CHAPTER 3

METHODOLOGY

This study has two objectives: (1) to present the new step-down dependent bootstrap min P procedure for comparing several means with a control, and (2) to compare the efficiency of the new step-down dependent bootstrap min P with the traditional step-down bootstrap min P and Dunnett's *t* statistic for the following situations;

1. There are 3 treatment groups and 1 control group.
2. Each group has equal sample size.
3. The sample size (*r*) is 3, 4, 5,…, 10.
4. The error term is normally distributed with equal variance.
5. The significance level of the two sided test is set to be 0.05.
6. For dependent bootstrap procedures, let *c* (the number of the copy sample data) be equal to 2 and 4.
7. The number of bootstrap resamples (*B*) is set to be 100, 1,000 and 10,000.
8. The number of Monte Carlo simulation is set to be 1,000 repetitions.

In this study, R version 2.5.1 was used for the Monte Carlo simulation with 1,000 repetitions. The criteria for efficiency comparisons on each procedure is determined by a pre-specified and acceptable level α for the Type I error rate and the goal is to seek the test that has maximum power within the class of tests with Type I error rate at most α.

3.1 Simulation Design

The null hypothesis in this study is $H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3$. It is also known as the "complete null hypothesis". Since the Type I error rate is defined as the probability of rejecting the null hypothesis when it is true and the power of the test is the probability of rejecting the null hypothesis when it is false. Then in order to obtain the Type I error rate, we assume the complete null hypothesis is true. On the other hand, we must also assume the complete null hypothesis is false in order to obtain the power of the test.

To estimate the Type I error rate, we generated the sample data from a $N(100, 25)$ population for each treatment group. The parameters of the normal population were chosen to have the coefficient of variation (CV.) equal to $\frac{\sqrt{25}}{100} = 5.00\%$.

To determine whether each procedure is able to control the Type I error rate, the empirical type I error rate should be in the range of [0.036, 0.064].

On the other hand, to estimate the power of the test, we generated the sample data from a $N(60, 25)$ population for a control group and from a $N(100, 25)$ population for other treatment groups. The population mean parameter for a control group was chosen to be 60 in order to get the empirical power from Monte Carlo simulation comparable (table 26) and to have the coefficient of variation (CV.) to be not too different from 5.00% ($CV. = \frac{\sqrt{25}}{60} = 8.33\%$).

Table 26: Empirical Power of the test of All procedures when *r* = 3 and *B* = 10

mue	Dunnett	IN	DE2	DE4
80	0.141	0.200	0.209	0.189
75	0.190	0.265	0.234	0.250
70	0.252	0.314	0.305	0.319
65	0.352	0.413	0.413	0.399
60	0.440	0.460	0.460	0.483
55	0.508	0.543	0.540	0.574
50	0.608	0.623	0.642	0.623

The sample size was set to be 3, 4, 5,…, 10 because if the sample sizes had been set to be some number larger than 10, the empirical power would have been indistinguishable (see the empirical power when *r* = 15 in table 27).

Table 27: Empirical Power of the test of All procedures when $\mu_0 = 50$ and *B* = 10

r	Dunnett	IN	DE2	DE4
5	0.878	0.850	0.865	0.853
10	0.999	0.991	0.992	0.989
15	1.000	1.000	1.000	1.000
20	1.000	1.000	1.000	1.000
25	1.000	1.000	1.000	1.000

From the above reasons, the simulation design considered sample data from a $N(100, 25)$ population for each treatment groups to estimate the Type I error rates, and sample data from a $N(60, 25)$ population for a control group and from a $N(100, 25)$ population for other treatment groups to estimate the power of the test.

3.2 Research Process

The main structure of the research has the following two parts;

3.2.1 Type I Error Rate Estimation

After the sample data from a $N(100,25)$ population for each treatment group were generated, we calculate some statistics and process the following steps;

1. Mean Square Error (MSE)

$$MSE = \frac{\sum_{i^*=0}^{k}\sum_{j=1}^{r}\left(y_{i^*j} - \bar{y}_{i^*}\right)^2}{(k+1)(r-1)}.$$

2. Dunnett's *t* two-sided test statistic (d_i)

$$d_i = \frac{\left|\bar{y}_0 - \bar{y}_i\right|}{\sqrt{MSE\left(\frac{2}{r}\right)}},$$

for the i^{th} hypothesis; $H_{0i}: \mu_0 = \mu_i$ versus $H_{1i}: \mu_0 \neq \mu_i$. We reject H_{0i} if $d_i > D_{\alpha,k,(k+1)(r-1)}$ where $D_{\alpha,k,(k+1)(r-1)}$ is the $1-\alpha$ quantile of Dunnett's two-sided range distribution.

3. Test statistic (t_i) for testing the i^{th} null hypothesis $H_{0i}: \mu_0 = \mu_i$.

$$t_i = \frac{\bar{y}_0 - \bar{y}_i}{\sqrt{MSE\left(\frac{2}{r}\right)}}.$$

If $|t_i| > |T_i|$ then we reject H_{0i} at significance level α, where T_i is the $1-\alpha$ quantile from a Student's *t* distribution with $(k+1)(r-1)$ degrees of freedom.

4. Unadjusted p-value for the i^{th} null hypothesis $H_{0i}: \mu_0 = \mu_i$

$$p_i = \Pr\left(|T_i| \geq |t_i| \mid H_0\right)$$

5. The step-down independent bootstrap min P adjusted p-values.

(5.1) For each treatment group, resample with replacement from the observed sample data.

(5.2) Calculate the test statistic $t_i = \frac{\bar{y}_0 - \bar{y}_i}{\sqrt{MSE\left(\frac{2}{r}\right)}}$ for the i^{th} null hypothesis $\left(H_{0i} : \mu_0 = \mu_i\right)$.

(5.3) Calculate p-values from $p_i = \Pr\left(\left|T_i\right| \geq \left|t_i\right| \mid H_0\right)$.

(5.4) Repeat step (5.1) to step (5.3) until the number of bootstrap resamples (B) equals 100 or 1,000 or 10,000. Then we have 100 or 1,000 or 10,000 p-values for each null hypothesis.

(5.5) Calculate the adjusted p-value from $\tilde{p}_{(i)} = \max_{m=1,\ldots,i}\left\{\Pr\left(\min_{l\in\{m,\ldots,3\}} P_l \leq p_{(m)} \mid H_0^c\right)\right\}$, therefore

$$\tilde{p}_{(1)} = \Pr\left(\min_{l\in\{1,2,3\}} P_l \leq p_{(1)}\right) = \Pr\left(\min\left(P_1, P_2, P_3\right) \leq p_{(1)}\right)$$

$$\tilde{p}_{(2)} = \max\left\{\Pr\left(\min_{l\in\{1,2,3\}} P_l \leq p_{(1)}\right), \Pr\left(\min_{l\in\{2,3\}} P_l \leq p_{(2)}\right)\right\}$$

$$\tilde{p}_{(3)} = \max\left\{\Pr\left(\min_{l\in\{1,2,3\}} P_l \leq p_{(1)}\right), \Pr\left(\min_{l\in\{2,3\}} P_l \leq p_{(2)}\right), \Pr\left(\min_{l\in\{3\}} P_l \leq p_{(3)}\right)\right\}.$$

6. The step-down dependent bootstrap min P adjusted p-values with resampling from $c = 2$ copies of the sample data.

(6.1) For each treatment group, resample from $c = 2$ copies of the observed sample data without replacement.

(6.2) Follow step (5.2) until step (5.5).

7. The step-down dependent bootstrap min P adjusted p-values with resampling from $c = 4$ copies of the sample data.

(7.1) For each treatment group, resample from $c = 4$ copies of the observed sample data without replacement.

(7.2) Follow step (5.2) until step (5.5).

8. At this step, we have empirical p-values from 4 methods i.e., (1) Dunnett's t two-sided, (2) step-down independent bootstrap min P adjusted p-values, (3) step-down dependent bootstrap min P adjusted p-values with resampling from $c = 2$ copies of the sample data and (4) step-down dependent bootstrap min P adjusted p-values with resampling from $c = 4$ copies of the sample data. All p-values were then analysed by the closure principle to obtain the adjusted p-value of the complete null hypothesis. The adjusted p-value is the minimum p-value of each null hypothesis of multiple comparisons as in figure 4.

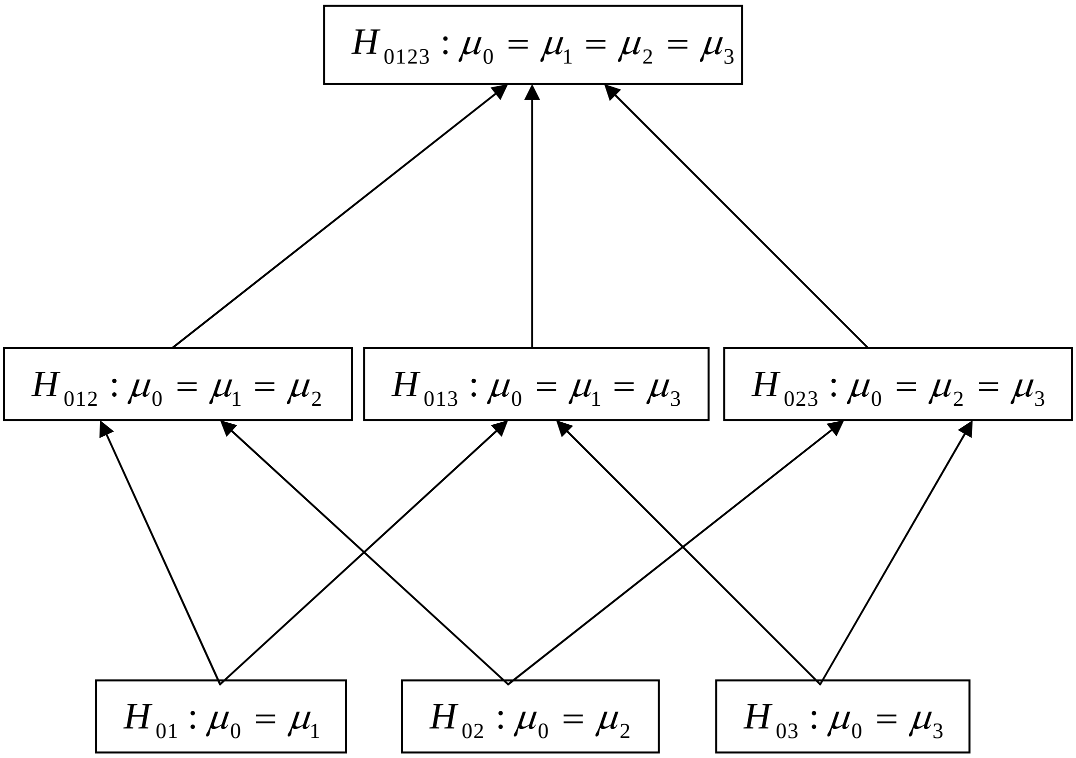

Figure 4: Closed Principle Chart

9. Repeat step 1 to step 8 for 1,000 Monte Carlo repetitions.
10. Calculate the empirical Type I error for each procedure.
11. Consider whether the empirical Type I error rate in the range of [0.036, 0.064].

3.2.2 Power of the Test Estimation

After the sample data from a $N(60, 25)$ population for a control group and from a $N(100, 25)$ population for other treatment groups were generated, we calculate some statistics and process the following steps;

1. Mean Square Error (MSE)

$$MSE = \frac{\sum_{i^*=0}^{k}\sum_{j=1}^{r}\left(y_{i^*j} - \bar{y}_{i^*}\right)^2}{(k+1)(r-1)}.$$

2. Dunnett's *t* two-sided test statistic (d_i)

$$d_i = \frac{\left|\bar{y}_0 - \bar{y}_i\right|}{\sqrt{MSE\left(\frac{2}{r}\right)}},$$

for the i^{th} hypothesis; $H_{0i} : \mu_0 = \mu_i$ versus $H_{1i} : \mu_0 \neq \mu_i$. We reject H_{0i} if $d_i > D_{\alpha,k,(k+1)(r-1)}$ where $D_{\alpha,k,(k+1)(r-1)}$ is the $1-\alpha$ quantile of Dunnett's two-sided range distribution.

3. Test statistic (t_i) for testing i^{th} null hypothesis $H_{0i} : \mu_0 = \mu_i$.

$$t_i = \frac{\bar{y}_0 - \bar{y}_i}{\sqrt{MSE\left(\frac{2}{r}\right)}}.$$

If $|t_i| > |T_i|$ then we reject H_{0i} at significance level α, where T_i is the $1-\alpha$ quantile from a Student's *t* distribution with $(k+1)(r-1)$ degrees of freedom.

4. Unadjusted p-value for i^{th} null hypothesis $H_{0i} : \mu_0 = \mu_i$

$$p_i = \Pr\left(|T_i| \geq |t_i| \mid H_0\right).$$

5. The step-down independent bootstrap min P adjusted p-values.

(5.1) For each treatment group, resample with replacement from the observed sample data.

(5.2) Calculate the test statistic $t_i = \frac{\bar{y}_0 - \bar{y}_i}{\sqrt{MSE\left(\frac{2}{r}\right)}}$ for the i^{th} null hypothesis $\left(H_{0i} : \mu_0 = \mu_i\right)$.

(5.3) Calculate p-values from $p_i = \Pr\left(\left|T_i\right| \geq \left|t_i\right| \mid H_0\right)$.

(5.4) Repeat step (5.1) to step (5.3) until the number of bootstrap resamples (B) equals 100 or 1,000 or 10,000. Then we have 100 or 1,000 or 10,000 p-values for each null hypothesis.

(5.5) Calculate the adjusted p-value from $\tilde{p}_{(i)} = \max_{m=1,\ldots,i}\left\{\Pr\left(\min_{l\in\{m,\ldots,3\}} P_l \leq p_{(m)} \mid H_0^c\right)\right\}$, therefore

$$\tilde{p}_{(1)} = \Pr\left(\min_{l\in\{1,2,3\}} P_l \leq p_{(1)}\right) = \Pr\left(\min\left(P_1, P_2, P_3\right) \leq p_{(1)}\right)$$

$$\tilde{p}_{(2)} = \max\left\{\Pr\left(\min_{l\in\{1,2,3\}} P_l \leq p_{(1)}\right), \Pr\left(\min_{l\in\{2,3\}} P_l \leq p_{(2)}\right)\right\}$$

$$\tilde{p}_{(3)} = \max\left\{\Pr\left(\min_{l\in\{1,2,3\}} P_l \leq p_{(1)}\right), \Pr\left(\min_{l\in\{2,3\}} P_l \leq p_{(2)}\right), \Pr\left(\min_{l\in\{3\}} P_l \leq p_{(3)}\right)\right\}.$$

6. The step-down dependent bootstrap min P adjusted p-values with resampling from $c = 2$ copies of the sample data.

(6.3) For each treatment group, resample from $c = 2$ copies of the observed sample data without replacement.

(6.4) Follow step (5.2) until step (5.5).

7. The step-down dependent bootstrap min P adjusted p-values with resampling from $c = 4$ copies of the sample data.

(7.3) Each treatment group, resample from $c = 4$ copies of the observed sample data without replacement.

(7.4) Follow step (5.2) until step (5.5).

8. At this step, we have empirical p-values from 4 methods i.e., (1) Dunnett's t two-sided, (2) step-down independent bootstrap min P adjusted p-values, (3) step-down dependent bootstrap min P adjusted p-values with resampling from $c = 2$ copies of the sample data and (4) step-down dependent bootstrap min P adjusted p-values with resampling from $c = 4$ copies of the sample data. All p-values were then analysed by the closure principle to obtain the adjusted p-value of the complete null hypothesis. The adjusted p-value is the minimum p-value of each null hypothesis of multiple comparisons as in figure 4.

9. Repeat step 1 to step 8 for 1,000 Monte Carlo repetitions.

10. Calculate the empirical power of the test for each procedure.

3.3 Test of Ability to Control Type I Error Rate

For each procedure we used a two-tailed z – test at significance level α to assess the ability to control the Type I error rate. The null hypothesis is

$$H_0 \;:\; \alpha = \alpha_0$$

$$H_1 \;:\; \alpha \neq \alpha_0$$

Test statistic is

$$Z = \frac{\alpha_* - \alpha_0}{\sqrt{\dfrac{\alpha_0(1-\alpha_0)}{N}}},$$

when α is the Type I error rate.

α_* is the empirical Type I error rate.

α_0 is the significance level.

N is the number of repetitions.

The null hypothesis is not rejected if $-1.96 \le Z \le 1.96$. Then, it would not be rejected if

$$-1.96 \le \frac{\alpha_* - 0.05}{\sqrt{\dfrac{0.05(1-0.05)}{1,000}}} \le 1.96$$

$$0.036 \le \alpha_* \le 0.064$$

In summary, the 95% confidence intervals of the Type I error rate is [0.036, 0.064].

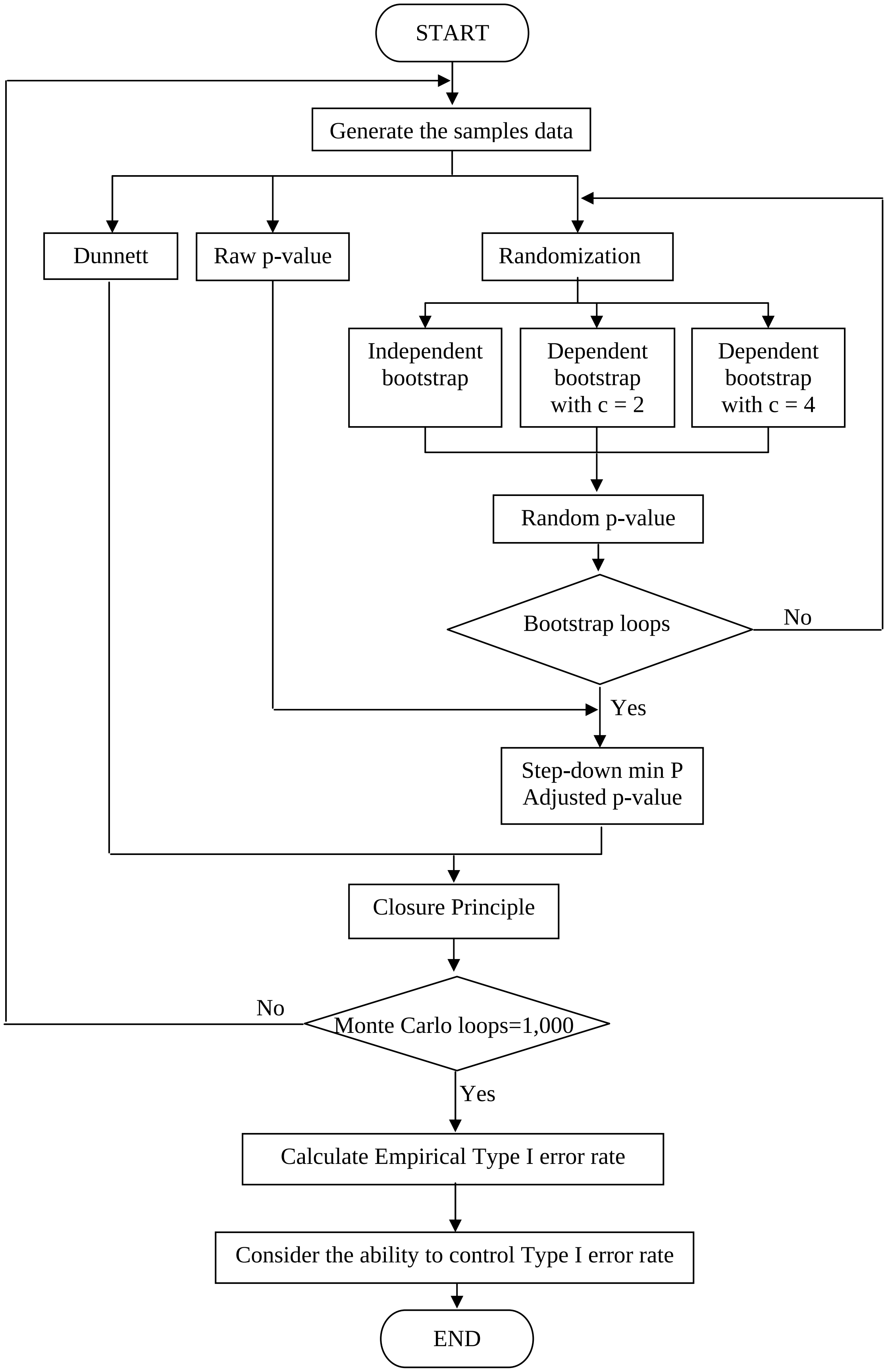

Figure 5: Flow Chart of Empirical Type I Error Rates

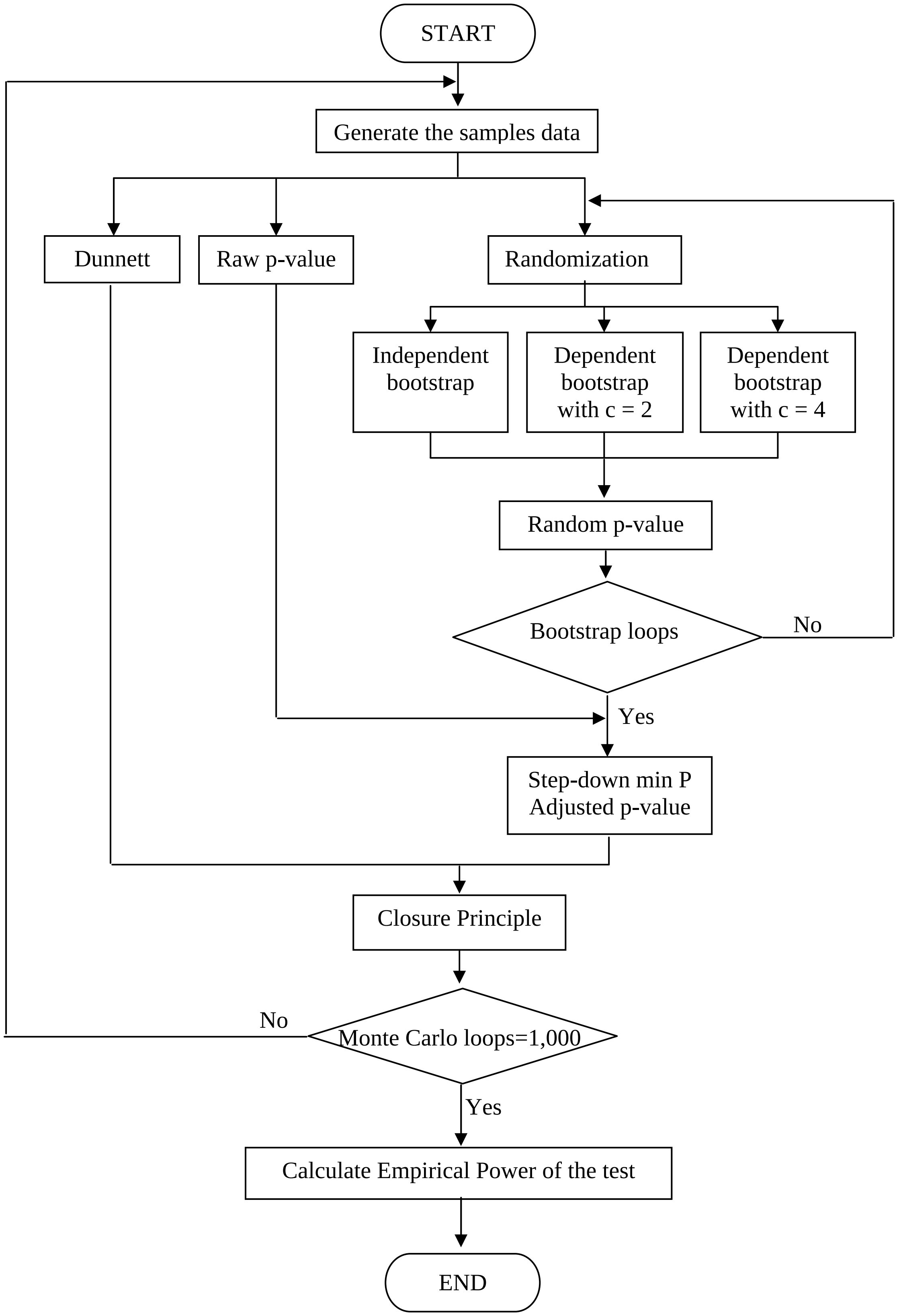

Figure 6: Flow Chart of Empirical Power of the Test