CHAPTER 2
REVIEW OF LITERATURE

This chapter is comprised of three parts: (1) the related theory (2) review of the related literature and (3) examples of the procedure.

2.1 The Related Theory

2.1.1 The One-way ANOVA Model

Let $\mu_0, \mu_1, ..., \mu_k$ be the k+1 treatment means. Assume a random sample of size r_{i*} $Y_{i*1}, Y_{i*2}, ..., Y_{i*r_{i*}}$ is taken, for the $i^{*\text{th}}$ treatment $\left(i* = 0,1,...,k\right)$, and the random samples are independent between treatments. Then, under the usual normality and equality of variance assumptions, the one-way ANOVA model is

$$Y_{i*j} = \mu_{i*} + \varepsilon_{i*j}; i* = 0,1,...,k,\ j = 1,2,...,r_{i*}$$

where $\varepsilon_{01}, ..., \varepsilon_{kr_k}$ are independent and identically normally distributed with mean 0 and variance σ^2 unknown. The estimators of μ_{i*} and σ^2 are defined as the following;

$$n = \sum_{i*=0}^{k} r_{i*}$$

$$\hat{\mu} = \bar{Y} = \sum_{i*=0}^{k}\sum_{j=1}^{r_{i*}} Y_{i*j} \Big/ \sum_{i*=0}^{k} r_{i*}$$

$$\hat{\mu}_{i*} = \bar{Y}_{i*} = \sum_{j=1}^{r_{i*}} Y_{i*j} / r_{i*}$$

$$\hat{\sigma}^2 = MSE = \sum_{i*=0}^{k}\sum_{j=1}^{r_{i*}} \left(Y_{i*j} - \bar{Y}_{i*}\right)^2 \Big/ \sum_{i*=0}^{k} \left(r_{i*} - 1\right)$$

where $\bar{Y}$ is the sample mean,

$\bar{Y}_{i*}$ is the sample mean for the $i^{*\text{th}}$ treatment,

$\hat{\sigma}^2$ is the pooled sample variance.

The sample means $\hat{\mu}_0,...,\hat{\mu}_k$ are independent identical normally distributed random variables with means $\mu_0,...,\mu_k$ and variances $\frac{\sigma^2}{r_0},...,\frac{\sigma^2}{r_k}$ and $\frac{\nu\hat{\sigma}^2}{\sigma^2}$ has a chi-square distribution with ν degrees of freedom with $\nu = \sum_{i*=0}^{k}\left(r_{i*}-1\right) = n-\left(k+1\right)$ (Hsu, 1996).

The null hypothesis test for differences among treatment means using the sample data is $H_0 : \mu_0 = \mu_1 = ... = \mu_k$ (overall null hypothesis) and the alternative hypothesis is H_a : not all treatment means are equal. The test statistic in this hypothesis is an F statistic, so that we reject $H_0 : \mu_0 = \mu_1 = ... = \mu_k$ at significance level α if

$$F^* = \frac{MSR}{MSE} = \frac{\sum_{i*=0}^{k}\sum_{j=1}^{r_{i*}}\left(\bar{Y}_{i*}-\bar{Y}\right)^2 \Big/ k}{\sum_{i*=0}^{k}\sum_{j=1}^{r_{i*}}\left(Y_{i*j}-\bar{Y}_{i*}\right)^2 \Big/ \sum_{i*=0}^{k}\left(r_{i*}-1\right)} \geq F^{(\alpha)}_{k,\sum_{i*=0}^{k}(r_{i*}-1)} .$$

When the null hypothesis is rejected, the inference made is that some differences exist among the treatment means. Multiple comparison methods (MCMs) are designed to investigate differences between specific pairs of treatment means (Rafter el al, 2002).

2.1.2 Multiple Comparisons Inference

Multiple comparisons methods (MCMs) can be classified by θ, a generic parametric of primary interest, and by p, the number of parameters of primary interest. A summary of MCMs is listed in Table 1 (Hsu, 1996).

Table 1: Parameter and Number of Parameters for Multiple Comparisons

MCMs	Parametric of primary interest	Number of parameters of primary interest
ACC *All-Contrast Comparisons*	$\theta = \sum_{i^*=0}^{k} c_{i^*}\mu_{i^*}$ with $\sum_{i^*=0}^{k} c_{i^*} = 0$ for all contrast	Infinite
MCA *All-Pairwise Comparisons*	$\theta = \mu_{i^*} - \mu_j$ for all $i^* < j$	$p = k\left(k+1\right)/2$
MCB *Multiple Comparisons with the Best*	$\theta = \mu_{i^*} - \max_{j \neq i^*} \mu_j$ for $i^* = 0,1,\ldots,k$	$p = k+1$
MCC *Multiple Comparisons with a Control*	$\theta = \mu_i - \mu_0$ with treatment 0 is the control for $i = 1,\ldots,k$	$p = k$

Reason for considering specific types of multiple comparisons (MCB: Multiple Comparisons with the Best and MCC: Multiple Comparisons with a Control) are (1) directed inference is sharper than deduced inference (Scheffé's method: confidence intervals for many linear combinations of $\mu_0,\ldots,\mu_k$ are interest) and (2) the result of a large random of simultaneous comparisons is difficult to comprehend and interpret (Tukey 1992 referred in Hsu, 1996).

Hypothesis of multiple comparisons may be classified into five categories: confidence intervals methods, confidence directions methods, confidence inequalities methods, tests of homogeneity and individual comparisons methods. A confidence level $1-\alpha$ asserts that (1) For confidence interval methods, $\theta_{i**} \in I_{i**}$ for $i** = 1,..., p$, (2) For confidence directions method, for each $i**$, $\theta_{i**} > 0$ or $\theta_{i**} < 0$ and (3) For confidence inequalities method, for each $i**$, $\theta_{i**} \neq 0$, while guaranteeing the simultaneous coverage probability of the intervals for each inference is at least $100(1-\alpha)\%$.Tests of homogeneity at level α asserts $\theta_{i**} \neq 0$ for some $i**$ if data so warrants, which corresponds to the rejection of the null hypothesis of homogeneity of means $H_0 : \mu_0 = ... = \mu_k$. Finally, the individual comparison method with confidence level $1-\alpha$ guarantees the simultaneous accuracy for each assertion is correct with a probability at least $1-\alpha$. Some test statistics to a test hypothesis for each multiple comparison procedure are shown in Table 2 (Hsu, 1996);

Table 2: Test for Multiple Comparisons

	MCA	MCB	MCC
Confidence intervals	(1) Tukey (2) Bofinger (3) Hayter (4) Steel / Dwass	(1) Hsu (2) Edwards - Hsu	(1) Dunnett (2) Bofinger / Steafansson – Kim – Hsu (3) Steel
Confidence directions			(1) Naik – Marcus – Peritz – Gabriel (2) Dunett – Tamhane (one-sided)
Confidence inequalities	(1) Ryan (2) Einot – Gabriel (3) Welch (4) Peritz (5) Finner / Royen		(1) Dunett – Tamhane (two-sided)
Test of homogeneity	(1) Newman – Keuls (2) Protected LSD (3) Dunn		
Individual comparisons	(1) Duncan (2) Unprotected LSD		

2.1.3 Multiple Comparisons with a Control

Multiple comparisons with a control are the comparison of *k* treatments with a control. For example, a control may be a placebo, or standard treatment or any other specified treatment. Suppose r_i , $i = 1,..,k$ is the number of observations for the i^{th} treatment, r_0 is number of observations for a control, μ_0 is the treatment mean of the control and $\mu_1,...,\mu_k$ are the treatment means of the other treatments. Then, in multiple comparisons with a control (MCC), the parameters of primary interest are $\mu_i - \mu_0$ for $i = 1,...,k$. Multiple comparison procedures with a control can be classified by a balanced or an unbalanced one-way ANOVA, and are summarized in Table 3.

Table 3: Multiple Comparisons with a Control (Hsu, 1996)

MCC	Balanced one-way ANOVA*	Unbalanced one-way ANOVA
One-sided	(1) Dunnett (2) Naik, Marcus, Peritz, Gabriel (3) Dunnett and Tamhane	(1) Dunnett (2)Miller - Winer (3) Based on probabilistic inequalities
Two-sided	(1) Dunnett	(1) Dunnett (2) Miller - Winer (3) Based on probabilistic inequalities

***In the balanced one–way ANOVA, each treatment group has equal sample size. That is, $r_0 = r_1 = ... = r_k = r$ and $Y_{i*j} = \mu_{i*} + \varepsilon_{i*j}$; $i* = 0,1,2,...,k,\ j = 1,2,...,r$.**

Dunnett (1955) proposed a method to compare of *k* treatments with a control. Suppose Y_{i^*j}; $i* = 0,1,...,k,\ \ j = 1,...,r_{i*}$ are independent and normally distributed with common variance σ^2 and means μ_{i*}. The estimator of μ_{i*} and σ^2 are defined as

$$\hat{\mu}_{i*} = \bar{Y}_{i*} = \sum_{j=1}^{r_{i*}} Y_{i*j} \Big/ r_{i*}$$

$$\hat{\sigma}^2 = s^2 = MSE = \sum_{i*=0}^{k} \sum_{j=1}^{r_{i*}} \left(Y_{i*j} - \bar{Y}_{i*}\right)^2 \Big/ \sum_{i*=0}^{k} \left(r_{i*} - 1\right).$$

The probability that all *k* confidence intervals will contain the corresponding $\mu_i - \mu_0$, is equal to $P\left(0 < P < 1\right)$, where P is the joint confidence for the *k* treatment effects $\mu_i - \mu_0$. In general, a comparison of the $i^{*\text{th}}$ treatment with a control can be made using $t_i = \dfrac{z_i}{s};\ \ i = 1,...,k$,

where $$z_i = \frac{\bar{Y}_i - \bar{Y}_o - (\mu_i - \mu_0)}{\sqrt{\dfrac{1}{r_i} + \dfrac{1}{r_0}}}.$$

The statistic t_i have a multivariate Student *t* distribution with $\rho_{ij} = 1\Big/\left(\dfrac{r_0}{r_1}+1\right)$ defined by Dunnett. Then lower confidence limits with will be given by

$$\bar{Y}_i - \bar{Y}_o - d_i^{'} s\sqrt{\frac{1}{r_i}+\frac{1}{r_0}};\ i = 1,\ldots,k\ .$$

Similarly, upper confidence limits will be given by

$$\bar{Y}_i - \bar{Y}_o + d_i^{'} s\sqrt{\frac{1}{r_i}+\frac{1}{r_0}},$$

if the *k* constants $d_i^{'}$ are chosen to satisfy

$$\Pr\left(t_1 < d_1^{'}, t_2 < d_2^{'}, \ldots, t_k < d_k^{'}\right) = P. \tag{1}$$

On the other hand, two-sided confidence limits have the desired joint confidence coefficient and will be given by

$$\bar{Y}_i - \bar{Y}_o \pm d_i^{''} s\sqrt{\frac{1}{r_i}+\frac{1}{r_0}};\ i = 1,\ldots,k,$$

if the *k* constants $d_i^{''}$ are chosen so that

$$\Pr\left(\left|t_1\right| < d_1^{'}, \left|t_2\right| < d_2^{'}, \ldots, \left|t_k\right| < d_k^{'}\right) = P. \tag{2}$$

Finding solutions to (1) and (2), requires a tabulation of the multivariate Student *t* distribution. It can be reduced to the multivariate normal distribution, so that the equation (1) and (2) can be written as the following:

$$P = \Pr\left(z_1 < d_1^{'}s, z_2 < d_2^{'}s, \ldots, z_k < d_k^{'}s\right)$$

$$= \int_{-\infty}^{+\infty} F\left(d_1^{'}s, d_2^{'}s, \ldots, d_k^{'}s\right) p(s)\, ds$$

and

$$P = \Pr\left(\left|z_1\right| < d_1^{''}s, \left|z_2\right| < d_2^{''}s, \ldots, \left|z_k\right| < d_k^{''}s\right)$$

$$= \int_{-\infty}^{+\infty} G\left(d_1^{''}s, d_2^{''}s, \ldots, d_k^{''}s\right) p(s)\, ds$$

where $F\left(z_1, z_2, \ldots, z_k\right)$ is the multivariate normal c.d.f of the $z_i, G\left(z_1, z_2, \ldots, z_k\right)$ is the c.d.f. of the $\left|z_i\right|$ and $p(s)$ is the probability density function of *s*.

2.1.4 The Family of Inferences

Statistical inferences can vary significantly depending on how the family is selected, and specifically, on how many tests or comparisons are included in the family. So family of inference is the number of test hypothesis on objective of the study (Westfall el al, 1999).

2.1.5 Error Rates

Consider the problem of simultaneously testing k null hypotheses $H_{0i}; \ i = 1, \ldots, k$ and let R denote the number of rejected hypotheses. The specific k hypotheses are assumed to be known in advance, the numbers k_0 and $k_1 = k - k_0$ of true and false null hypotheses are unknown parameters, R is an observable random variable and S, T, U and V shown in Table 4 are unobservable random variables.

Table 4: Number of Rejected or Not Rejected

Number of	**Number not rejected**	**Number rejected**	
True null hypotheses	U	V	k_0
Non-true null hypotheses	T	S	k_1
	$k - R$	R	k

The error rate in multiple comparisons is explained below (Hsu, 1996; Westfall, 1999; Rafter et al, 2002; Dudoit et al, 2003);

- **Type I error**, *V*, is the number of false positives or the number of errors of incorrectly rejecting a null hypothesis when it is actually true, is usually controlled at some designated level α. The probability of a Type I error is the Type I error rate or significance level α. Weak control refers to control of the Type I error rate only when all the null hypotheses are true. Strong control refers to control of the Type I error rate under any combination of true and false null hypotheses.
- **Per-Comparison level of significance** is the error level of significance α which each of several hypotheses tests is done.
- **Type II error**, *T,* is the number of false negatives or the number of errors of incorrectly failing to reject a null hypothesis when the alternative hypothesis is true. The probability of a Type II error is β.
- **Type III error or Directional error** is the probability of misclassifying the sign of an effect**.**
- **CER** (Comparison-wise Error Rate) is the probability that an interval statistic does not contain the parameter when null hypothesis is true.
- **PCER** (Per-Comparison Error Rate) is the probability of incorrectly rejecting each of the null hypotheses that makes up the family or expected value of the number of Type I errors divided by the number of hypotheses, that is $PCER = E\left(V\right)/k$.
- **PFER** (Per-Family Error Rate) is the expected number of Type I errors, that is, $PFER = E\left(V\right)$ or the expected number of errors in the family.
- **FWE** (Family-wise Error Rate) or Experiment-wise Error Rate is the probability of at least one Type I error, that is, $FWE = \Pr\left(V \geq 1\right)$ or the probability of incorrectly rejecting at least one of the null hypotheses that make up the family. The definition of FWE depends on whether the inferences are interval-based or testing-based.

(1) For interval-based: simultaneous confidence interval

FWE = Pr (at least one interval is incorrect)

= 1-Pr (all intervals are correct)

(2) For testing-based: multiple tests of hypotheses

If the null hypotheses corresponding to $H_{01},...,H_{0m}$ are true, and all other null hypotheses are false ($H_{0(m+1)},...,H_{0k}$), then

FEW = Pr(reject at least one of $H_{01},...,H_{0m} \mid H_{01},...,H_{0m}$ all are true)

- **FDR** (False Discovery Rate) of Benjamini and Hochberg (1995) is the expected proportion of Type I errors among the rejected hypotheses, that is, $FDR = E(Q)$, where, by definition, $Q = V/R$ if R>0 and Q = 0 if R = 0

2.1.6 Power of the Test

The definition of power of a single test is

Power = Pr (reject the null hypothesis given the null hypothesis false) $= 1-\beta$, while, on the other hand power, in multiple testing and multiple comparisons, the definitions of power include (Westfall, 1999)

- Complete power = Pr (reject all H_{0i} that are false)
- Minimal Power = Pr (reject at least one H_{0i} that is false)
- Individual Power = Pr (reject a particular H_{0i} that is false)
- Proportional Power = average proportion of false H_{0i} that are rejected

In the definitions mentioned above, each definition of power is based on three items in common: (1) the probability of rejecting at least one false null hypothesis, $\Pr(S \geq 1) = \Pr(T \leq k_1 - 1)$, (2) the average probability of rejecting the false null hypotheses, $E(S)/k_1$, or average power, and (3) the probability of rejecting all false null hypotheses, $\Pr(S = k_1) = \Pr(T = 0)$ (Shaffer, 1995 referred in Dudoit et al., 2003).

2.1.7 The Closure Principle

Multiple comparisons based on the closure principle are called "closed testing" methods because the families of hypotheses are closed under intersection. A closed family is defined as one for which any subset intersection of hypotheses involving members of the family of tests are also a member of the family (Westfall, 1999).

The closure concept is shown in the following example: suppose the goal is to test the pair-wise equality of four means. There are six such pair-wise comparisons to be made. To form a closed family, we take all possible intersections among the pair-wise hypotheses. Remember that a hypothesis that is formed by an intersection of two or more hypotheses is true if and only if all the components are true. Then all possible intersections in this family are the following:

- Original pair-wise homogeneity hypotheses: $H_{12}:\mu_1=\mu_2$, $H_{13}:\mu_1=\mu_3$, $H_{14}:\mu_1=\mu_4$, $H_{23}:\mu_2=\mu_3$, $H_{24}:\mu_2=\mu_4$, $H_{34}:\mu_3=\mu_4$
- Three means homogeneity hypotheses: $H_{123}:\mu_1=\mu_2=\mu_3$, $H_{124}:\mu_1=\mu_2=\mu_4$, $H_{134}:\mu_1=\mu_3=\mu_4$, $H_{234}:\mu_2=\mu_3=\mu_4$
- Four means homogeneity hypothesis: $H_{1234}:\mu_1=\mu_2=\mu_3=\mu_4$
- Subset intersection (disjoint) hypotheses: $H_{(12)\cap(34)}:\mu_1=\mu_2$ and $\mu_3=\mu_4$, $H_{(13)\cap(24)}:\mu_1=\mu_3$ and $\mu_2=\mu_4$, $H_{(14)\cap(23)}:\mu_1=\mu_4$ and $\mu_2=\mu_3$

The closed family is presented as figure 1.

Another point to be noted in closed families is the hierarchical structure; if $H_{(12)\cap(34)}$ is true, then necessarily H_{12} and H_{34} are both true. We have an implication relationship among the hypotheses; we say that $H_{(12)\cap(34)}$ implies both H_{12} and H_{34} . The implication relations lead to a property of MCPs, termed ***coherence.*** Coherence is related to another property of multiple comparisons, called ***consonance***. Where $H^{'}$ is rejected, at least one of its components will be rejected too. (Westfall, 1999)

The methodology is not closed in the following cases (1) multiple comparisons of the means in several treatment groups by using permutation resampling and (2) multiple comparisons of proportions (binary data) in several treatment groups by using either permutation resampling or bootstrap resampling. (Westfall, 2000)

2.1.8 P-value

A P-value of a statistical test may be classified into 2 categories: unadjusted p-value (or raw p-value) and adjusted p-value. Details of an unadjusted p-value and an adjusted p-value are described below (Dudoit et al, 2003);

An unadjusted p-value (or raw p-value) is the probability of obtaining a result at least as extreme as the given data, under the null hypothesis. It can be written as: $p_i = \inf\left\{\alpha : t_i \in S_\alpha\right\} = \Pr\left(\left|T_i\right| \geq \left|t_i\right| \mid H_0\right)$, when t_i is the observed value of the test statistic, S_α is a rejection region and T_i is the test statistic for the i^{th} null hypothesis.

An adjusted p-value is the probability of obtaining a result at least as extreme as the given data, under the null hypothesis when the null hypothesis is rejected at a pre-specified level α . This is chosen for the purpose of controlling a Type I error rate (FWE, PCER, PFER or FDR). If we are interested in controlling the FWE, the adjusted p-value for hypothesis H_i is $\tilde{p}_i = \inf\left\{\alpha \in [0,1] \mid H_i \text{ is rejected at nominal FWE} = \alpha\right\}$, where the nominal FWE is the significance level α at which the specified procedure is performed.

2.1.9 Control of the Family-wise Error Rate

The procedures to control of the family-wise error rate (FWE) are single step and step-wise procedures. Step-wise procedures can be classified into the two categories of step-down and step-up procedures. Table 5 shows the formulas for adjusted p-values $\left(\tilde{p}_i\right)$ by single step and step-wise procedures.

Note: 1. p_i is the unadjusted p-value or the observation p-value for i^{th} null hypothesis.

2. $p_{(i)}$ is the observed ordered unadjusted p-value.

3. $\tilde{p}_i$ is the adjusted p-value for i^{th} null hypothesis.

4. P_l is the l^{th} random unadjusted p-value.

5. $P_{(i)}$ is the random ordered unadjusted p-value.

5. H_0^c is the complete null hypothesis. In this study, the complete null hypothesis is $H_0^c : \mu_0 = \mu_1 = \mu_2 = \ldots = \mu_k$.

6. T_l is the l^{th} random test statistic value.

7. t_i is the observed test statistics value for the i^{th} null hypothesis. In this study

$$t_i = \frac{\overline{y}_i - \overline{y}_0}{\sqrt{MSE\left(\frac{1}{r_i} + \frac{1}{r_0}\right)}} .$$

In single-step procedures, equivalent multiplicity adjustments are performed for all hypotheses. They will now be described.

- **Bonferroni** adjusted p-values: Reject i^{th} null hypothesis$\left(H_{0i}\right)$ when $p_i \le \frac{\alpha}{k}$. Then the Bonferroni adjusted p-values is $\tilde{p}_i = \min\left(kp_i, 1\right)$. See Table 6 for details.

Table 6: Example of Bonferroni Adjusted P-values

Hypothesis	Algorithm	Adjusted p-value
$\mu_0 = \mu_1$	If $p_1 \le \frac{\alpha}{k}$ then reject H_{01} else fail to reject H_{01}	$\tilde{p}_1 = \min\left(kp_1, 1\right)$
$\mu_0 = \mu_2$	If $p_2 \le \frac{\alpha}{k}$ then reject H_{02} else fail to reject H_{02}	$\tilde{p}_2 = \min\left(kp_2, 1\right)$
$\mu_0 = \mu_i$	If $p_i \le \frac{\alpha}{k}$ then reject H_{0i} else fail to reject H_{0i}	$\tilde{p}_i = \min\left(kp_i, 1\right)$
$\mu_0 = \mu_k$	If $p_k \le \frac{\alpha}{k}$ then reject H_{0k} else fail to reject H_{0k}	$\tilde{p}_k = \min\left(kp_k, 1\right)$

- **Šidák** adjusted p-values: reject H_{0i} when $p_i \leq 1-1(1-\alpha)^{\frac{1}{k}}$. The adjusted p-value is $\tilde{p}_i = 1-(1-p_i)^k$. See Table 7 for details.

Table 7: Example of Šidák Adjusted P-values

Hypothesis	Algorithm	Adjusted p-value
$\mu_0 = \mu_1$	If $p_1 \leq 1-1(1-\alpha)^{\frac{1}{k}}$ then reject H_{01} else fail to reject H_{01}	$\tilde{p}_1 = 1-(1-p_1)^k$
$\mu_0 = \mu_2$	If $p_2 \leq 1-1(1-\alpha)^{\frac{1}{k}}$ then reject H_{02} else fail to reject H_{02}	$\tilde{p}_2 = 1-(1-p_2)^k$
$\mu_0 = \mu_i$	If $p_i \leq 1-1(1-\alpha)^{\frac{1}{k}}$ then reject H_{0i} else fail to reject t H_{0i}	$\tilde{p}_i = 1-(1-p_i)^k$
$\mu_0 = \mu_k$	If $p_k \leq 1-1(1-\alpha)^{\frac{1}{k}}$ then reject H_{0k} else fail to reject H_{0k}	$\tilde{p}_k = 1-(1-p_k)^k$

- **Min P** adjusted p-values is $\tilde{p}_i = \Pr\left(\min_{1 \leq l \leq k} P_l \leq p_i \mid H_0^c\right)$. See Table 8 for details.

Table 8: Example of min P Adjusted P-values

Hypothesis	Adjusted p-value
$\mu_0 = \mu_1$	$\tilde{p}_1 = \Pr\left(\min\left(P_1, P_2, ..., P_k\right) \leq p_1 \mid H_o^c\right)$
$\mu_0 = \mu_2$	$\tilde{p}_2 = \Pr\left(\min\left(P_1, P_2, ..., P_k\right) \leq p_2 \mid H_o^c\right)$
$\mu_0 = \mu_i$	$\tilde{p}_i = \Pr\left(\min\left(P_1, P_2, ..., P_k\right) \leq p_i \mid H_o^c\right)$
$\mu_0 = \mu_k$	$\tilde{p}_k = \Pr\left(\min\left(P_1, P_2, ..., P_k\right) \leq p_k \mid H_o^c\right)$

- **Max T** adjusted p-value is $\tilde{p}_i = \Pr\left(\max_{1\le l\le k}|T_l| \ge |t_i| \mid H_0^c\right)$. See Table 9 for details.

Table 9: Example of max T Adjusted P-values

Hypothesis	Adjusted p-value
$\mu_0 = \mu_1$	$\tilde{p}_1 = \Pr\left(\max\left(\lvert T_1\rvert, \lvert T_2\rvert, ..., \lvert T_k\rvert\right) \ge \lvert t_1\rvert \mid H_o^c\right)$
$\mu_0 = \mu_2$	$\tilde{p}_2 = \Pr\left(\max\left(\lvert T_1\rvert, \lvert T_2\rvert, ..., \lvert T_k\rvert\right) \ge \lvert t_2\rvert \mid H_o^c\right)$
$\mu_0 = \mu_i$	$\tilde{p}_i = \Pr\left(\max\left(\lvert T_1\rvert, \lvert T_2\rvert, ..., \lvert T_k\rvert\right) \ge \lvert t_i\rvert \mid H_o^c\right)$
$\mu_0 = \mu_k$	$\tilde{p}_k = \Pr\left(\max\left(\lvert T_1\rvert, \lvert T_2\rvert, ..., \lvert T_k\rvert\right) \ge \lvert t_k\rvert \mid H_o^c\right)$

In step-down procedures, the hypotheses that correspond to the most significant test statistics (i.e., smallest unadjusted p-values or largest absolute test statistics) are considered successively. Let $p_{(1)} \le p_{(2)} \le \cdots \le p_{(i)} \le \cdots \le p_{(k)}$ denote the observed ordered unadjusted p-values. In other words $t_{(1)} \ge t_{(2)} \ge \cdots \ge t_{(i)} \ge \cdots \ge t_{(k)}$ denotes the observed ordered test statistics and let $H_{(1)}, H_{(2)}, ..., H_{(i)}, ..., H_{(k)}$ denote the corresponding null hypotheses. The step-down procedures are now described;

- **Bonferroni – Holm** adjusted p-values is $\tilde{p}_{(i)} = \max(\tilde{p}_{(i-1)}, (k-i+1)\, p_{(i)})$.

  See Table 10 for details.

Table 10: Algorithm of Bonferroni – Holm Adjusted P-values

Step	Algorithm	Adjusted p-value
1	If $p_{(1)} \le \frac{\alpha}{k}$ then reject $H_{(01)}$ else fail to reject all hypotheses	$\tilde{p}_{(1)} = kp_{(1)}$
2	If $p_{(2)} \le \frac{\alpha}{k-1}$ then reject $H_{(02)}$ else fail to reject $H_{(02)},\ldots,H_{(0k)}$	$\tilde{p}_{(2)} = \max\left(\tilde{p}_{(1)}, (k-1)\, p_{(2)}\right)$
i	If $p_{(k-i+1)} \le \frac{\alpha}{k-i+1}$ then reject $H_{(0i)}$ else fail to reject $H_{(0i)},\ldots,H_{(0k)}$	$\tilde{p}_{(i)} = \max(\tilde{p}_{(i-1)}, (k-i+1)\, p_{(i)})$
k	If $p_{(k)} \le \alpha$ then reject $H_{(0k)}$ else fail to reject $H_{(0k)}$	$\tilde{p}_{(k)} = \max\left(\tilde{p}_{(k-1)}, p_{(k)}\right)$

- **Šidák – Holm** adjusted p-values is $\tilde{p}_{(i)} = \max\left(\tilde{p}_{(i-1)}, 1-\left(1-p_{(i)}\right)^{k}\right)$.

  See Table 11 for details.

Table 11: Algorithm of Šidák – Holm Adjusted P-values

Step	Adjusted p-value
1	$\tilde{p}_{(1)} = 1-\left(1-p_{(1)}\right)^{k}$
2	$\tilde{p}_{(2)} = \max\left(\tilde{p}_{(1)}, 1-\left(1-p_{(2)}\right)^{k}\right)$
i	$\tilde{p}_{(i)} = \max\left(\tilde{p}_{(i-1)}, 1-\left(1-p_{(i)}\right)^{k}\right)$
k	$\tilde{p}_{(k)} = \max\left(\tilde{p}_{(k-1)}, p_{(k)}\right)$

- **Simes Modified Bonferroni** procedure controls the global Type I error rate under independence, and performed some simulations indicating that control of the FWE weakly under a variety of situations. Recent research indicates that the Simes test provides good satisfactory control of the FWE, but can sometimes exceed α (Hochberg and Rom, 1995; Sarkar and Chang, 1998, referred Westfall, 1999). For the Simes Modified Bonferroni procedure $H_{(0i)}$ is rejected $p_{(i)} \le i\alpha/k$. See Table 12.

Table 12: Algorithm of Simes Modified Bonferroni Adjusted P-values

Step	Algorithm	Adjusted p-value
1	If $p_{(1)} \le \alpha/k$ then reject $H_{(1)}$ else fail to reject all hypotheses	$\tilde{p}_{(1)} = kp_{(1)}$
2	If $p_{(2)} \le 2\alpha/k$ then reject $H_{(2)}$ else fail to reject $H_{(2)},\ldots,H_{(k)}$	$\tilde{p}_{(2)} = \max\left(\tilde{p}_{(1)}, \frac{kp_{(2)}}{2}\right)$
i	If $p_{(i)} \le i\alpha/k$ then reject $H_{(i)}$ else fail to reject $H_{(i)},\ldots,H_{(k)}$	$\tilde{p}_{(i)} = \max(\tilde{p}_{(i-1)}, \frac{kp_{(i)}}{i})$
k	If $p_{(k)} \le \alpha$ then reject $H_{(k)}$ else fail to reject $H_{(k)}$	$\tilde{p}_{(k)} = \max\left(\tilde{p}_{(k-1)}, p_{(k)}\right)$

- **Step-down min P** adjusted p-value is $\tilde{p}_{(i)} = \max_{m=1,\ldots,i}\left\{\Pr\left(\min_{l\in\{m,\ldots,k\}} P_l \le p_{(m)} \mid H_0^c\right)\right\}$.

  See Table 13 for the algorithm.

- **Step-down max T** adjusted p-value is $\tilde{p}_{(i)} = \max_{m=1,\ldots,i}\left\{\Pr\left(\max_{l\in\{m,\ldots,k\}} |T_l| \ge |t_m| \mid H_0^c\right)\right\}$.

  See Table 14 for the algorithm.

In Step-up procedures, the hypotheses that correspond to the least significant test statistics (i.e., largest unadjusted p-values or smallest absolute test statistics) are considered successively. Let $p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(i)} \geq \cdots \geq p_{(k)}$ denote the observed ordered unadjusted p-values. Alternatively let $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(i)} \leq \cdots \leq t_{(k)}$ denote the observed ordered test statistics and let $H_{(1)}, H_{(2)}, \ldots, H_{(k)}$ denote the corresponding null hypotheses. The algorithm to find adjusted p-values for two step-up procedures are shown in Table 15 and Table 16.

- **Hochberg' s Method** adjusted p-values.

Table 15: Algorithm of Hochberg's Method

Step	Algorithm	Adjusted p-value
1	If $p_{(k)} \leq \alpha$ then all hypotheses are rejected, else fail to reject $H_{(k)}$	$\tilde{p}_{(k)} = p_{(k)}$
2	If $p_{(k-1)} \leq \frac{\alpha}{2}$ then rejected $H_{(k-1)}, \ldots, H_{(1)}$, else fail to reject $H_{(k-1)}$	$\tilde{p}_{(k-1)} = \min\left(\tilde{p}_{(k)}, 2p_{(k-1)}\right)$
i	If $p_{(k-i+1)} \leq \frac{\alpha}{i+1}$ then rejected $H_{(k-i+1)}, \ldots, H_{(1)}$, else fail to reject $H_{(k-i+1)}$	$\tilde{p}_{(k-i)} = \min(\tilde{p}_{(k-i+1)}, (i+1)\, p_{(k-i+1)})$
k	If $p_{(1)} \leq \frac{\alpha}{k}$ then rejected $H_{(1)}$, else fail to reject $H_{(1)}$	$\tilde{p}_{(1)} = \min\left(\tilde{p}_{(2)}, kp_{(1)}\right)$

- **Rom's Method**; while Hochberg's procedure does not control FWE for all correlation structures, it does control FWE at a level slightly lower than α when the p-values are independent. Rom (1990) devised another step-up procedures similar to Hochberg's procedure, but with slightly more power because it has FWE = α exactly under independence of p-values. See Table 16.

Table 16: Algorithm of Rom's Method

Step	Algorithm
1	If $p_{(k)} < C_k$ then all hypotheses are rejected, else fail to reject $H_{(k)}$
2	If $p_{(k-1)} \le C_{k-1}$ then rejected $H_{(k-1)},...,H_{(1)}$, else fail to reject $H_{(k-1)}$
i	If $p_{(k-i+1)} < C_{(k-i+1)}$ then rejected $H_{(k-i+1)},...,H_{(1)}$, else fail to reject $H_{(k-i+1)}$
k	If $p_{(1)} \le C_1$ then rejected $H_{(1)}$, else fail to reject $H_{(1)}$

Step-up and step-down procedures are more powerful than single step procedures, and step-up procedures are more powerful than step-down procedures when Simes inequality $\Pr\left(P_{(1)} > \frac{\alpha}{k}, P_{(2)} > \frac{2\alpha}{k},..., P_{(i)} > \frac{i\alpha}{k},..., P_{(k)} > \alpha \mid H_0^C\right) \ge 1-\alpha$ assumption holds.

The min P and max T adjusted p-values procedures provide weak control of FWE. On the other hand, they provide strong control of FWE under the assumption of "subset pivotality". The subset pivotality is also called distribution of unadjusted p-value or joint distribution of the test statistic (or raw p-value) of the true complete null hypothesis. When the joint and even marginal distribution of the test statistic (subset pivotality) is unknown, we can use resampling methods (e.g., bootstrap, permutation) to estimate raw p-values and adjusted p-values.

The min P adjusted p-value procedure is less powerful than or equal to the Šidák adjusted p-value procedure when under the "Šidák inequality" also called "orthant dependence property": $\Pr\left(\left|T_1\right| \le c_1,...,\left|T_k\right| \le c_k\right) \ge \prod_{j=1}^{k} \Pr\left(\left|T_j\right| \le c_j\right)$.

2.1.10 The Traditional Bootstrap

Traditional bootstrapping involves resampling from the sample observations with replacement, and produces independent and identically distributed random variables (Smith and Taylor, 2001)

The basic ideas of the traditional bootstrap or independent bootstrap is to calculate the properties of estimator $\hat{\theta}\left(x_1,...,x_n\right)$ through $F_n\left(x\right)=\frac{1}{n}\sum_{i=1}^{n}\Pi X_i \leq x,$ the empirical cdf. of the sample $X_1,...,X_n$. If an estimate of $\theta\left(F\right)=\int h\left(x\right)dF\left(x\right)$ for some function n is desired, an obvious candidate is $\theta\left(F_n\right)=\int h\left(x\right)dF_n\left(x\right)$. It has become common to denote a bootstrap sample with a superscript "*", so can draw bootstrap samples with $X^{*i}=\left(X_1^*,...,X_n^*\right)\overset{iid}{\sim}F_n$. Based on drawing $X^{*i},...,X^{*B},\theta\left(F_n\right)$ can be approximated by the bootstrap estimator $\hat{\theta}\left(F_n\right)\approx\frac{1}{B}\sum_{i-1}^{B}h\left(X^{*i}\right)$ (Robert and Casella, 2004). The procedure is outlined in Figure 2.

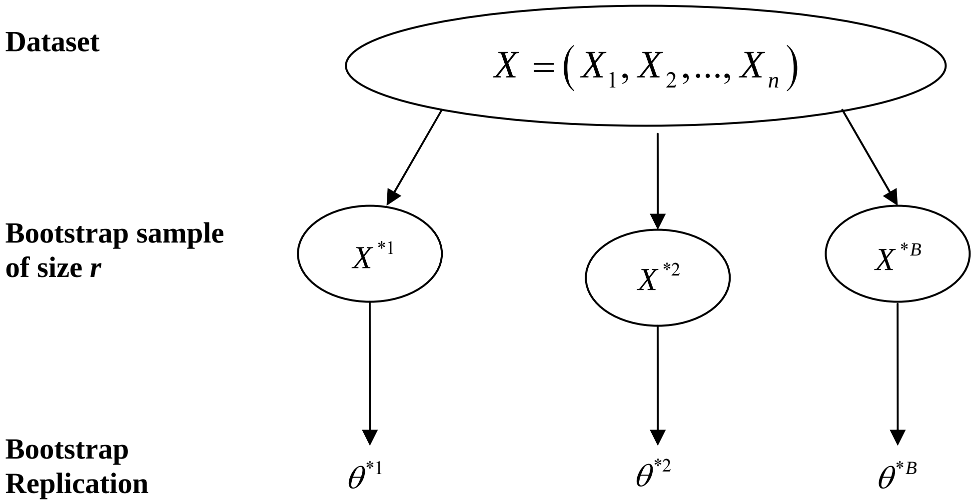

Figure 2: Independent Bootstrap Replication (Efron and Tibshirani, 1993)

2.1.11 Dependent Bootstrap

A dependent bootstrap sample is defined as the sample of size *r*, draw *r* without replacement from the collection of cn items made up of *c* copies each of the sample data $X_1, ..., X_n$, where $r \le cn$ (Smith and Taylor, 2001). The procedure is outlined in Figure 3.

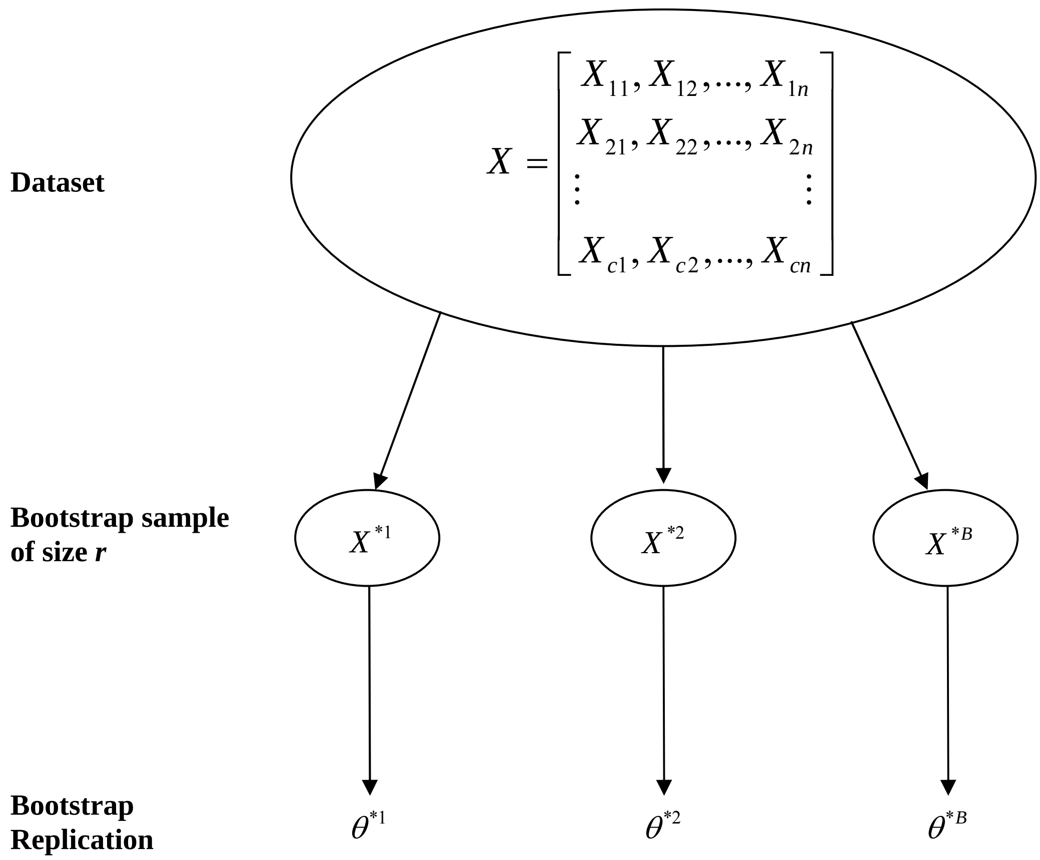

Figure 3: Dependent Bootstrap Replication

Related literatures regarding to dependent bootstrap are given below (Smith and Taylor, 2001);

In 1989, Gross introduced dependent bootstrap concepts, including the dependent bootstrap as a resampling without replacement procedure having dependent random variables which are still identically distributed.

In 1990, Wu studied resampling without replacement from sample data for the independent identically distribution (iid.) case for statistics which are asymptotically normal.

Afterward in 1993, Praestgaard and Wellner studied the "r out of cn" dependent bootstrap where $r\left(<< n\right)$ is the bootstrap sample size and n is the sample size of the original sample. This procedure could allow large bootstrap sample sizes and some asymptotic results using exchangeability arguments were given.

In 1994, Polities and Romano examined resampling without replacement from a data set to approximate the sample distribution of a statistic. Under weak assumptions, they showed that the empirical distribution of the suitably normalized value of the statistic computed for all subsamples of size r from the original data is first order asymptotically normal.

In 1997, Bertail shows second order correctness of this method for an adequately chosen resample size, but they sample without replacement from the original data with a fixed number of copies of each observation.

2.1.12 Distributions

The relevant distributions are normal distribution, multivariate normal distribution, chi-square distribution, student's t distribution and F (variance ratio) distribution. The definition and some properties of each distribution are now summarized;

2.1.12.1 Normal Distribution

The normal distribution, also called the Gaussian distribution, was first described by the French mathematician de Moivre in 1733. Properties of a normal distribution include:

Variate $N : \mu, \sigma$

Range $-\infty < x < \infty.$

Location parameter μ , the mean.

Scale parameter $\sigma > 0$, the standard deviation

Probability density function	$\dfrac{1}{\sigma(2\pi)^{1/2}} \exp\left(\dfrac{-(x-\mu)^2}{2\sigma^2}\right)$
Mean	μ
Variance	σ^2

2.1.12.2 Multivariate Normal Distribution

A multivariate normal distribution is a multivariate extension of the normal distribution. Properties of a multivariate normal distribution include:

Multivariate $MN : \underline{\mu}, \Sigma$

Quantile $x = \left[x_1, ..., x_k\right]'$ a $k \times 1$ vector.

Range $-\infty < x_i < \infty$, for $i = 1, ..., k.$

Location parameter, the $k \times 1$ mean vector, $\mu = \left[\mu_1, ..., \mu_k\right]'$,with $-\infty < \mu_i < \infty$.

Parameter Σ, the $k \times k$ positive definite variance-covariance matrix, with elements $\Sigma_{ij} = \sigma_{ij}$.

Probability density function	$(2\pi)^{-(1/2)k} \lvert\Sigma\rvert^{-1/2} \exp\left[\frac{-1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)\right]$
Mean	μ
Variance-Covariance	Σ
For individual elements MN_i	
Probability density function	$(2\pi)^{-(1/2)} \lvert\Sigma_{ii}\rvert^{-1/2} \exp\left[\frac{-1}{2}(x_i-\mu_i)' \Sigma_{ii}^{-1} (x_i-\mu_i)\right]$
Mean	μ_i
Variance	$\Sigma_{ii} = \sigma_i^2$
Covariance	$\Sigma_{ij} = \sigma_{ij}$

2.1.12.3 Chi-square Distribution

The chi-squared variate arises from the sum of the squares of a number of independent standard normal variates. Propertied of a chi-square distribution include:

Variate $\chi^2 : \nu$

Range $0 \le x < \infty$.

Shape parameter ν, degrees of freedom.

Probability density function $\dfrac{x^{(\nu-2)/2}\exp(-x/2)}{2^{\nu/2}\Gamma(\nu/2)}$ where $\Gamma(\nu/2)$ is the gamma function with argument $\nu/2$

Mean ν

Variance 2ν

2.1.12.4 Student's *t* Distribution

The Student's *t* distribution arises from the ratio of a standard normal random variable divided by the square root of a scaled chi-square random variable with ν degree of freedom. The student's *t* distribution is used to test the difference between the means of two independent samples of observations. Its properties include:

Variate $t : \nu$.

Range $-\infty < x < \infty$.

Shape parameter ν ,degrees of freedom, ν a positive integer.

Probability density function $\dfrac{\{\Gamma[(\nu+1)/2]\}}{(\pi\nu)^{1/2}\,\mathrm{T}(\nu/2)\left[1+\left(x^2/\nu\right)\right]^{(\nu+1)/2}}$

Mean 0, $\nu > 1$

Variance $\nu/(\nu-2), \nu > 2$

2.1.12.5 *F* (Variance Ratio) Distribution

The *F* variate is the ratio of two chi-squared variate scaled by their degrees of freedom. It is often used compare the ratio of variances associated with different factors. This technique is called analysis of variance (ANOVA). Properties of the *F* distribution include:

Variate $F : \nu, \omega$.

Range $0 \le x < \infty$.

Shape parameters ν, ω, positive integers, referred to as degrees of freedom.

Probability density function

$$\frac{\Gamma\left(\frac{1}{2}(\nu+\omega)\right)(\nu/\omega)^{\nu/2}\, x^{(\nu-2)/2}}{\Gamma\left(\frac{1}{2}\nu\right)\Gamma\left(\frac{1}{2}\omega\right)\left[1+(\nu/\omega)x\right]^{(\nu+\omega)/2}}$$

Mean

$$\frac{\omega}{\omega-2}, \omega > 2$$

Variance

$$\frac{2\omega^2(\nu+\omega-2)}{\nu(\omega-2)^2(\omega-4)}, \omega > 4$$

2.2 Review of the Related Literature

Reviews of the related literatures of research are included as the following;

Dunnett el al, (2001) addressed the problem of sample size determination in multiple comparisons of *k* treatments with a control for step-down and step-up testing, assuming normal data and homogeneous variances. They defined power as the probability of correctly rejecting all hypotheses for which the treatment versus control difference exceeds a specified value. They provided expressions that allow computer evaluation of the power and necessary the sample sizes for which the sample sizes to guarantee a specified power can be determined.

Romano and Wolf (2005) considered the problem of using *k* hypotheses simultaneously. They discussed finite-sample and large-sample theory of step-down methods that provide control of the family-wise error rate (FWE). To improve on the Bonferroni method or on Holm's step-down method, Westfall and Young made effective use of resampling to construct step-down methods that implicitly estimate the dependence structure of the test statistics. However, their methods depend on an assumption known as "subset pivotality". Their goal is to construct general step-down methods that do not require such an assumption. To accomplish this, they took a close look at what makes step-down procedures work; a key component is a monotonicty requirement of critical values. By imposing monotonicity on estimated critical values (which is not an assumption on the model but rather is an assumption on the model), they showed how to construct step-down tests that can be applied in a stagewise fashion so that at most *k* tests need to be computed. Moreover, at each stage, an intersection test that controls the usual probability of Type I error is calculated, which allows then to draw on an enormous resampling literature as a general means of test construction. In addition, it is possible to carry out this method using the same set of resamples for each of the intersection tests.

Hathairat (2006) studied the efficiency of closed multiple test methods with step-wise procedures in controlling the error rates for testing the difference between two population means: Hotelling's T^2 method, Bonferroni-Holm method, Hommel's method based on Simes test, Westfall-Young bootstrap method and Exact Permutation method under multivariate normal distributions with a covariance matrix which is equal to the correlation matrix dependent, variables equal sample size, equal and unequal correlation design matrices with correlation coefficient at 0.01 and 0.05 significance level. Monte Carlo simulations were performed and repeated 1,000 times for each scenario. The results showed that in most situations, Hotelling's T^2 method has an empirical Type I error rate less than lower bound of the tolerance Type I error rate controllable criterion, while others have empirical Type I error rate that lies in the interval of the tolerance criterion. The sample size and the correlation design matrix do not affect the capacity of controlling the Type I error rate but the number of dependent variables affects the capacity of controlling the Type I error rate. In addition, the Westfall-Young bootstrap method and the Exact Permutational method have empirical Type I error rates that lie in the interval of the tolerance criterion. *Under an equal correlation design matrix,* the Exact Permutational method has the highest empirical power for almost every situation, but the Hotelling's T^2 method has the highest empirical power when the correlation between dependent variables is close to zero and the sample size is greater than or equal 30. *Under an unequal correlation design matrix,* for almost every situation, the Westfall-Young bootstrap method has the highest empirical power, but Hommel's method based on Simes test has the highest empirical power when the correlation coefficient is small. Hotelling's T^2 method has the lowest empirical power in all situations. Westfall-Young bootstrap method and Exact Permutational method has similar empirical power on all situations. In addition, empirical power varies according to the sample size and the significance level but varies inversely with the correlation coefficient. The empirical power also varies according to the number of dependent variables under unequal correlation design matrix.

Somervilly and Hemmelmann (2008) proposed a new procedure to control a generalized family-wise error rate (k-FWE), defined as the probability of at least (k+1) Type I errors (k = 0 for the usual FWE). It is very powerful to control the "false discovery rate" when it come to large number of hypotheses. According to the study, they used data under the assumption of a multivariate normal distribution of the test statistic, for a fixed number of steps in step-down or step-up procedures. They showed considerable improvement in the reduction of the number of PFP: the proportion of false positives.

2.3 Examples of the Procedure

The following examples show how to construct the (1) balanced one-way ANOVA (2) unadjusted p-value (3) step-down independent bootstrap min P (4) step-down dependent bootstrap min P and (5) simes inequality procedures. This example was taken from Westfall and Wolfinger, 2000; there are 3 treatment groups and 1 control group. In each treatment groups, there are 4 replications. The sample data are presented in Table 17.

Table 17: The Sample Data

REP	GROUP			
	0: control	**1**	**2**	**3**
1	89.8	84.4	64.4	75.2
2	93.8	116.0	79.8	62.4
3	88.4	84.0	88.0	62.4
4	112.6	68.6	69.4	73.8
MEAN	**96.2**	**88.3**	**75.4**	**68.5**

2.3.1 Balanced One-way ANOVA

Model: $y_{i^*j} = \mu_{i^*} + \varepsilon_{i^*j}$; i^* = 0,1, 2,3, j = 1, 2,3, 4

Sums of Squares: $\sum_{i^*=0}^{3}\sum_{j=1}^{4}\left(y_{i^*j} - \bar{y}\right)^2 = \sum_{i^*=0}^{3}\sum_{j=1}^{4}\left(\bar{y}_{i^*} - \bar{y}\right)^2 + \sum_{i^*=0}^{3}\sum_{j=1}^{4}\left(y_{i^*j} - \bar{y}_{i^*}\right)^2$

SST = SSTrt + SSE

Null hypothesis: $H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3$

From the ANOVA presented in Table 18, the null hypothesis $\left(H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3\right)$ is rejected at significance level 0.05, because the p-value = 0.045 is less than $\alpha = 0.05$. We conclude that the means of several treatment groups are different. Then multiple comparisons with a control were used to further analyze the data.

Table 18: ANOVA Table

SV	df.	SS	MS	F	P-value
Treatment	3	$SSTrt = 1865.7$	$MSTrt = 621.9$	$F^* = \frac{MSTrt}{MSE} = 3.644$	0.045
Residual	12	$SSE = 2048.1$	$MSE = 170.6$		
Total	**15**	$\mathbf{SST = 3913.8}$			

2.3.2 Unadjusted P-values

In multiple comparisons with a control for this example, number of inferences to be tested is 3 which are $H_{01}: \mu_0 = \mu_1$, $H_{02}: \mu_0 = \mu_2$ and $H_{03}: \mu_0 = \mu_3$. Unadjusted p-values can be calculated from;

Hypothesis: $H_{0i}: \mu_0 = \mu_i;\ H_{1i}: \mu_0 \neq \mu_i;\ i = 1,2,3$

Test statistic: $$t_i^* = \frac{\overline{y}_0 - \overline{y}_i}{\sqrt{MSE\left(\frac{1}{r_0} + \frac{1}{r_i}\right)}} \sim t_{(k+1)(r-1)}(\alpha/2)$$

Therefore;

$H_{01}: \mu_0 = \mu_1;\ H_{11}: \mu_0 \neq \mu_1$

$$t_1^* = \frac{\overline{y}_0 - \overline{y}_1}{\sqrt{MSE\left(\frac{1}{r_0} + \frac{1}{r_1}\right)}} \therefore raw_{10} = 0.4092 .$$

$H_{02}: \mu_0 = \mu_2;\ H_{12}: \mu_0 \neq \mu_2$

$$t_2^* = \frac{\overline{y}_0 - \overline{y}_2}{\sqrt{MSE\left(\frac{1}{r_0} + \frac{1}{r_2}\right)}} \therefore raw_{20} = 0.0443 .$$

$H_{03}: \mu_0 = \mu_3;\ H_{13}: \mu_0 \neq \mu_3$

$$t_3^* = \frac{\overline{y}_0 - \overline{y}_3}{\sqrt{MSE\left(\frac{1}{r_0} + \frac{1}{r_3}\right)}} \therefore raw_{30} = 0.0111 .$$

2.3.3 Step-down Independent Bootstrap min P

The formula for the step-down independent bootstrap min P procedure is

$$\tilde{p}_{(i)} = \max_{m=1,\ldots,i}\left\{\Pr\left(\min_{l\in\{m,\ldots,k\}} P_l \le p_{(m)}\right)\right\},$$

when P_l is a random p-value of the test statistic, which is calculated from resampling the sample data with replacement under the complete null hypothesis $(H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3)$ and $p_{(m)}$ is the ordered observed p-value. In this example the ordered observed p-values are $p_{(1)} = raw_{30} = 0.0111$, $p_{(2)} = raw_{20} = 0.0443$ and $p_{(3)} = raw_{10} = 0.4092$. If we used the bootstrap technique with 3 repetitions of resampling the sample data, the result of the step-down independent bootstrap min P are shown in Table 19.

Repetition 1:

(1) Resampling the sample data with replacement under the complete null hypothesis, the new sample data is shown in Table 19.

Table 19: Resampling the Sample Data for the *1*st Repetition

REP	GROUP			
	0: control	**1**	**2**	**3**
1	68.6	64.4	79.8	62.4
2	89.8	69.4	116.0	64.4
3	89.8	68.6	62.4	84.4
4	84.0	79.8	79.8	79.8
MEAN	**83.1**	**70.6**	**84.5**	**72.8**

(2) Calculate the test statistic $t_i^* = \dfrac{\bar{y}_0 - \bar{y}_i}{\sqrt{MSE\left(\dfrac{1}{r_0} + \dfrac{1}{r_i}\right)}} \sim t_{(k+1)(r-1)}(\alpha/2)$

to test the null hypothesis $H_{0i} : \mu_0 = \mu_i$;

$H_{01}:\mu_0=\mu_1;\; H_{11}:\mu_0\neq\mu_1$

$$t_1^*=\frac{\bar{y}_0-\bar{y}_1}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_1}\right)}}\therefore P_{10}^{(1)}=0.2273$$

$H_{02}:\mu_0=\mu_2;\; H_{12}:\mu_0\neq\mu_2$

$$t_2^*=\frac{\bar{y}_0-\bar{y}_2}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_2}\right)}}\therefore P_{20}^{(1)}=0.8851$$

$H_{03}:\mu_0=\mu_3;\; H_{13}:\mu_0\neq\mu_3$

$$t_3^*=\frac{\bar{y}_0-\bar{y}_3}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_3}\right)}}\therefore P_{30}^{(1)}=0.3151$$

Repetition 2:

(1) Resampling the sample data with replacement under the complete null hypothesis, the new sample data is shown in Table 20.

Table 20: Resampling the Sample Data for the *2*nd Repetition

REP	**GROUP**			
	Control	**1**	**2**	**3**
1	88.0	79.8	75.2	64.4
2	88.0	84.0	62.4	75.2
3	84.4	79.8	93.8	68.6
4	62.4	73.8	64.4	84.4
MEAN	**80.7**	**79.4**	**74.0**	**73.2**

(2) Calculate the test statistic $t_i^*=\dfrac{\bar{y}_0-\bar{y}_i}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_i}\right)}}\sim t_{(k+1)(r-1)}\left(\alpha/2\right)$

to test the null hypothesis $H_{0i}:\mu_0=\mu_i$;

$H_{01}:\mu_0=\mu_1;\ \ H_{11}:\mu_0\neq\mu_1$

$$t_1^*=\frac{\bar{y}_0-\bar{y}_1}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_1}\right)}}\therefore P_{10}^{(2)}=0.8605$$

$H_{02}:\mu_0=\mu_2;\ \ H_{12}:\mu_0\neq\mu_2$

$$t_2^*=\frac{\bar{y}_0-\bar{y}_2}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_2}\right)}}\therefore P_{20}^{(2)}=0.3869$$

$H_{03}:\mu_0=\mu_3;\ \ H_{13}:\mu_0\neq\mu_3$

$$t_3^*=\frac{\bar{y}_0-\bar{y}_3}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_3}\right)}}\therefore P_{30}^{(2)}=0.3350$$

Repetition 3:

(1) Resampling the sample data with replacement under the complete null hypothesis, the new sample data is shown in Table 21.

Table 21: Resampling the Sample Data for the *3*rd Repetition

REP	GROUP			
	Control	1	2	3
1	93.8	79.8	69.4	116.0
2	116.0	62.4	84.4	93.8
3	62.4	62.4	88.4	68.6
4	62.4	84.4	69.4	62.4
MEAN	**83.7**	**72.3**	**77.9**	**85.2**

(2) Calculate the test statistic $t_i^*=\dfrac{\bar{y}_0-\bar{y}_i}{\sqrt{MSE\left(\frac{1}{r_0}+\frac{1}{r_i}\right)}}\sim t_{(k+1)(r-1)}\left(\alpha/2\right)$

to test the null hypothesis $H_{0i}:\mu_0=\mu_i$;

$H_{01} : \mu_0 = \mu_1;\ H_{11} : \mu_0 \neq \mu_1$

$$t_1^* = \frac{\bar{y}_0 - \bar{y}_1}{\sqrt{MSE\left(\frac{1}{r_0} + \frac{1}{r_1}\right)}} \therefore P_{10}^{(3)}=0.4247$$

$H_{02} : \mu_0 = \mu_2;\ H_{12} : \mu_0 \neq \mu_2$

$$t_2^* = \frac{\bar{y}_0 - \bar{y}_2}{\sqrt{MSE\left(\frac{1}{r_0} + \frac{1}{r_2}\right)}} \therefore P_{20}^{(3)}=0.6841$$

$H_{03} : \mu_0 = \mu_3;\ H_{13} : \mu_0 \neq \mu_3$

$$t_3^* = \frac{\bar{y}_0 - \bar{y}_3}{\sqrt{MSE\left(\frac{1}{r_0} + \frac{1}{r_3}\right)}} \therefore P_{30}^{(3)}=0.9124$$

Then, the step-down independent bootstrap min P adjusted p-values $\left(\tilde{p}_{(i)} = \max_{m=1,\ldots,i}\left\{\Pr\left(\min_{l\in\{m,\ldots,k\}} P_l \le p_{(m)}\right)\right\}\right)$ are shown in Table 22.

Table 22: Adjusted P-values by Step-down Independent Bootstrap min P

Detail	Ordered of observed unadjusted p-value		
	1	**2**	**3**
Hypothesis	$H_{03} : \mu_0 = \mu_3$	$H_{02} : \mu_0 = \mu_2$	$H_{01} : \mu_0 = \mu_1$
Raw p-value	$p_{(1)} = raw_{30}$=0.0111	$p_{(2)} = raw_{20} = 0.0443$	$p_{(3)} = raw_{10} = 0.4092$
$P^{(1)}$	$P_{30}^{(1)}$=0.3151	$P_{20}^{(1)}$=0.8851	$P_{10}^{(1)}$=0.2273
$P^{(2)}$	$P_{30}^{(2)}$=0.3350	$P_{20}^{(2)}$=0.3869	$P_{10}^{(2)}$=0.8605
$P^{(3)}$	$P_{30}^{(3)}$=0.9124	$P_{20}^{(3)}$=0.6841	$P_{10}^{(3)}$=0.4247
Adjusted p-values	$\tilde{p}_{(1)} = 0.0000$	$\tilde{p}_{(2)} = 0.0000$	$\tilde{p}_{(3)} = 0.3333$

Note: $P^{(1)}$ is the random p-value for the *1*st repetition.

From table 22; the adjusted p-values are

$$\tilde{p}_{(1)} = \Pr\left(\min\left(P_{10}^{(i)}, P_{20}^{(i)}, P_{30}^{(i)}\right) \le p_{(1)}\right) = 0.0000$$

$$\tilde{p}_{(2)} = \max\left\{\Pr\left(\min\left(P_{10}^{(i)}, P_{20}^{(i)}, P_{30}^{(i)}\right) \le p_{(1)}\right), \Pr\left(\min\left(P_{10}^{(i)}, P_{20}^{(i)}\right) \le p_{(2)}\right)\right\}$$

$$= 0.0000$$

$$\tilde{p}_{(3)} = \max\left\{\begin{array}{l} \Pr\left(\min\left(P_{10}^{(i)}, P_{20}^{(i)}, P_{30}^{(i)}\right) \le p_{(1)}\right) \\ \Pr\left(\min\left(P_{10}^{(i)}, P_{20}^{(i)}\right) \le p_{(2)}\right) \\ \Pr\left(P_{10}^{(i)} \le p_{(3)}\right) \end{array}\right\} = \max\left\{0, 0, \frac{1}{3}\right\}$$

$$= \frac{1}{3} = 0.3333$$

2.3.4 Step-down Dependent Bootstrap min P

The Algorithm for the step-down dependent bootstrap min P is similar to the algorithm for the step-down independent bootstrap min P, except for the method to randomize the sample data. This example showed how to resample the sample data without replacement, when the number of copies of the sample data, *c*, equals 2.

(1) Copy the sample data as below;

Table 23: Two Copies Sample Data

	REP	GROUP			
		0: control	**1**	**2**	**3**
Copies 1	1	89.8	84.4	64.4	75.2
	2	93.8	116.0	79.8	62.4
	3	88.4	84.0	88.0	62.4
	4	112.6	68.6	69.4	73.8
Copies 2	1	89.8	84.4	64.4	75.2
	2	93.8	116.0	79.8	62.4
	3	88.4	84.0	88.0	62.4
	4	112.6	68.6	69.4	73.8

(2) Resampling from 2 copies of the sample data without replacement under complete null hypothesis, the new sample data set is shown in table 24.

Table 24: Data without Replacement 2 Copies the Sample Data

REP	GROUP			
	0: control	**1**	**2**	**3**
1	75.2	84.0	62.4	62.4
2	84.4	84.4	79.8	68.6
3	89.8	112.6	64.4	69.4
4	64.4	93.8	116.0	62.4
MEAN	**78.5**	**93.7**	**80.7**	**65.7**

2.3.5 Simes Inequality

The Simes inequality: $\left(\Pr\left(P_{(i)} > \frac{\alpha i}{k}, \forall i = 1, \ldots, k \mid H_0^C\right) \geq 1-\alpha\right)$, when $P_{(i)}$ is a random ordered unadjusted p-value. Using the results in Table 22, an example of Simes inequality is shown in table 25.

Table 25: Example of Simes Inequality

Repetition	Simes Inequality
1	$\Pr\left(P_{20}^{(1)} = 0.8851 > \frac{\alpha}{3}, P_{30}^{(1)} = 0.3350 > \frac{2\alpha}{3}, P_{10}^{(1)} = 0.2273 > \alpha,\right) \geq 1-\alpha$
2	$\Pr\left(P_{10}^{(2)} = 0.8605 > \frac{\alpha}{3}, P_{20}^{(2)} = 0.3869 > \frac{2\alpha}{3}, P_{30}^{(2)} = 0.3350 > \alpha,\right) \geq 1-\alpha$
3	$\Pr\left(P_{30}^{(3)} = 0.9124 > \frac{\alpha}{3}, P_{20}^{(3)} = 0.6841 > \frac{2\alpha}{3}, P_{10}^{(3)} = 0.4247 > \alpha,\right) \geq 1-\alpha$