	0.358	0.191			60	0.281	2.683	0.373	0.137
		90	0.241	2.83	0.353	0.198			90	0.335	2.698	0.371	0.158
		120	0.248	2.88	0.347	0.205			120	0.356	2.73	0.366	0.166
	15	30	0.192	4.529	0.221	0.131		15	30	0.203	2.412	0.415	0.089
		60	0.25	3.171	0.315	0.172			60	0.306	1.876	0.533	0.129
		90	0.271	3.137	0.319	0.189			90	0.363	2.638	0.379	0.147
		120	0.28	3.173	0.315	0.196			120	0.39	2.686	0.372	0.161
	20	60	0.271	4.079	0.245	0.155		20	60	0.323	2.559	0.391	0.111
		80	0.293	4.13	0.242	0.172			80	0.378	2.562	0.39	0.13
		100	0.305	4.106	0.244	0.182			100	0.409	2.594	0.386	0.142
		120	0.312	4.062	0.246	0.189			120	0.43	2.647	0.378	0.153
	25	60	0.294	2.615	0.382	0.139		25	60	0.339	2.493	0.401	0.101
		80	0.32	2.647	0.378	0.159			80	0.404	2.53	0.395	0.121
		100	0.333	2.711	0.369	0.17			100	0.446	2.554	0.392	0.136
		120	0.342	2.748	0.364	0.179			120	0.474	2.598	0.385	0.149
	30	60	0.317	3.696	0.271	0.124		30	60	0.351	1.79	0.559	0.094
		90	0.364	3.898	0.257	0.156			90	0.455	2.715	0.368	0.123
		120	0.381	3.59	0.279	0.171			120	0.507	2.751	0.364	0.142
		150	0.389	3.543	0.282	0.182			150	0.529	2.788	0.359	0.153

Table 1.10: Rejection region when $\sigma_y^2 = 1$, $\rho_0 = 2$ and $\alpha = 0.1$

α_1	n_1	n	power	C_q	$1/C_q$	C_d	α_1	n_1	n	power	C_q	$1/C_q$	C_d
0.2	10	30	0.136	2.216	0.451	0.189	0.3	10	30	0.14	1.997	0.501	0.125
		60	0.167	1.833	0.546	0.197			60	0.189	1.838	0.544	0.134
		90	0.186	2.032	0.492	0.198			90	0.229	1.877	0.533	0.142
		120	0.203	2.018	0.496	0.207			120	0.26	1.862	0.537	0.151
	15	30	0.142	2.124	0.471	0.157		15	30	0.149	1.889	0.529	0.112
		60	0.177	1.798	0.556	0.181			60	0.19	1.786	0.56	0.121
		90	0.202	1.989	0.503	0.192			90	0.239	1.622	0.617	0.138
		120	0.217	1.978	0.506	0.198			120	0.271	1.881	0.532	0.144
	20	60	0.186	1.805	0.554	0.167		20	60	0.194	1.79	0.559	0.109
		80	0.203	1.985	0.504	0.175			80	0.227	1.671	0.598	0.122
		100	0.216	1.981	0.505	0.182			100	0.26	1.587	0.63	0.132
		120	0.228	1.97	0.507	0.191			120	0.286	1.847	0.541	0.138
	25	60	0.184	2.68	0.373	0.147		25	60	0.196	1.728	0.579	0.1
		80	0.207	2.452	0.408	0.161			80	0.239	1.618	0.618	0.116
		100	0.222	2.247	0.445	0.171			100	0.264	1.55	0.645	0.124
		120	0.234	2.185	0.458	0.18			120	0.29	1.887	0.53	0.129
	30	60	0.186	1.747	0.572	0.13		30	60	0.2	1.783	0.561	0.089
		90	0.222	1.894	0.528	0.154			90	0.26	1.633	0.613	0.111
		120	0.247	1.904	0.525	0.168			120	0.302	1.824	0.548	0.124
		150	0.26	1.906	0.525	0.181			150	0.34	1.817	0.55	0.137

Table 1.11: Rejection region when $\sigma_y^2 = 1$, $\rho_0 = 3$ and $\alpha = 0.1$

α_1	n_1	n	power	C_q	$1/C_q$	C_d	α_1	n_1	n	power	C_q	$1/C_q$	C_d
0.2	10	30	0.138	3.465	0.289	0.178	0.3	10	30	0.142	1.766	0.566	0.135
		60	0.168	3.069	0.326	0.191			60	0.186	3.237	0.309	0.128
		90	0.192	3.007	0.333	0.2			90	0.227	3.009	0.332	0.14
		120	0.204	3.002	0.333	0.202			120	0.26	2.945	0.34	0.152
	15	30	0.14	4.177	0.239	0.154		15	30	0.141	1.946	0.514	0.106
		60	0.177	4.334	0.231	0.178			60	0.193	1.791	0.558	0.122
		90	0.199	4.235	0.236	0.189			90	0.242	3.222	0.31	0.136
		120	0.214	4.14	0.242	0.197			120	0.273	3.106	0.322	0.144
	20	60	0.18	1.865	0.536	0.162		20	60	0.193	4.109	0.243	0.106
		80	0.198	3.705	0.27	0.17			80	0.227	4.131	0.242	0.118
		100	0.211	3.489	0.287	0.178			100	0.255	4.11	0.243	0.128
		120	0.225	3.388	0.295	0.186			120	0.281	4.019	0.249	0.137
	25	60	0.185	2.858	0.35	0.14		25	60	0.197	1.833	0.546	0.097
		80	0.206	2.851	0.351	0.157			80	0.233	3.066	0.326	0.11
		100	0.227	2.852	0.351	0.172			100	0.261	2.975	0.336	0.119
		120	0.237	2.872	0.348	0.178			120	0.29	2.901	0.345	0.129
	30	60	0.193	2.94	0.34	0.129		30	60	0.197	1.788	0.559	0.087
		90	0.226	2.923	0.342	0.153			90	0.255	2.858	0.35	0.105
		120	0.249	2.907	0.344	0.172			120	0.309	2.791	0.358	0.127
		150	0.261	2.934	0.341	0.18			150	0.343	2.785	0.359	0.139

Chapter 2
Multiple Testing of Bio-equivalence

2.1 Background

In biochemical analysis, consistent reporting of test results is important to patient care. As clinical laboratories periodically change instruments, reagents or methods, determining whether patient results have shifted or remained the same because of these changes is part of the validation process. If significant changes have occurred, then the normal reference ranges may require adjustments. Several statistical tools exist for determining significant differences, but relatively few determine equivalence. Frequently, statistical differences exist without clinical significance (makes no difference in patient care). Determining equivalence may provide an easier tool for validations in a clinical laboratory.

In Appendix 3.1, data of two types of Iron evaluation: Serum Iron and Heparin Iron, are collected from two instruments: RXL1 and RXL2. In addition, two types of reagent: current and new reagent are adopted in this test. The following 8 questions were raised:

1. For instrument RXL1, are Serum Iron level by current reagent and Serum Iron level by new reagent equivalent?

2. For instrument RXL1, are Serum Iron level by current reagent and Heparin Iron level by new reagent equivalent?

3. For instrument RXL1, are Serum Iron level by new reagent and Heparin Iron level by new reagent equivalent?

4. For instrument RXL2, are Serum Iron level by current reagent and Serum Iron level by new reagent equivalent?

5. For instrument RXL2, are Serum Iron level by current reagent and Heparin Iron level by new reagent equivalent?

6. For instrument RXL2, are Serum Iron level by new reagent and Heparin Iron level by new reagent equivalent?

7. For Serum Iron level by new reagent, are instrument RXL1 and RXL2 equivalent?

8. For Heparin Iron level by new reagent, are instrument RXL1 and RXL2 equivalent?

Our plan is to test these 8 pairwise differences for bio-equivalence. Before doing so however we note that each of the 6 variables measuring iron levels listed in Appendix 3.1 were tested for log-normality. Table 3.2 lists test results of marginal normality for each variable. Both test results and QQ plots in Figure 3.1 indicate marginal normality holds for all 6 variables after log-transformation. In addition, we performed multivariate normality test and the results are listed in Table 3.3. P-values indicate that multivariate normality assumption does not hold at level 5%. However, We want to remark at this time that all statistics in our next discussion will be based on sample means of 38 observations which under the multivariate central limit theorem should be approximately multivariate normal.

Based on results from Table 3.2, we first studied the bio-equivalence questions considering 2 different step-down multiple testing procedures, namely one based on the TOST method of Schuirmann (1987) and the other based on the unbiased test method of Brown, Hwang and Munk (1997). For each of these methods the basic assumption is that differences in the logs of the variables of interest are normally distributed. The test methods involve only the marginal distribution of each sample mean difference. In

addition we use a third method which requires a stronger assumption of multivariate normality of the logs of the 6 variables measured. Since marginal normality is a special case of multivariate normality we develop all 3 methods with the basic assumption of multivariate normality realizing that such an assumption is not necessary for the first 2 methods. This leads us into Section 2.2.

2.2 Statistical Model

Suppose

$$\begin{aligned} X_1 &= \log(\text{Serum Iron level by current reagent for instrument RXL1}) \\ X_2 &= \log(\text{Serum Iron level by new reagent for instrument RXL1}) \\ X_3 &= \log(\text{Heparin Iron level by new reagent for instrument RXL1}) \\ X_4 &= \log(\text{Serum Iron level by current reagent for instrument RXL2}) \\ X_5 &= \log(\text{Serum Iron level by new reagent for instrument RXL2}) \\ X_6 &= \log(\text{Heparin Iron level by new reagent for instrument RXL2}) \end{aligned}$$

Now consider the following statistical scenario:

Let X_{jk}, $j = 1, \cdots, 6$, $k = 1, \cdots, 38$, be a random sample from $\mathcal{N}(\theta, \Sigma)$, where Σ is unknown. Then let $\mu_i = E(Y_i), i = 1, \cdots, 8$, where

1. $Y_{1k} = X_{1k} - X_{2k}$
2. $Y_{2k} = X_{1k} - X_{3k}$
3. $Y_{3k} = X_{2k} - X_{3k}$
4. $Y_{4k} = X_{4k} - X_{5k}$
5. $Y_{5k} = X_{4k} - X_{6k}$
6. $Y_{6k} = X_{5k} - X_{6k}$

7. $Y_{7k} = X_{2k} - X_{5k}$

8. $Y_{8k} = X_{3k} - X_{6k}$

It is clear that the 8 questions in Section 2.1 can be converted as a series of bio-equivalence testing hypotheses

$$H_i : |\mu_i| \geq \Delta \text{ vs } K_i : |\mu_i| < \Delta$$

where $\Delta > 0$ is the tolerance limit usually preset by regulatory agencies.

2.3 Step-wise Multiple testing procedure

2.3.1 Bio-equivalence test on single hypothesis

In Schuirmann (1987), an α-level two one-sided tests(TOST) procedure was proposed and then widely adopted in practice. The rejection region of this procedure is as follows

$$|\bar{Y}| < \Delta - \frac{t_{n-1;1-\alpha} S}{\sqrt{n}} \tag{2.3.1}$$

where $\bar{Y}$ is the sample mean, S is the sample standard deviation, and $t_{n-1;1-\alpha}$ is the lower $(1-\alpha)th$ quantile of a t-distribution with degree of freedom $n-1$.

Despite the simplicity of the TOST procedure, it is a biased test. As a matter of fact, when $\mu = 0$, the power of this test $\to 0$ as $S \to \infty$. To overcome this flaw, Brown, Hwang and Munk (1997) constructed an unbiased, level α test, which is uniformly more powerful than TOST. The construction of their rejection region is recursive with a general form as follows

$$|\bar{Y}| < B(\Delta; \alpha; n) S \tag{2.3.2}$$

where B is some positive function which is never smaller than the right-hand side of (2.3.1). And B increases as α increases as well as Δ increases.

Notice, when deviance is small, the unbiased rejection region and TOST's rejection region are pretty close. The unbiased test is an improvement to TOST when deviance is large.

2.3.2 Step-down testing procedure

A. TOST based step-down procedure

For pre-selected constants $0 < C_1 < \cdots < C_8$,

- *Step* 0. Calculate $T_i = \frac{\Delta - |\bar{Y}_i|}{S_{ii}}, i = 1, \cdots, 8$. Here S_{ii} is a sample standard deviation. Order T_i's and denote as $T_{[1]} \leq T_{[2]} \leq \cdots \leq T_{[8]}$. Also denote $H_{[i]}$ as the hypothesis with respect to statistics $T_{[i]}$.

- *Step* 1. If $T_{[8]} > C_8$ then reject $H_{[8]}$ and go to *Step* 2; Otherwise stop and accept all H_i's.

...

- *Step* q. If $T_{[8-q+1]} > C_{8-q+1}$ then reject $H_{[8-q+1]}$ and go to *Step* q+1; Otherwise stop and accept all rest H_q's.

B. Unbiased test based step-down procedure

For pre-selected constants $0 < C_1 < \cdots < C_8$,

- *Step* 0. Calculate $T_i = \frac{|\bar{Y}_i|}{S_{ii}}, i = 1, \cdots, 8$. Order T_i's and denote as $T_{[1]} \leq T_{[2]} \leq \cdots \leq T_{[8]}$. Also denote $H_{[i]}$ as the hypothesis with respect to statistics $T_{[i]}$.

- *Step* 1. If $T_{[1]} < C_1$ then reject $H_{[1]}$ and go to *Step* 2; Otherwise stop and accept all H_i's.

...

- *Step* q. If $T_{[q]} < C_q$ then reject $H_{[q]}$ and go to *Step* q+1; Otherwise stop and accept all rest H_q's.

Table 2.1: Holm's Proposal

Nominal α	FWER	FDR
0.01	0.011	0.004
0.02	0.018	0.007
0.03	0.028	0.011
0.04	0.035	0.013
0.05	0.043	0.017
0.06	0.047	0.019
0.07	0.055	0.023
0.08	0.062	0.025
0.09	0.069	0.028
0.10	0.073	0.031
0.11	0.082	0.034
0.12	0.086	0.035
0.13	0.093	0.037
0.14	0.102	0.041
0.15	0.105	0.043
0.16	0.109	0.045
0.17	0.116	0.049
0.18	0.122	0.053
0.19	0.128	0.053
0.20	0.137	0.058

Table 2.2: Benjamini's Proposal

Nominal α	FDR	FWER
0.01	0.01	0.032
0.02	0.021	0.06
0.03	0.029	0.088
0.04	0.039	0.104
0.05	0.05	0.127
0.06	0.057	0.148
0.07	0.067	0.169
0.08	0.074	0.188
0.09	0.083	0.204
0.10	0.091	0.217
0.11	0.097	0.231
0.12	0.105	0.244
0.13	0.111	0.26
0.14	0.119	0.276
0.15	0.124	0.288
0.16	0.134	0.294
0.17	0.135	0.309
0.18	0.141	0.324
0.19	0.148	0.33
0.20	0.152	0.34

2.3.3 Application

The testing procedures in Section 2.3.2 were applied to actual data collected in lab to answer questions listed in Section 2.1. Δ was preset at ln1.25 as recommended by regulatory agencies.

To control Type I error at a nominal level of α, we studied the critical values proposed by Holm (1979), which sets $\alpha_i = \alpha/(m+1-i)$ at step $i = 1, \cdots, m$, as well as the critical values proposed by Benjamini and Gavrilov (2009) which sets $\alpha_i = i\alpha/(m+1-i(1-\alpha))$ at step i. Table 2.1 and Table 2.2 list the actual FWER and FDR given nominal α level from 0.01 to 0.2. Table 2.1 shows that if the testing procedures want to control the family-wise error rate (FWER), Holm's proposal should work well. On the other hand, if FWER appears to be too conservative and the false discovery rate (FDR) as a measurement is preferred, then the results in Table 2.2

justified Benjamini and Gavrilov's proposal.

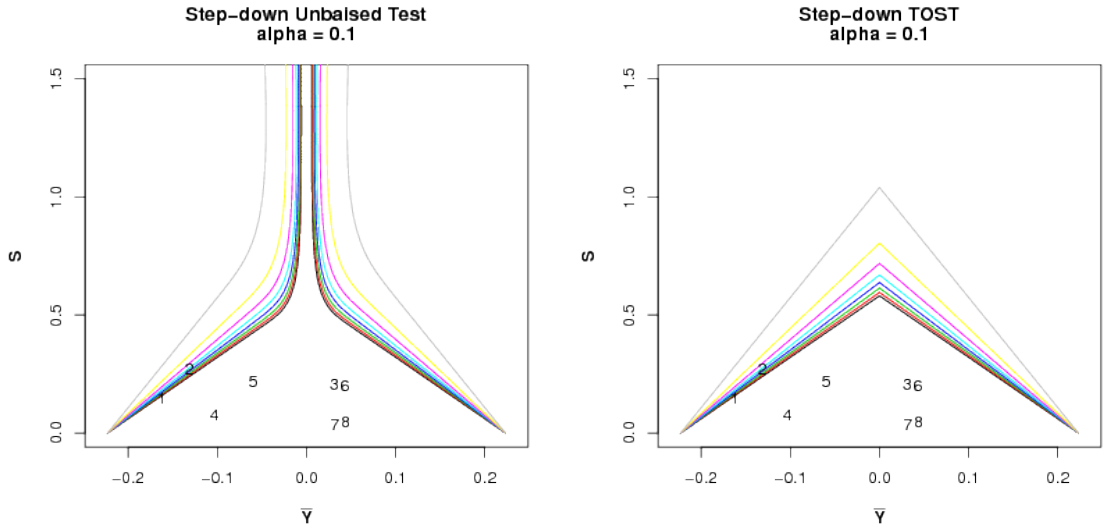


Figure 2.1: Step-down Multiple Testing.(α= 0.1)

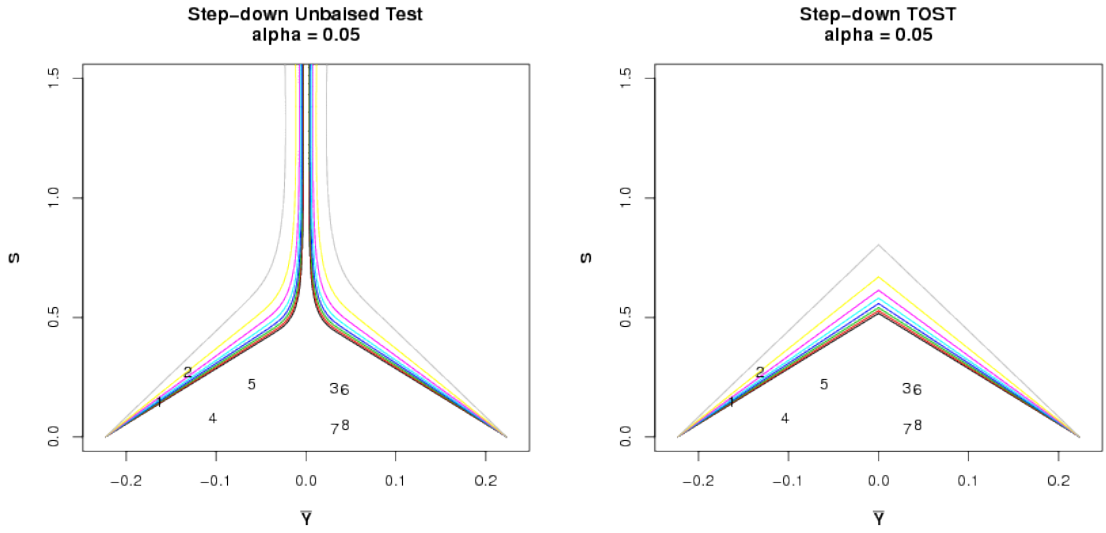


Figure 2.2: Step-down Multiple Testing.(α= 0.05)

Next, we demonstrate the testing procedure by controlling FWER, i.e. we use Holm's critical values. The left side of Figure 2.1 and Figure 2.2 are step-wise procedure based on Brown's unbiased test when α is set at 0.1 and 0.05. And the right side of Figure 2.1 and Figure 2.2 are step-wise procedure based on TOST when nominal α is set at 0.1 and 0.05 respectively. The rejection region in each figure is the area

between two curves/lines that are symmetric w.r.t $\bar{Y} = 0$. Since the step down procedure makes it easier to reject in later steps when fewer hypotheses are left, 8 pairs of curves/lines are plotted in each figure to show a series of enlarging rejection regions. In other words, at step i, from $\bar{Y} = 0$, the i^{th} curve/line from inside to outside on the left and right side represent the rejection region for that step.

Also the observed statistics for each hypothesis are marked in the plots. If the observed statistics fall into the rejection region, then the corresponding hypothesis is rejected.

At level 0.1 and 0.05, all hypotheses were rejected, in an order of 7, 8, 4, 3, 6, 5, 1, 2.

2.4 Two stage multiple testing procedure

2.4.1 Individualized 2-stage testing procedure

In Cohen, Ma and Sackrowitz(2012), a new individualized 2-stage multiple testing procedure was introduced. This procedure treats all hypotheses equally, and tests each in a two-stage fashion. Under the statistical model of section 2, the procedure is as follows: For each $i, i = 1, \cdots, 8$, a 2-stage testing procedure is conducted w.r.t. hypothesis

$$H_i : |\mu_i| \geq \Delta \text{ vs } K_i : |\mu_i| < \Delta$$

- *Stage* 1.
  - *Step* 0. Define

$$U_q = \frac{\sqrt{n}|\bar{Y}_q - \frac{S_{iq}}{S_{ii}}\bar{Y}_i|}{(S_{qq} - \frac{S_{iq}^2}{S_{ii}})^{\frac{1}{2}}}, \quad q = 1, \cdots, 8, q \neq i$$

    where S_{iq} is the sample covariance and S_{ii} is the sample variance.

    Let $0 < C_1 < \cdots < C_7$ be pre-selected constants.

    Order U_q's and denote as $U_{q_{[1]}} \leq U_{q_{[2]}} \leq \cdots \leq U_{q_{[7]}}$

  - *Step* 1. If $U_{q_{[1]}} < C_1$ then reject $H_{q_{[1]}}$ and go to *Step* 2; Otherwise stop and accept all H_q's and set $R_i = 0$.
  - $\cdots$
  - *Step* k. If $U_{q_{[k]}} < C_k$ then reject $H_{q_{[k]}}$ and go to *Step* k+1; Otherwise stop and accept all other H_q's and set $R_i = k - 1$.

- *Stage* 2. Based on R_i, which represents the number of rejections in *Stage* 1, form a confidence interval estimate for μ_i as

$$\bar{Y}_i \pm B(R_i)\sqrt{\frac{S_{ii}}{n}} \tag{2.4.1}$$

  where $B(R_i)$ is decreasing as R_i increases.

  H_i is rejected if the confidence interval estimate above lies entirely in the preset tolerance interval $(-\Delta, \Delta)$.

2.4.2 Application

To apply the two-stage multiple testing procedure on the bio-equivalence problem, a series of simulations were studied to control FDR based on actual data in hand. In stage 1, inspired by Wellek (2002) and Benjamini and Gavrilov (2009), the following critical values perform well

$$C_k = \sqrt{\chi^2_{1;\alpha_{1k}}(((2\Delta)^2)}, k = 1, \cdots, 7$$

where $\chi^2_{1;\alpha_{1k}}(((2\Delta)^2)$ denotes the α_{1k}th quantile of a χ^2-distribution with $df = 1$ and noncentrality parameter $(2\Delta)^2$. And $\alpha_{1k} = \frac{k\alpha_1}{8-k(1-\alpha_1)}$.

In stage 2, a modified version of Holm's critical values works well as follows

$$B(r_i) = t_{n-1;1-\frac{\alpha_2/2}{8-r_i}}$$

where $t_{m;\alpha}$ denotes the αth quantile of a t-distribution with $df = m$.

Table 2.3: Preliminary study when $\alpha_2 = 0.1$

α_1	α_2	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	FDR	FWER
0.2	0.1	$-\Delta$	-2Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	Δ	0.043	0.043
		$-\Delta$	0	Δ	2Δ	Δ	$-\Delta$	Δ	$-\Delta$	0.025	0.042
		2Δ	Δ	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.024	0.045
		Δ	0	$-\Delta$	0	Δ	Δ	$-\Delta$	Δ	0.016	0.042
		$-\Delta$	$-\Delta$	0	Δ	Δ	0	$-\Delta$	$-\Delta$	0.015	0.044
		-2Δ	$-\Delta$	Δ	$-\Delta$	Δ	2Δ	Δ	2Δ	0.015	0.015
		$-\Delta$	0	Δ	$-\Delta$	$-\Delta$	0	Δ	0	0.011	0.038
		$-\Delta$	$-\Delta$	0	Δ	0	$-\Delta$	Δ	0	0.011	0.039
		0	$-\Delta$	$-\Delta$	Δ	$-\Delta$	-2Δ	2Δ	Δ	0.006	0.011
		0	Δ	Δ	0	$-\Delta$	$-\Delta$	$-\Delta$	-3Δ	0.005	0.014
									max	0.043	0.045
0.3	0.1	Δ	3Δ	2Δ	$-\Delta$	Δ	2Δ	Δ	Δ	0.042	0.042
		Δ	0	$-\Delta$	$-\Delta$	Δ	2Δ	-4Δ	$-\Delta$	0.025	0.040
		Δ	Δ	0	$-\Delta$	-3Δ	-2Δ	Δ	$-\Delta$	0.022	0.044
		$-\Delta$	$-\Delta$	0	Δ	Δ	0	$-\Delta$	$-\Delta$	0.017	0.048
		Δ	2Δ	Δ	Δ	$-\Delta$	-2Δ	$-\Delta$	-4Δ	0.014	0.014
		Δ	0	$-\Delta$	Δ	Δ	0	0	Δ	0.011	0.035
		Δ	Δ	0	Δ	0	$-\Delta$	0	$-\Delta$	0.011	0.039
		Δ	2Δ	Δ	Δ	Δ	0	$-\Delta$	-2Δ	0.008	0.014
		0	Δ	Δ	Δ	Δ	0	Δ	0	0.005	0.017
									max	0.042	0.048
0.4	0.1	Δ	2Δ	Δ	2Δ	Δ	$-\Delta$	Δ	$-\Delta$	0.043	0.043
		Δ	0	$-\Delta$	$-\Delta$	Δ	2Δ	-2Δ	Δ	0.029	0.040
		$-\Delta$	0	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.028	0.044
		0	$-\Delta$	$-\Delta$	Δ	0	$-\Delta$	$-\Delta$	$-\Delta$	0.018	0.044
		$-\Delta$	Δ	2Δ	$-\Delta$	Δ	2Δ	$-\Delta$	$-\Delta$	0.014	0.014
		$-\Delta$	0	Δ	Δ	Δ	0	0	$-\Delta$	0.011	0.039
		Δ	0	$-\Delta$	Δ	$-\Delta$	-2Δ	2Δ	Δ	0.007	0.011
		0	Δ	Δ	0	Δ	Δ	$-\Delta$	$-\Delta$	0.005	0.013
		$-\Delta$	$-\Delta$	0	0	$-\Delta$	$-\Delta$	0	$-\Delta$	0.004	0.010
		$-\Delta$	$-\Delta$	0	$-\Delta$	0	Δ	0	Δ	0.004	0.013
									max	0.043	0.044

Table 2.4: Preliminary study when $\alpha_2 = 0.2$

α_1	α_2	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	FDR	FWER
0.2	0.2	$-\Delta$	-2Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	Δ	0.080	0.080
		$-\Delta$	0	Δ	-2Δ	$-\Delta$	Δ	Δ	Δ	0.048	0.084
		2Δ	Δ	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.045	0.087
		$-\Delta$	-2Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	-3Δ	$-\Delta$	0.038	0.038
		Δ	0	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.030	0.084
		Δ	Δ	0	0	$-\Delta$	$-\Delta$	0	$-\Delta$	0.018	0.059
		0	$-\Delta$	$-\Delta$	Δ	$-\Delta$	-2Δ	2Δ	Δ	0.018	0.033
		$-\Delta$	$-\Delta$	0	Δ	0	$-\Delta$	Δ	0	0.017	0.061
		Δ	Δ	0	0	Δ	Δ	0	Δ	0.013	0.039
		0	Δ	Δ	0	Δ	Δ	$-\Delta$	$-\Delta$	0.012	0.029
									max	0.080	0.087
0.3	0.2	Δ	2Δ	Δ	2Δ	Δ	$-\Delta$	Δ	$-\Delta$	0.077	0.077
		Δ	0	$-\Delta$	$-\Delta$	Δ	2Δ	-4Δ	$-\Delta$	0.045	0.076
		Δ	$-\Delta$	-2Δ	Δ	Δ	0	$-\Delta$	Δ	0.043	0.080
		-2Δ	$-\Delta$	Δ	Δ	$-\Delta$	-2Δ	2Δ	$-\Delta$	0.037	0.037
		$-\Delta$	0	Δ	Δ	0	$-\Delta$	Δ	$-\Delta$	0.032	0.086
		$-\Delta$	0	Δ	Δ	Δ	0	Δ	0	0.018	0.061
		0	$-\Delta$	$-\Delta$	Δ	Δ	0	0	Δ	0.017	0.062
		$-\Delta$	Δ	2Δ	0	Δ	Δ	$-\Delta$	-2Δ	0.015	0.027
		0	$-\Delta$	$-\Delta$	$-\Delta$	$-\Delta$	0	Δ	2Δ	0.013	0.032
		$-\Delta$	Δ	2Δ	0	Δ	Δ	Δ	0	0.012	0.032
									max	0.077	0.086
0.4	0.2	$-\Delta$	-2Δ	$-\Delta$	2Δ	Δ	$-\Delta$	$-\Delta$	$-\Delta$	0.087	0.087
		$-\Delta$	0	Δ	2Δ	Δ	$-\Delta$	Δ	$-\Delta$	0.044	0.077
		-2Δ	$-\Delta$	Δ	Δ	0	$-\Delta$	Δ	$-\Delta$	0.041	0.079
		$-\Delta$	-2Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	Δ	3Δ	0.036	0.036
		$-\Delta$	0	Δ	Δ	0	$-\Delta$	Δ	$-\Delta$	0.029	0.082
		2Δ	Δ	$-\Delta$	0	Δ	Δ	-3Δ	$-\Delta$	0.020	0.036
		Δ	Δ	0	0	$-\Delta$	$-\Delta$	0	$-\Delta$	0.018	0.060
		0	$-\Delta$	$-\Delta$	Δ	Δ	0	$-\Delta$	0	0.017	0.061
		Δ	2Δ	Δ	Δ	Δ	0	0	$-\Delta$	0.013	0.032
		$-\Delta$	$-\Delta$	0	0	$-\Delta$	$-\Delta$	Δ	0	0.010	0.030
									max	0.087	0.087

Table 2.5: Preliminary study when $\alpha_2 = 0.3$

α_1	α_2	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	FDR	FWER
0.2	0.3	Δ	2Δ	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.120	0.120
		2Δ	Δ	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.063	0.120
		Δ	$-\Delta$	-2Δ	$-\Delta$	-2Δ	$-\Delta$	-2Δ	$-\Delta$	0.058	0.058
		Δ	0	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.043	0.120
		0	$-\Delta$	$-\Delta$	Δ	$-\Delta$	-2Δ	2Δ	Δ	0.030	0.055
		Δ	-2Δ	-3Δ	Δ	0	$-\Delta$	$-\Delta$	Δ	0.029	0.056
		Δ	Δ	0	0	$-\Delta$	$-\Delta$	0	$-\Delta$	0.026	0.086
		Δ	Δ	0	$-\Delta$	0	Δ	0	Δ	0.025	0.093
		$-\Delta$	-2Δ	$-\Delta$	$-\Delta$	$-\Delta$	0	0	Δ	0.019	0.047
		Δ	Δ	0	Δ	0	$-\Delta$	0	$-\Delta$	0.015	0.055
									max	0.120	0.120
0.3	0.3	Δ	2Δ	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.130	0.130
		$-\Delta$	0	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.062	0.111
		$-\Delta$	Δ	2Δ	2Δ	Δ	$-\Delta$	Δ	-2Δ	0.050	0.050
		0	Δ	Δ	Δ	0	$-\Delta$	Δ	$-\Delta$	0.040	0.108
		0	Δ	Δ	Δ	$-\Delta$	-2Δ	4Δ	Δ	0.039	0.075
		$-\Delta$	$-\Delta$	0	0	$-\Delta$	$-\Delta$	0	$-\Delta$	0.028	0.088
		$-\Delta$	$-\Delta$	0	$-\Delta$	-2Δ	$-\Delta$	2Δ	Δ	0.026	0.047
		Δ	Δ	0	0	Δ	Δ	-2Δ	$-\Delta$	0.023	0.051
		0	Δ	Δ	Δ	Δ	0	$-\Delta$	-2Δ	0.021	0.051
		Δ	Δ	0	0	Δ	Δ	$-\Delta$	0	0.017	0.053
									max	0.130	0.130
0.4	0.3	-2Δ	$-\Delta$	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.119	0.119
		Δ	0	$-\Delta$	$-\Delta$	-2Δ	$-\Delta$	$-\Delta$	$-\Delta$	0.060	0.106
		$-\Delta$	Δ	2Δ	$-\Delta$	3Δ	4Δ	$-\Delta$	Δ	0.053	0.053
		$-\Delta$	0	Δ	0	$-\Delta$	$-\Delta$	Δ	$-\Delta$	0.043	0.109
		$-\Delta$	0	Δ	Δ	0	$-\Delta$	Δ	$-\Delta$	0.039	0.110
		0	$-\Delta$	$-\Delta$	2Δ	$-\Delta$	-3Δ	Δ	$-\Delta$	0.028	0.051
		0	$-\Delta$	$-\Delta$	Δ	Δ	0	$-\Delta$	0	0.027	0.097
		$-\Delta$	$-\Delta$	0	0	$-\Delta$	$-\Delta$	$-\Delta$	-2Δ	0.023	0.052
		Δ	Δ	0	2Δ	Δ	$-\Delta$	Δ	0	0.019	0.055
		0	Δ	Δ	Δ	Δ	0	0	$-\Delta$	0.017	0.052
									max	0.119	0.119

Table 2.6: Control FDR/FWER when fix $\alpha_1 = 0.2$

α_1	α_2	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	FDR	FWER
0.20	0.11	-2Δ	$-\Delta$	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.044	0.044
		$-\Delta$	0	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.025	0.040
		$-\Delta$	-2Δ	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.022	0.041
		Δ	2Δ	Δ	2Δ	Δ	$-\Delta$	$-\Delta$	-3Δ	0.021	0.021
		$-\Delta$	0	Δ	Δ	0	$-\Delta$	Δ	$-\Delta$	0.016	0.043
		0	Δ	Δ	Δ	$-\Delta$	-2Δ	4Δ	Δ	0.013	0.026
		$-\Delta$	Δ	2Δ	0	Δ	Δ	2Δ	Δ	0.013	0.022
		0	$-\Delta$	$-\Delta$	Δ	Δ	0	0	Δ	0.010	0.036
		$-\Delta$	$-\Delta$	0	0	$-\Delta$	$-\Delta$	2Δ	Δ	0.008	0.020
		Δ	Δ	0	0	Δ	Δ	$-\Delta$	0	0.007	0.022
									max	0.044	0.044
0.20	0.13	Δ	2Δ	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.052	0.052
		-2Δ	$-\Delta$	Δ	2Δ	Δ	$-\Delta$	$-\Delta$	-3Δ	0.027	0.027
		Δ	0	$-\Delta$	2Δ	Δ	$-\Delta$	Δ	Δ	0.027	0.043
		0	Δ	Δ	$-\Delta$	-2Δ	$-\Delta$	Δ	$-\Delta$	0.024	0.046
		$-\Delta$	0	Δ	0	$-\Delta$	$-\Delta$	Δ	$-\Delta$	0.019	0.049
		0	Δ	Δ	Δ	$-\Delta$	-2Δ	4Δ	Δ	0.019	0.036
		$-\Delta$	0	Δ	$-\Delta$	Δ	2Δ	Δ	2Δ	0.018	0.027
		2Δ	Δ	$-\Delta$	0	Δ	Δ	-3Δ	$-\Delta$	0.015	0.027
		Δ	0	$-\Delta$	Δ	Δ	0	0	Δ	0.012	0.041
		0	Δ	Δ	$-\Delta$	$-\Delta$	0	0	$-\Delta$	0.011	0.041
		Δ	Δ	0	0	Δ	Δ	Δ	2Δ	0.010	0.025
		0	$-\Delta$	$-\Delta$	Δ	$-\Delta$	-2Δ	0	$-\Delta$	0.010	0.027
		Δ	Δ	0	0	Δ	Δ	$-\Delta$	0	0.008	0.024
									max	0.052	0.052
0.20	0.15	$-\Delta$	-2Δ	$-\Delta$	Δ	2Δ	Δ	$-\Delta$	Δ	0.053	0.053
		-2Δ	$-\Delta$	Δ	2Δ	Δ	$-\Delta$	3Δ	Δ	0.031	0.031
		Δ	0	$-\Delta$	Δ	2Δ	Δ	$-\Delta$	Δ	0.031	0.052
		2Δ	Δ	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.028	0.055
		0	Δ	Δ	Δ	$-\Delta$	-2Δ	4Δ	Δ	0.023	0.044
		Δ	0	$-\Delta$	$-\Delta$	0	Δ	$-\Delta$	Δ	0.022	0.060
		0	$-\Delta$	$-\Delta$	Δ	$-\Delta$	-2Δ	2Δ	Δ	0.016	0.029
		Δ	0	$-\Delta$	Δ	Δ	0	0	Δ	0.013	0.045
		0	$-\Delta$	$-\Delta$	Δ	Δ	0	$-\Delta$	0	0.013	0.046
		Δ	Δ	0	0	Δ	Δ	-2Δ	$-\Delta$	0.011	0.028
		$-\Delta$	$-\Delta$	0	0	$-\Delta$	$-\Delta$	Δ	0	0.009	0.028
									max	0.053	0.060

To control FDR and FWER by adjusting α_1 in stage 1 and α_2 in stage 2, a series of preliminary simulations with 1000 sample observations were generated from a multivariate normal distribution. Here the sample covariance from actual data is used as covariance matrix. Different combinations of means from null and alternative are selected to observe the effects of changing α_1 and α_2.

From Table 2.3 - Table 2.5, it indicates that for fixed α_2, even as α_1 increased dramatically, the change of FDR or FWER was quite insignificant. Therefore, to control FDR at a level of 5%, with fixed $\alpha_1 = 0.2$, further simulations were conducted while α_2 varied.

Table 2.6 shows simulated FDR and FWER under different combinations of null and alternative hypotheses. To increase accuracy, 5000 observations were generated under each combination.

Table 2.6 indicates that when $\alpha_1 = 0.2$ and $\alpha_2 = 0.13$, using

$$C_k = \sqrt{\chi^2_{1;\alpha_{1k}}(((2\Delta)^2)}, \quad \alpha_{1k} = \frac{k\alpha_1}{8 - k(1-\alpha_1)} \tag{2.4.2}$$

And

$$B(r_i) = t_{n-1;1-\frac{\alpha_2/2}{8-r_i}} \tag{2.4.3}$$

FDR can be controlled at 5% level.

Apply the above results to actual data to answer questions in section 1. In stage 1, the observed $U = (U_1, U_2, \cdots, U_8)'$ is

$$U = \begin{pmatrix} * & 2.153 & 2.153 & 4.158 & 1.708 & 2.408 & 2.322 & 2.481 \\ 6.387 & * & 6.387 & 7.520 & 3.467 & 6.459 & 4.683 & 4.477 \\ 6.962 & 6.962 & * & 7.984 & 6.465 & 1.976 & 5.327 & 5.683 \\ 1.009 & 2.030 & 1.907 & * & 2.078 & 2.078 & 2.439 & 2.263 \\ 6.699 & 4.221 & 6.318 & 7.909 & * & 7.909 & 5.082 & 5.053 \\ 6.981 & 6.964 & 1.733 & 7.970 & 7.970 & * & 5.412 & 5.440 \\ 4.678 & 1.502 & 1.013 & 6.221 & 0.589 & 1.683 & * & 2.930 \\ 4.773 & 0.715 & 2.254 & 6.166 & 0.437 & 1.810 & 2.953 & * \end{pmatrix}$$

Then order U_i's from smallest to largest and denote as

$$V = \begin{pmatrix} 1.708 & 2.153 & 2.153 & 2.322 & 2.408 & 2.481 & 4.158 \\ 3.467 & 4.477 & 4.683 & 6.387 & 6.387 & 6.459 & 7.520 \\ 1.976 & 5.327 & 5.683 & 6.465 & 6.962 & 6.962 & 7.984 \\ 1.009 & 1.907 & 2.030 & 2.078 & 2.078 & 2.263 & 2.439 \\ 4.221 & 5.053 & 5.082 & 6.318 & 6.699 & 7.909 & 7.909 \\ 1.733 & 5.412 & 5.440 & 6.964 & 6.981 & 7.970 & 7.970 \\ 0.589 & 1.013 & 1.502 & 1.683 & 2.930 & 4.678 & 6.221 \\ 0.437 & 0.715 & 1.810 & 2.254 & 2.953 & 4.773 & 6.166 \end{pmatrix}$$

When $\alpha_1 = 0.2$, using formula of (2.4.2) where $k = 7$, the vector of constants is calculated as $C = (0.038, 0.087, 0.149, 0.232, 0.352, 0.539, 0.895)'$. Since $V_{i1} > C_1$ for all $i = 1, 2, \cdots, 8$, there is no rejection for any H_i, therefore, $R_i = 0$ for all i's.

In stage 2, When $\alpha_2 = 0.13$, using formula of (2.4.3), $B(r_i) = B(0) = 2.518$ for all i's. The confidence interval estimates for μ_i's is calculated using (2.4.1) as follows:

(1) (-0.2239, -0.1014)

(2) (-0.2446, -0.0181)

(3) (-0.0537, 0.1162)

(4) (-0.1367, -0.0700)

(5) (-0.1516, 0.0314)

(6) (-0.0389, 0.1254)

(7) (0.0168, 0.0470)

(8) (0.0230, 0.0648)

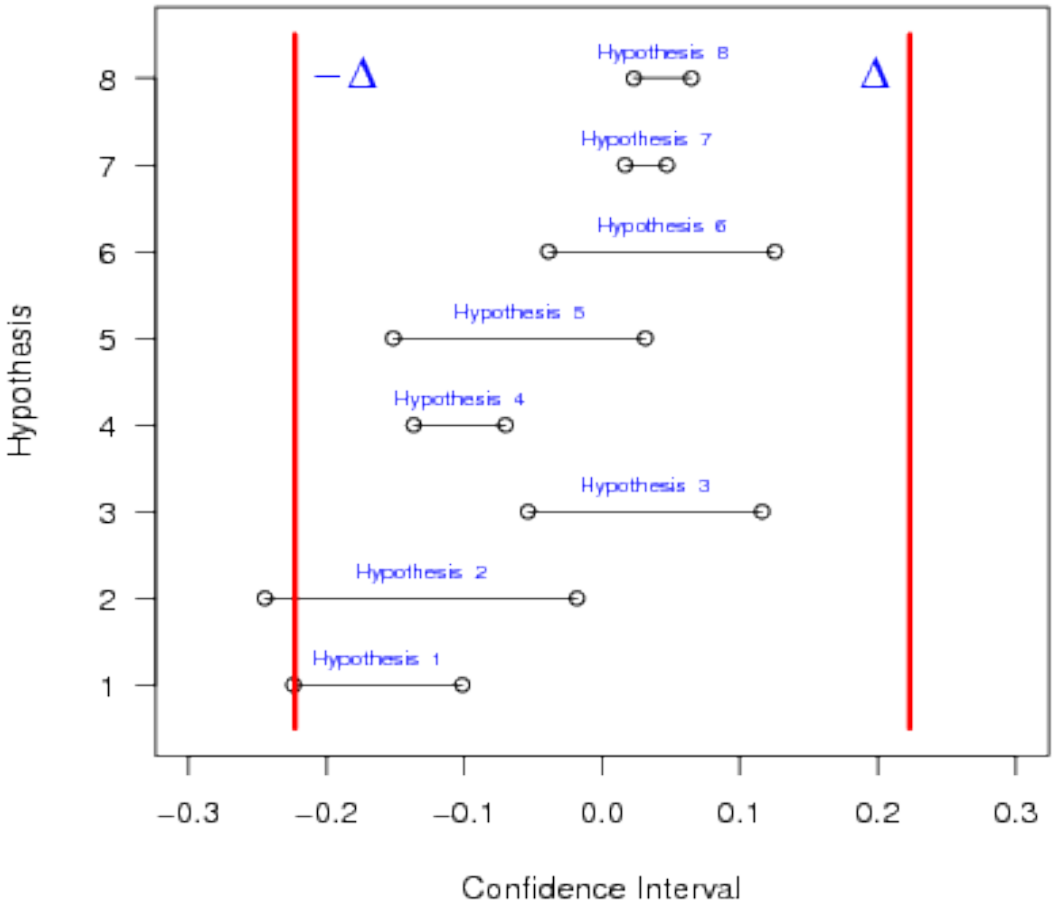


Figure 2.3: 2-stage multiple testing confidence interval estimate

From Figure 2.3, it is clear that confidence interval estimates for $\mu_i, i = 3, 4, \cdots, 8$, falls into $(-\Delta, \Delta)$ where $\Delta = ln1.25 \simeq 0.223$. Therefore null hypothesis $H_i, i = 3, 4, \cdots, 8$, was rejected. However, since the confidence interval estimates for $\mu_i, i = 1, 2$, was not completely included in $(-\Delta, \Delta)$, null hypothesis $H_i, i = 1, 2$, was accepted.

Chapter 3

Appendix

3.1 Derivation of (1.2.4) and (1.2.5)

3.1.1 Proof of (1.2.4)

Since $\sqrt{n_1}\bar{X}_1 \sim \mathcal{N}(\sqrt{n_1}\theta, 1)$,

$$\begin{aligned} P_\theta(R) &= P_\theta(\sqrt{n_1}|\bar{X}_1| < C_{\alpha_1;\sqrt{n_1}\epsilon}) \\ &= \Phi(C_{\alpha_1;\sqrt{n_1}\epsilon} - \sqrt{n_1}\theta) - \Phi(-C_{\alpha_1;\sqrt{n_1}\epsilon} - \sqrt{n_1}\theta) \\ &= \Phi(C_{\alpha_1;\sqrt{n_1}\epsilon} - \sqrt{n_1}\theta) + \Phi(C_{\alpha_1;\sqrt{n_1}\epsilon} + \sqrt{n_1}\theta) - 1 \end{aligned}$$

3.1.2 Proof of (1.2.5)

Notice that

$$P(R|\bar{x}) = \frac{\int_R f(\bar{x}_1, \bar{x}) d\bar{x}_1}{\int f(\bar{x}_1, \bar{x}) d\bar{x}_1}$$

Since

$$\bar{X}_1 \sim \mathcal{N}\left(\theta, \frac{1}{n_1}\right) \quad \text{and} \quad \bar{X}_2 \sim \mathcal{N}\left(\theta, \frac{1}{n_2}\right),$$

also $\bar{X}_1$ and $\bar{X}_2$ are independent,

$$f(\bar{x}_1, \bar{x}_2) = \frac{n_1 n_2}{2\pi} \exp(-\frac{n_1}{2}(\bar{x}_1 - \theta)^2 - \frac{n_2}{2}(\bar{x}_2 - \theta)^2).$$

Let

$$\bar{x}_2 = \frac{n\bar{x} - n_1\bar{x}_1}{n_2}, \quad |J| = \frac{n}{n_2},$$

then

$$\begin{aligned}\Longrightarrow f(\bar{x}_1,\bar{x}) &= \frac{n_1 n}{2\pi}\exp(-\frac{n_1}{2}(\bar{x}_1-\theta)^2-\frac{n_2}{2}(\frac{n\bar{x}-n_1\bar{x}_1}{n_2}-\theta)^2)\\ &= \frac{n_1 n}{2\pi}\exp(-\frac{n_1+n_2}{2}\theta^2)\exp(n\bar{x}\theta)\exp(-\frac{n}{2}\bar{x}^2)\exp(-\frac{nn_1}{2n_2}(\bar{x}_1-\bar{x})^2)\end{aligned}$$

$$\Longrightarrow P(R|\bar{x}) = \frac{\int_R \exp(-\frac{nn_1}{2n_2}(\bar{x}_1-\bar{x})^2)d\bar{x}_1}{\int \exp(-\frac{nn_1}{2n_2}(\bar{x}_1-\bar{x})^2)d\bar{x}_1} \tag{3.1.1}$$

From (3.1.1), for each fixed $\bar{X}=\bar{x}$, $\bar{X}_1 \sim \mathcal{N}(\bar{x}, \frac{n_2}{nn_1})$, thus

$$P(R|\bar{x}) = \Phi\left(\frac{C_{\alpha_1;\sqrt{n_1}\epsilon}-\sqrt{n_1}\bar{x}}{\sqrt{n_2/n}}\right)+\Phi\left(\frac{C_{\alpha_1;\sqrt{n_1}\epsilon}+\sqrt{n_1}\bar{x}}{\sqrt{n_2/n}}\right)-1$$

3.2 Two versions of hypotheses when comparing two binomial distributions

Let

$$\gamma_1 = \log(\text{odds ratio}) = \log \frac{p_1(1-p_2)}{p_2(1-p_1)}$$

Consider the following hypothesis

$$H^{'} : |\gamma_1| \geq \gamma_0 \quad \text{vs} \quad K^{'} : |\gamma_1| < \gamma_0$$

Let $\beta_1 = e^{-\gamma_0}$, $\beta_2 = e^{\gamma_0}$ ($\gamma_0 > 0$), then $0 < \beta_1 < 1$, $\beta_2 > 1$. And

$$\begin{aligned}
K^{'} : \beta_1 < &\ \tfrac{p_1(1-p_2)}{p_2(1-p_1)} &< \beta_2 \\
\Longrightarrow \beta_1(\frac{p_2}{1-p_2}) < &\ \tfrac{p_1}{1-p_1} &< \beta_2(\frac{p_2}{1-p_2}) \\
\Longrightarrow \frac{\beta_1 p_2}{1-p_2+\beta_1 p_2} < &\ p_1 &< \frac{\beta_2 p_2}{1-p_2+\beta_2 p_2} \\
\Longrightarrow g(\beta_1, p_2) < &\ p_1 - p_2 &< g(\beta_2, p_2)
\end{aligned}$$

where

$$g(\beta, p_2) = \frac{p_2(\beta-1)(1-p_2)}{\beta + (1-\beta)(1-p_2)}$$

Notice that the first derivative of $g(\beta, p_2)$ when $0 < p_2 < 1$

$$g(\beta, p_2)\prime = \frac{dg(\beta, p_2)}{dp_2} = -1 + \frac{\beta}{[1+(\beta-1)p_2]^2}$$

is a strictly increasing function when $0 < \beta_1 < 1$ and is a strictly decreasing function when $\beta_2 > 1$. In addition, since $g(\beta_1, 0)\prime = \beta_1 - 1 < 0$, $g(\beta_2, 0)\prime = \beta_2 - 1 > 0$ and $g(\beta_1, \frac{1}{\sqrt{\beta_1}+1})\prime = g(\beta_2, \frac{1}{\sqrt{\beta_2}+1})\prime = 0$, it is clear that $g(\beta_1, p_2)$ is a decreasing then increasing function with a minimum at $p_2 = \frac{1}{\sqrt{\beta_1}+1}$ and $g(\beta_2, p_2)$ is an increasing then decreasing function with a maximum $p_2 = \frac{1}{\sqrt{\beta_2}+1}$. And

$$\min_{p_2} g(\beta_1, p_2) = \frac{\sqrt{\beta_1}-1}{\sqrt{\beta_1}+1} \qquad \max_{p_2} g(\beta_2, p_2) = \frac{\sqrt{\beta_2}-1}{\sqrt{\beta_2}+1}$$

Since

$$-\frac{\sqrt{\beta_1}-1}{\sqrt{\beta_1}+1}=-\frac{e^{-\gamma_0/2}-1}{e^{-\gamma_0/2}+1}=\frac{e^{\gamma_0/2}-1}{e^{\gamma_0/2}+1}=\frac{\sqrt{\beta_2}-1}{\sqrt{\beta_2}+1}>0$$

let

$$\epsilon=\frac{\sqrt{\beta_2}-1}{\sqrt{\beta_2}+1}=\frac{e^{\gamma_0/2}-1}{e^{\gamma_0/2}+1}$$

then

$$H:|p_1-p_2|\geq\epsilon \quad \text{vs} \quad K:|p_1-p_2|<\epsilon \tag{3.2.1}$$

is equivalent to

$$H^{'}:|\gamma_1|\geq\gamma_0 \quad \text{vs} \quad K^{'}:|\gamma_1|<\gamma_0 \tag{3.2.2}$$

From the above derivation, we can see that for every given ϵ, if let $\gamma_0=2\log(\frac{1+\epsilon}{1-\epsilon})$, then the alternative space of the hypothesis (3.2.2) would be contained in the alternative space of the hypothesis (3.2.1). Therefore when we formulate a size α test based on the hypothesis (3.2.2) with $\gamma_0=2\log(\frac{1+\epsilon}{1-\epsilon})$, it is also a level α test for the hypothesis (3.2.1).

3.3 Data for Multiple Testing on Bio-equivalence

3.3.1 Iron Evaluation for Blood Serum

Table 3.1: Iron level after log-transformation

RXL1			RXL2		
Current reagent	New reagent	New reagent	Current reagent	New reagent	New reagent
Serum Iron	Serum Iron	Heparin Iron	Serum Iron	Serum Iron	Heparin Iron
3.807	3.912	3.850	3.850	3.912	3.784
4.419	4.477	4.454	4.443	4.489	4.454
3.761	3.871	3.892	3.807	3.871	3.850
5.226	5.247	5.182	5.236	5.257	5.182
4.700	4.727	4.754	4.710	4.736	4.762
3.807	3.932	3.714	3.850	3.912	3.714
1.792	2.398	2.485	2.079	2.398	2.398
2.639	2.890	2.944	2.708	2.773	2.833
3.555	3.664	3.714	3.611	3.664	3.738
3.638	3.761	3.784	3.664	3.738	3.761
3.951	4.043	4.025	3.970	4.025	3.989
2.833	3.045	3.091	2.944	3.045	2.996
4.489	4.533	4.663	4.500	4.533	4.663
2.197	2.639	2.639	2.398	2.565	2.565
2.890	3.178	3.258	2.944	3.091	3.178
3.784	3.850	3.850	3.807	3.850	3.829
3.401	3.526	3.555	3.466	3.526	3.526
3.555	3.738	3.584	3.555	3.664	3.555
1.946	2.485	2.485	2.079	2.398	2.485
5.652	5.576	5.313	5.624	5.631	5.313
3.401	3.555	3.045	3.367	3.466	2.890
3.714	3.807	3.807	3.689	3.761	3.784
4.190	4.248	4.605	4.159	4.205	4.595
4.905	4.934	4.564	4.852	4.913	4.564
3.784	3.871	3.892	3.738	3.850	3.871
4.466	4.489	4.369	4.407	4.443	4.344
2.079	2.565	2.398	2.197	2.485	2.303
3.434	3.584	3.555	3.434	3.555	3.526
3.664	3.761	3.761	3.611	3.738	3.738
4.205	4.248	3.689	4.159	4.220	3.664
3.091	3.258	3.434	3.045	3.219	3.367
3.951	3.989	3.555	3.912	3.951	3.611
2.485	2.833	2.708	2.485	2.773	2.565
3.401	3.555	3.611	3.434	3.497	3.526
2.996	3.178	3.714	2.996	3.135	3.555
3.434	3.611	3.807	3.434	3.526	3.689
2.639	2.890	2.890	2.708	2.890	2.890
2.944	3.135	3.178	2.996	3.091	3.091

3.3.2 Normality test

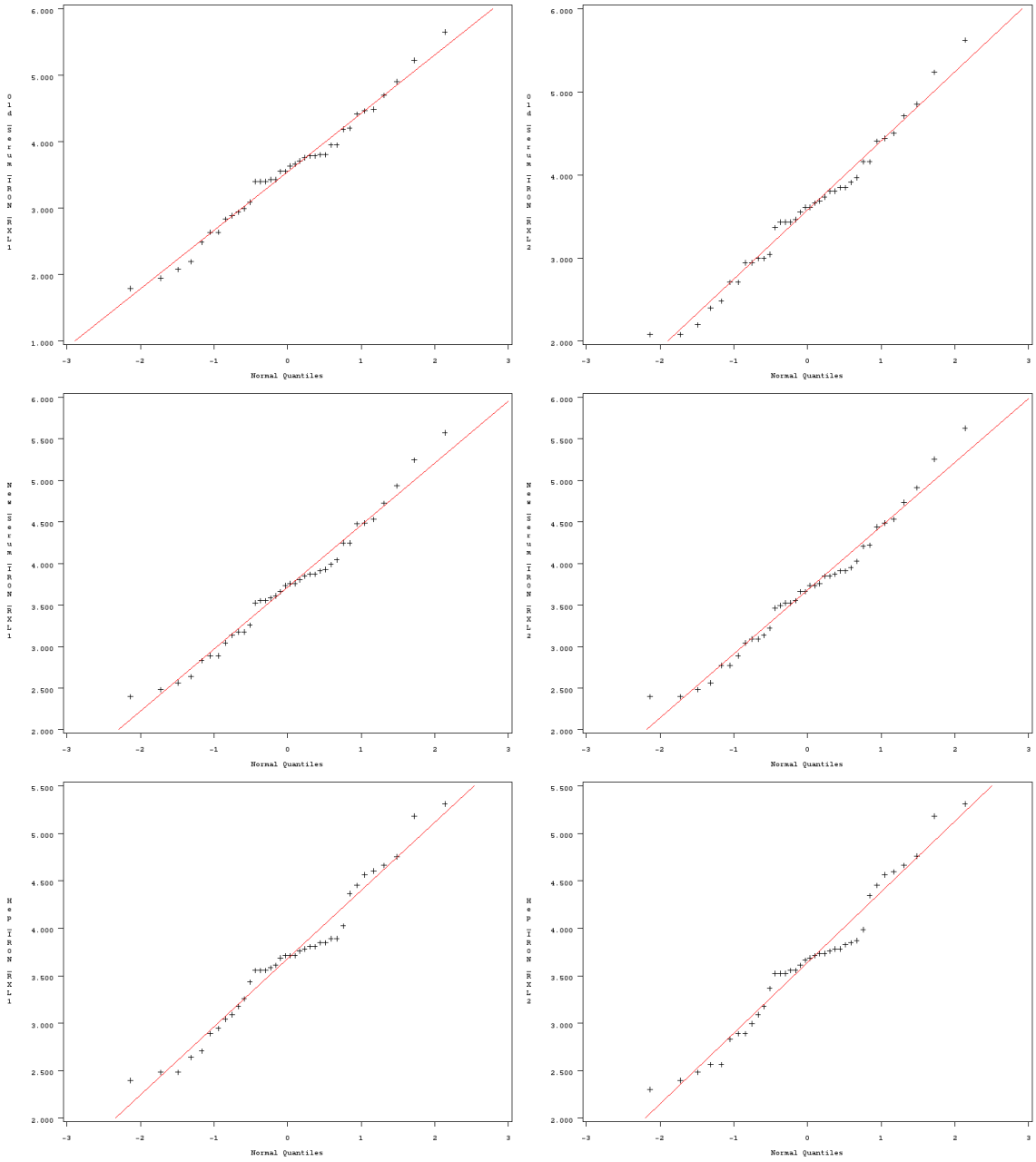
Table 3.2: Marginal Normality Test for Iron measurements

Old Serum Iron: RXL1				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.984205	Pr < W	0.8579
Kolmogorov-Smirnov	D	0.118033	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.050172	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.264051	Pr > A-Sq	>0.2500
New Serum Iron: RXL1				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.97583	Pr < W	0.5708
Kolmogorov-Smirnov	D	0.094108	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.05113	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.300418	Pr > A-Sq	>0.2500
New Serum Heparin Iron: RXL1				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.965021	Pr < W	0.2752
Kolmogorov-Smirnov	D	0.14684	Pr > D	0.0379
Cramer-von Mises	W-Sq	0.10164	Pr > W-Sq	0.1039
Anderson-Darling	A-Sq	0.527639	Pr > A-Sq	0.1743
Old Serum Iron: RXL2				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.979571	Pr < W	0.7023
Kolmogorov-Smirnov	D	0.090567	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.044517	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.260062	Pr > A-Sq	>0.2500
New Serum Iron: RXL2				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.97355	Pr < W	0.4957
Kolmogorov-Smirnov	D	0.098297	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.046148	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.289917	Pr > A-Sq	>0.2500
New Serum Heparin Iron: RXL2				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.963794	Pr < W	0.2516
Kolmogorov-Smirnov	D	0.139093	Pr > D	0.0630
Cramer-von Mises	W-Sq	0.10671	Pr > W-Sq	0.0906
Anderson-Darling	A-Sq	0.556123	Pr > A-Sq	0.1450

Table 3.3: Multivariate Normality Test for Iron measurements

Test Statistic	Value	Prob
Mardia Skewness	114.1	<.0001
Mardia Kurtosis	2.67	0.0076
Henze-Zirkler T	7.05	<.0001

Figure 3.1: Q-Q plot: RXL1 vs RXL2



References

[1] ANDERSON, S. and HAUCK, W. W. (1990). Considerations of individual bioequivalence. *J. Pharmacokinetics and Biopharmaceutics*, **8**, 259–273.

[2] BENJAMINI, Y. and GAVRILOV, Y. (2009). A simple forward selection procedure based on false discovery rate control. *Ann. Appl. Stat.*, **3**, 179–198.

[3] BERGER, R. L. and HSU, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statist. Sci.*, **11**, 283–302.

[4] BROWN, L. D., HWANG, J. T. G. and MUNK, A. (1997). An unbiased test for the bioequivalence problem. *Ann. Statist.*, **25**, 2345–2367.

[5] CHOW, S. C., SHAO, J. and WANG, H. (2003). Statistical tests for population bioequivalence. *Statistica Sinica*, **13**, 539–554.

[6] COHEN A., MA, Y. and SACKROWITZ H. B. (2012). Individualized two-stage multiple testing procedures with corresponding interval estimates. *Biometrical Journal*

[7] COHEN A. and SACKROWITZ H. B. (1996). Lower confidence bounds using pilot samples with an application to auditing. *J. Amer. Statist. Assoc.*, **91**, 338–342.

[8] FDA (2001). *Guidance for Industry: Statistical Approaches to Establishing Bioequivalence*. Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.

[9] HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, **6**, 65–70.

[10] LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*. Springer, New York.

[11] LIU, J. P. and CHOW, S. C. (1992). On assessment of variability in bioavailability=bioequivalence studies. *Commun. Statist.- Theory Meth.*, **21**, 2591–2607.

[12] SCHALL, R. (1995). Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. *Biometrics*, **51**, 615–626.

[13] SCHALL, R. and LUUS, H. G. (1993) On population and individual bioequivalence. *Statist. Medicine*, **12**, 1109–1124.

[14] SCHUIRMANN, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinetics and Biopharmaceutics*, **15**, 657–680.

[15] SHEINER, L. B. (1992). Bioequivalence revisited. *Statist. Medicine*, **11**, 1777–1788.

[16] WELLEK, S. (2000). On a reasonable disaggregate criterion of population bioequivalence admitting of resampling-free testing procedures. *Statist. Medicine*, **19**, 2755–2767.

[17] WELLEK, S. (2002). *Testing Statistical Hypotheses of Equivalence*. Chapman and Hall/CRC.