Heteroscedasticity Problems in Comparison of Several Means

Wei-Ming Wang

Advisor: Professor Miin-Jye Wen

Dissertation Defense

June 23, 2022

Outline

Motivation

Effect of Treatments Data from Juneau (2003)

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Responses	99.891	113.408	115.138	104.133
	99.651	111.645	108.581	116.034
	97.486	107.345	109.565	110.294
	101.120	108.763	110.846	93.026
Mean	99.537	110.290	111.033	105.872
SD	1.511	2.742	2.890	9.847

An experiment from Juneau (2003) was conducted to test the effects of different treatments. A patient received one of four different treatments and measured the effect. The aim of the experiment is to test

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

The p-value computed by the one-way ANOVA is 0.038 and the p-value computed by the Brown-Forsythe test is 0.107.

Motivation

Pairwise Comparisons Test for Scheffé and Games-Howell.

Scheffé Test			
Contrast	Mean Difference	Std. Error	*p*-value
1 vs 2	–10.753	3.793	0.098
1 vs 3	–11.496	3.793	0.069
1 vs 4	–6.335	3.793	0.456
2 vs 3	–0.742	3.793	0.998
2 vs 4	4.419	3.793	0.772
3 vs 4	5.161	3.793	0.617
Games-Howell Test			
Contrast	Mean Difference	Std. Error	*p*-value
1 vs 2	–10.753	1.566	**0.005***
1 vs 3	–11.496	1.631	**0.005***
1 vs 4	–6.335	4.981	0.631
2 vs 3	–0.742	1.992	0.981
2 vs 4	4.419	5.111	0.823
3 vs 4	5.161	5.131	0.757

The results are contradicted between the test equality of means and the pairwise comparison differences.

Motivation

Mutagenicity Data from Djira et al. (2020)

Treatment group	Mean	Standard deviation	Sample size
Vehicle control	2.571	1.272	7
30 mg/kg	3.800	1.095	5
50 mg/kg	6.200	1.483	5
75 mg/kg	14.000	3.937	5
100 mg/kg	20.000	4.062	5

Mutagenicity assay for 4 doses of a compound (hydroquinone) against a vehicle control. After 24h, counts of micronuclei are used in polychromatic erythrocytes as a measure for the potency to cause chromosome damage.

- The multiple comparisons of primary interest for medical research may be the comparison of each treatment with a control group. For example, the control may be a placebo, a standard treatment, or a new treatment.

Motivation

- The assumption of homogeneous variances between groups is not always realistic and may lead to incorrect decisions. For example, dose finding research may have the problem of heteroscedasticity since the data's variance relies on the dose effects.

Introduction

- Comparison of several means in the presence of heteroscedasticity has been a long-standing problem in statistics. Studies have shown that the distribution of a test statisitcs depends heavily on the unknown variances and is not robust under heteroscedastic variances.

- Welch (1938) was the first who published a test for the comparison of means of two normally distributed populations with heterogeneous variances. Welch approximated the degree of freedom of the resulting t-distribution.

- The t-distribution depends on unknown population variances σ_1^2, σ_2^2 and sample size n_1, n_2, it is impossible to derive a exact distribution of test statistics (Dudewicz and Mishra (1988), pp. 500-501)

Introduction

- When the variances are unknown and unequal, Bishop and Dudewicz (1978) developed an exact analysis of variance (ANOVA) for the means of k independent normal populations by using a two-stage sampling procedure which was originally proposed by Stein (1945).

- The two-stage sampling procedure is design-oriented. The number of required samples can be large and perhaps make the procedure non-viable during the second stage. It may not be practical because we often has only one single sample available while working on statistical data analyses.

- Chen and Lam (1989) adapted Bishop's (1976) concept of two-stage sampling procedure for a single sample by single splitting them into two data sets to work out these problems.

Introduction

- Lam (1992) extended the single-stage to the subset selection procedure.
- Chen and Chen (1998) expanded the single-stage sampling procedure to one-way and two-way layout ANOVA under heteroscedasticity.
- Chen (2001) compared typical one-stage with two-stage sampling procedures for testing the equality of means hypothesis.
- Chen et al. (2004) tested the order treatment means in the one-way layout using a one-stage procedure..
- Chen and Wen (2006) proposed an optimal confidence interval for the largest normal mean under heteroscedasticity.
- Chen and Wu (2012) develop a simultaneous confidence interval for both the biggest and lowest means of several independent normal populations under heteroscedasticity.
- Wen et al. (2017) proposed a single-stage sampling procedure of the t best populations under heteroscedasticity.

Introduction

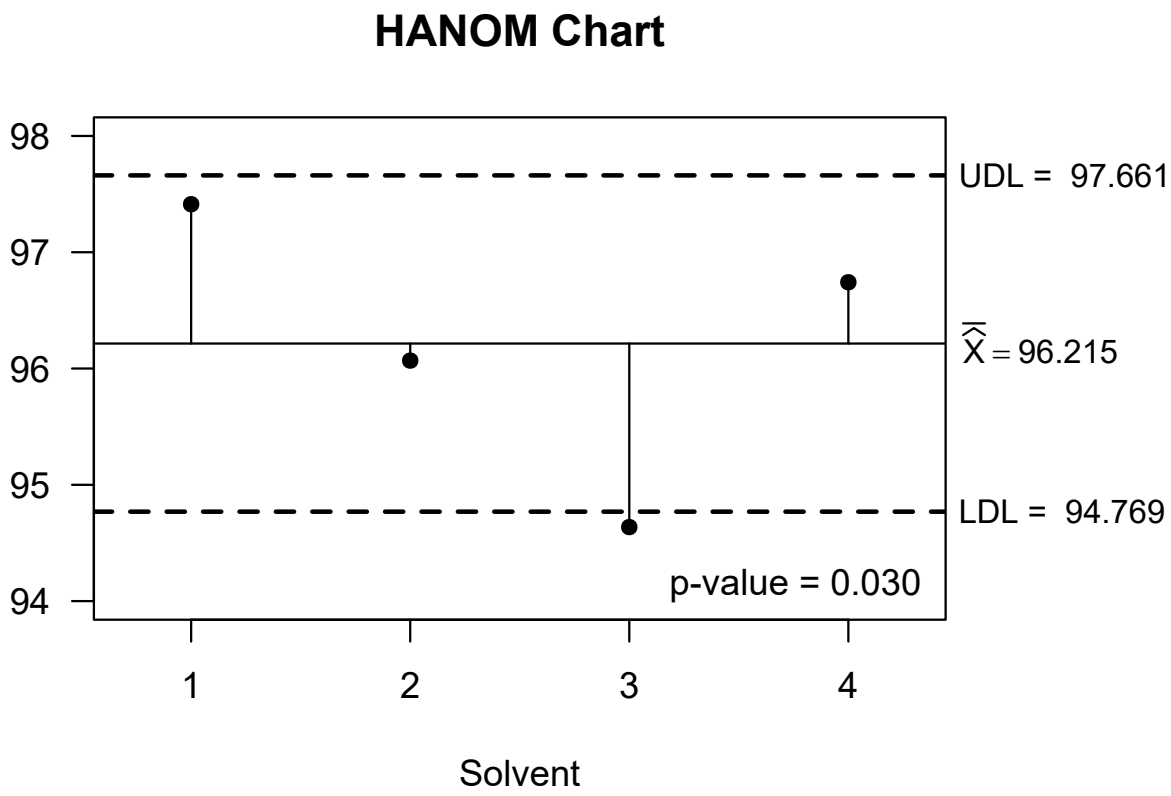
In this thesis, we applied Chen and Lam's (1989), Chen and Chen's (1998) and Chen's (2001) single-stage sampling procedure, respectively, to find critical values of the analysis of means (ANOM) in the one-way layout and apply to multiple comparison with a control (MCC) under heteroscedasticity with equal and unequal sample sizes.

Heteroscedastic ANOM

- When the null hypothesis that all the means are equal is rejected in a one way ANOVA, it indicates that at least one of the treatment mean is statistically significant. In order to find the difference of among treatment groups, we do the post hoc multiple comparisons and sometimes contradict to each other in ANOVA.

- Analysis of means (ANOM) can be considered as a useful alternative to the ANOVA to solve the above problems. ANOM is a graphical procedure that initially introduced by Ott (1967) for comparing k treatment means to specify if any of them has significant difference from the overall mean.

ANOM Chart Example



Heteroscedastic ANOM

- ANOM is similar to a control chart that displays the upper and lower decision limits, each treatment means and the overall mean. If a treatment mean falls outside of the decision limits, it indicates that this particular mean is statistically different from the overall mean.

- One assumption of the classical ANOM model is that the variances are equal, which is the same as an ANOVA model. However, equal variances is not always true for the practice.

Single-Stage Sampling Procedure

The single-stage procedure proposed by Chen and Lam's (1989) and Chen and Chen's (1998) is detailed as follows:

- Let X_{ij} $(i = 1, \ldots, k; j = 1, \ldots, n_i)$ be random samples from a normal distribution with unknown mean, μ_i, and unknown and unequal variance, σ_i^2.
- Among each i-th population, we selected the first $n_i - 1$ observations to define the sample mean and the sample variance, respectively, by

$$\overline{X}_i = \frac{\sum_{j=1}^{n_i-1} X_{ij}}{n_i - 1} \quad \text{and} \quad S_i^2 = \frac{\sum_{j=1}^{n_i-1} (X_{ij} - \overline{X}_i)^2}{n_i - 2}.$$

- Calculate the weighted sample mean

$$\widetilde{X}_i = \sum_{j=1}^{n_i} \omega_{ij} X_{ij} = U_i \sum_{j=1}^{n_i-1} X_{ij} + V_i X_{in_i} \tag{1}$$

Single-Stage Sampling Procedure

- The weights of U_i and V_i of the n_i observations satisfy the following two conditions

$$(n_i - 1)U_i + V_i = 1 \tag{2}$$

$$(n_i - 1)U_i^2 + V_i^2 = \frac{S_{[k]}^2}{n_i S_i^2} \tag{3}$$

where $S_{[k]}^2$ is the maximum of $\left\{S_1^2, \ldots, S_k^2\right\}$.

- The advantage of Equation (2) implies $\sum_{j=1}^{n_i-1} U_i + V_i = \sum_{j=1}^{n_i} \omega_{ij} = 1$ that makes the weighted sample mean, $\widetilde{X}_i$, an unbiased estimate of the group mean, μ_i, and Equation (3) implies $\sum_{j=1}^{n_i} \omega_{ij}^2 S_i^2 = S_{[k]}^2/n_i$ to eliminate the influence of the unknown and unequal variances according to Stein (1945).

Single-Stage Sampling Procedure

- The solutions of the weights of U_i and V_i are derived as

$$U_i = \frac{1}{n_i} + \frac{1}{n_i}\sqrt{\frac{1}{n_i - 1}[S^2_{[k]}/S^2_i - 1]},$$
$$V_i = \frac{1}{n_i} - \frac{1}{n_i}\sqrt{(n_i - 1)[S^2_{[k]}/S^2_i - 1]}.$$

- The sampling distribution of the weighted sample mean is denoted by

$$\widetilde{X}_i \sim N\left(\mu_i, \sum_{j=1}^{n_i} \omega_{ij}^2 \, \sigma_i^2\right).$$

Single-Stage Sampling Procedure

- Given the sample variance, S_i^2 $(i = 1, \ldots, k)$, transformation of the weighted sample mean

$$T_i = \frac{\widetilde{X}_i - \mu_i}{\sqrt{\sum_{j=1}^{n_i} \omega_{ij}^2 \, S_i^2}} = \frac{\widetilde{X}_i - \mu_i}{S_{[k]}/\sqrt{n_i}}, \quad i = 1, \ldots, k,$$

  is a conditional normal distribution with mean zero and variance σ_i^2/S_i^2 $(i = 1, \ldots, k)$ and is also an unconditional Student's t distribution with degrees of freedom $n_i - 2$ (Wen and Chen, 1994; Chen and Chen, 1998).

Single-Stage Sampling Procedure with Balanced Design

The One-Way Layout Model

- Let $X_{ij} = \mu_i + \varepsilon_{ij} \sim N\left(0,\ \sigma_i^2\right), i = 1, \ldots, k; j = 1, \ldots, n.$
  We want to test

$$H_0 : \mu_1 = \ldots = \mu_i = \ldots = \mu_k$$

- A HANOM decision chart has three lines:
  - Centerline

$$\overline{\widetilde{X}} = \frac{\sum_{i=1}^{k} \widetilde{X}_i}{k}$$

  - Lower Decision Line (LDL) and Upper Decision Line (UDL)

$$\overline{\widetilde{X}} \pm d^* \cdot \frac{S_{[k]}}{\sqrt{n}}$$

where $d^* = d^*(\frac{\alpha}{2}, k, df = n - 2) > 0$ is the critical value for HANOM.

Determination of Critical Values (Balanced Design)

- The rejection region of treatment i as

$$\left| \frac{\widetilde{X}_i - \overline{\widetilde{X}}}{S_{[k]}/\sqrt{n}} \right| > d^*, \ i = 1, \ldots, k.$$

- Determination of Critical Values:

$$\begin{aligned}
1 - \alpha =& P\left(\left| \frac{\widetilde{X}_i - \overline{\widetilde{X}}}{S_{[k]}/\sqrt{n}} \right| \leq d^*, \ i = 1, \ldots, k \right) \\
=& P\Bigg(- d^* \leq \frac{\widetilde{X}_i - \mu_i}{S_{[k]}/\sqrt{n}} - \frac{\overline{\widetilde{X}} - \mu_\cdot}{S_{[k]}/\sqrt{n}} + \frac{\mu_i - \mu_\cdot}{S_{[k]}/\sqrt{n}} \leq d^*, \ i = 1, \ldots, k \Bigg). \\
=& P\left(-d^* \leq T_i - \bar{T} \leq d^*, \ i = 1, \ldots, k \right) \quad \text{where } T_i \sim t(n-2) \\
=& P\left(-d^* < \widetilde{T}_{[1]} \leq \widetilde{T}_{[k]} < d^* \right) \quad \text{where } \widetilde{T}_i = T_i - \bar{T}
\end{aligned}$$

Since $(\widetilde{T_1}, \ldots, \widetilde{T_k})$ is symmetric by Wu and Chen (1998), we let

$$P\left(\widetilde{T}_{[1]} < -d^* \right) = \frac{\alpha}{2} = P\left(\widetilde{T}_{[k]} > d^* \right)$$

Critical Values

The simulated values of d^* when $1-\alpha = 0.9, 0.95, 0.99$ are given in Tables A1-A3. We also developed an interface using *R* shiny for HANOM.

`https://weiming3524.shinyapps.io/HANOM/`

	Number of Treatments, k																		
df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	12.69	19.65	26.13	32.54	38.71	45.05	51.42	57.88	63.81	70.03	76.31	82.48	88.65	94.64	101.38	107.79	113.84	120.71	126.73
2	3.25	4.39	5.27	6.04	6.74	7.38	7.98	8.57	9.11	9.61	10.10	10.58	11.01	11.44	11.84	12.24	12.61	12.99	13.39
3	2.29	2.99	3.48	3.89	4.24	4.55	4.83	5.08	5.32	5.54	5.74	5.94	6.13	6.29	6.47	6.62	6.77	6.92	7.07
4	1.97	2.54	2.92	3.22	3.48	3.70	3.89	4.06	4.23	4.37	4.50	4.63	4.75	4.86	4.96	5.05	5.15	5.24	5.33
5	1.81	2.32	2.65	2.91	3.12	3.30	3.45	3.59	3.72	3.83	3.94	4.03	4.12	4.20	4.28	4.35	4.43	4.49	4.56
6	1.73	2.19	2.50	2.72	2.92	3.07	3.21	3.33	3.43	3.53	3.62	3.69	3.77	3.84	3.91	3.96	4.03	4.08	4.13
7	1.66	2.11	2.39	2.61	2.78	2.92	3.04	3.15	3.25	3.33	3.41	3.48	3.55	3.61	3.66	3.72	3.77	3.81	3.86
8	1.62	2.06	2.33	2.53	2.69	2.82	2.93	3.03	3.12	3.20	3.27	3.33	3.39	3.45	3.50	3.55	3.60	3.64	3.68
9	1.59	2.01	2.28	2.47	2.62	2.75	2.86	2.94	3.03	3.10	3.17	3.23	3.28	3.33	3.39	3.43	3.47	3.51	3.55
10	1.57	1.98	2.23	2.42	2.57	2.69	2.79	2.88	2.96	3.03	3.09	3.15	3.20	3.25	3.30	3.34	3.38	3.41	3.45
11	1.55	1.96	2.21	2.38	2.53	2.64	2.74	2.83	2.90	2.97	3.03	3.08	3.13	3.18	3.22	3.27	3.30	3.34	3.37
12	1.54	1.94	2.18	2.36	2.49	2.61	2.70	2.78	2.86	2.92	2.98	3.03	3.08	3.13	3.17	3.21	3.24	3.28	3.31
13	1.52	1.92	2.16	2.33	2.47	2.58	2.67	2.75	2.82	2.88	2.94	2.99	3.04	3.08	3.12	3.16	3.19	3.22	3.25
14	1.51	1.91	2.14	2.31	2.44	2.55	2.64	2.72	2.79	2.85	2.91	2.95	3.00	3.04	3.08	3.12	3.15	3.18	3.21
15	1.50	1.89	2.13	2.30	2.43	2.53	2.62	2.70	2.77	2.83	2.88	2.92	2.97	3.01	3.05	3.08	3.12	3.15	3.17
16	1.49	1.88	2.11	2.28	2.41	2.51	2.60	2.68	2.74	2.80	2.85	2.90	2.94	2.98	3.02	3.05	3.09	3.11	3.14
17	1.49	1.87	2.10	2.27	2.40	2.50	2.58	2.66	2.72	2.78	2.83	2.88	2.92	2.96	2.99	3.03	3.06	3.09	3.11
18	1.48	1.86	2.09	2.26	2.38	2.48	2.57	2.64	2.71	2.76	2.81	2.86	2.90	2.94	2.97	3.00	3.03	3.06	3.09
19	1.47	1.86	2.08	2.24	2.37	2.47	2.55	2.63	2.69	2.74	2.80	2.84	2.88	2.92	2.95	2.98	3.01	3.04	3.07
20	1.47	1.85	2.08	2.24	2.36	2.46	2.54	2.61	2.68	2.73	2.78	2.82	2.86	2.90	2.94	2.97	3.00	3.02	3.05
40	1.43	1.79	2.01	2.15	2.27	2.36	2.43	2.50	2.56	2.60	2.65	2.68	2.72	2.75	2.78	2.81	2.83	2.86	2.88
60	1.41	1.77	1.99	2.13	2.24	2.33	2.40	2.46	2.52	2.56	2.60	2.64	2.67	2.70	2.73	2.76	2.78	2.81	2.83
120	1.40	1.76	1.96	2.10	2.21	2.30	2.37	2.43	2.48	2.52	2.56	2.60	2.63	2.66	2.69	2.71	2.73	2.76	2.78
∞	1.39	1.74	1.94	2.08	2.18	2.27	2.33	2.39	2.44	2.48	2.52	2.55	2.59	2.61	2.64	2.66	2.69	2.71	2.73

The interface by *R* shiny

Single-stage Sampling Procedure for Heteroscedasticity ANOM

File input

Browse... treatment_c.csv

Upload complete

Download Toy Dataset

Significant Level (α):

0.05

Run

Data type: Balanced sample sizes

Describe statistic | Test for normality | Test for variances | HANOM

HANOM Chart

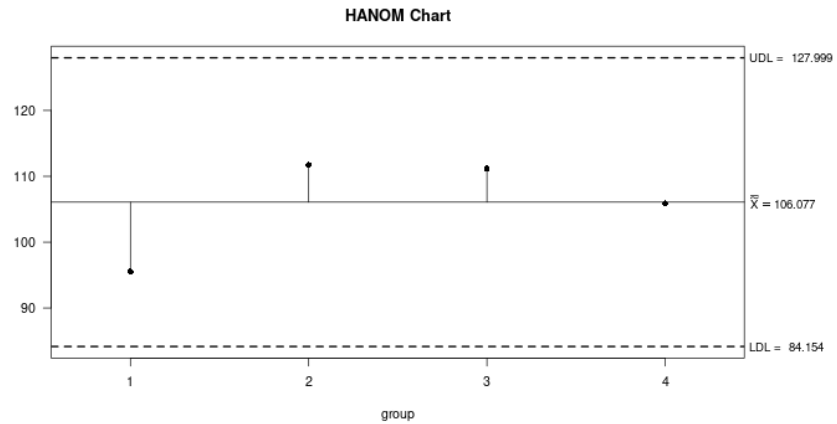


Table. Summary Statistics

	Treat.1	Treat.2	Treat.3	Treat.4
n	4.000	4.000	4.000	4.000
X_bar_i	99.009	110.799	111.095	110.154
S_i	1.325	3.119	3.536	5.952
U_i	0.882	0.485	0.445	0.250
V_i	-1.647	-0.454	-0.336	0.250
X_tilde_i	95.534	111.723	111.178	105.872

Critical Value (standard error)

7.367 (0.031)

P-value

0.227

p-value (Balanced Design)

- The p-value for HANOM can be identified as a value that the weighted sample mean, $\widetilde{X}_i$, is greatest distance away from the weighted overall mean, $\overline{\widetilde{X}}$ (Nelson et al. (2005))

$$
\begin{aligned}
& P\left(\left|\frac{\widetilde{X}_i-\overline{\widetilde{X}}}{S_{[k]}/\sqrt{n}}\right| \leq d^*\left(\frac{p}{2}, k, n-2\right),\ i=1,\ldots,k\right) \\
&= P\left(\max_i\left|\frac{\widetilde{X}_i-\overline{\widetilde{X}}}{S_{[k]}/\sqrt{n}}\right| \leq d^*\left(\frac{p}{2}, k, n-2\right)\right) \\
&= 1-p.
\end{aligned}
$$

Using the R simulation program, we increase p by 0.001 from 0 to 1. By running the HANOM procedure, we try to find the $(100\frac{p}{2})$ quantile of d^* distribution and could find the minimized p as the p-value for HANOM.

Single-Stage Sampling Procedure $\mathcal{P}_1$ with Unbalanced Design

Under an unbalanced design, the degrees of freedom are different among the k Student's t distribution, and this joint t distribution may be not symmetric which may cause troubles in solving the critical value.

- We consider

$$X_{ij} = \mu_i + \varepsilon_{ij} \sim N\left(0,\ \sigma_i^2\right),\ \ i = 1, \ldots, k; j = 1, \ldots, n_i.$$

- We used Chen and Chen's (1998) single-stage sampling procedure that is referred to as $\mathcal{P}_1$ to solve the HANOM and its decision line as follows:

$$\overline{\overline{X}} \pm h^* \cdot \frac{S_{[k]}}{\sqrt{n_i}}$$

where $h^* = h^*(\frac{\alpha}{2}, k, n_1 - 2, \ldots, n_k - 2)(> 0)$ is the critical value for HANOM.

Critical Values (Unbalanced Design)

Determination of Critical Values for HANOM:

$$
\begin{aligned}
1-\alpha =& P\left(\left|\frac{\widetilde{X}_i-\overline{\widetilde{X}}}{S_{[k]}/\sqrt{n_i}}\right| \leq h^*,\ i=1,\ldots,k\right) \\
=& P\left(-h^* \leq \left(\frac{k-1}{k}\right) T_i - \sqrt{n_i}\left(\frac{1}{k}\sum_{z\neq i}^{k}\frac{T_z}{\sqrt{n_z}}\right) \leq h^*,\ i=1,\ldots,k\right) \\
=& P\left(-h^* \leq \widetilde{T}_{[1]} \leq \widetilde{T}_{[k]} \leq h^*\right) \qquad \text{where } T_i \sim t(n_i-2) \text{ and } T_z \sim t(n_z-2) \\
\geq& 1-2\max\left\{P\left(\widetilde{T}_{[1]} < -h^*\right), P\left(\widetilde{T}_{[k]} > h^*\right)\right\}
\end{aligned}
$$

Therefore, we selected the larger critical value as h^* according to $\max\left\{P\left(\widetilde{T}_{[1]} < -h^*\right), P\left(\widetilde{T}_{[k]} > h^*\right)\right\} \leq \frac{\alpha}{2}$.

Modified Single-Stage Sampling Procedure $\mathcal{P}_2$ with Unbalanced Design

The modified single-stage sampling procedure for HANOM is briefly described as follows:

1. For each i population of X_{ij} $(i = 1, \ldots, k; j = 1, \ldots, n_i)$, we randomly select n_0 observations to be the minimum of $n_i - 1$ to calculate the initial sample mean and the sample variance, respectively.
2. Calculate the weighted sample mean

$$\widehat{X}_i = \sum_{j=1}^{n_i} \omega_{ij} X_{ij} = U_i \sum_{j=1}^{n_0} X_{ij} + V_i \sum_{j=n_0+1}^{n_i} X_{ij}$$

where the weights are derived as

$$U_i = \frac{1}{n_i} + \frac{1}{n_i}\sqrt{\frac{n_i - n_0}{n_0}\left[n_i z^* / S^2_{i,n_0} - 1\right]}$$

$$V_i = \frac{1}{n_i} - \frac{1}{n_i}\sqrt{\frac{n_0}{n_i - n_0}\left[n_i z^* / S^2_{i,n_0} - 1\right]}$$

and z^* is the maximum of $\left\{\frac{S_1^2}{n_1}, \ldots, \frac{S_k^2}{n_k}\right\}$.

3. Determinate the critical value

$$\begin{aligned}\alpha =& P\left(\left|\frac{\widehat{X}_i - \overline{\widehat{X}}}{\sqrt{z^*}}\right| > d^*,\ i = 1, \ldots, k\right)\\ \overset{H_0}{=}& P\left(\left|T_i^* - \bar{T}^*\right| > d^*,\ i = 1, \ldots, k\right)\\ =& P\left(\left|\widehat{T}_i\right| > d^*,\ i = 1, \ldots, k\right)\end{aligned}$$

, where $T_i^*, \quad i = 1, \ldots, k$, following the t distribution with degrees of freedom $n_0 - 1$. The critical value, $d^* = d^*(\frac{\alpha}{2}, k, n_0 - 1)$ is obtained by $P(\widehat{T}_{[k]} < d^*) = 1 - \alpha/2$.

4. The LDL and UDL for HANOM decision chart are $\overline{\widehat{X}} \pm d^* \cdot \sqrt{z^*}$, respectively.

The interface by *R* shiny

Modified Single-stage Sampling Procedure for Heteroscedasticity ANOM (HANOM)

HANOM | Calculate Power | Calculate Sample Size

The HANOM Chart and Summary Statistics

Toy Dataset

Download file

File input

Browse... Bishop.csv

Upload complete

Type I Error Rate (α):

0.05

Run

Describe statistic | Test for normality | Test for variances | HANOM(P1) | HANOM(P2)

HANOM Chart

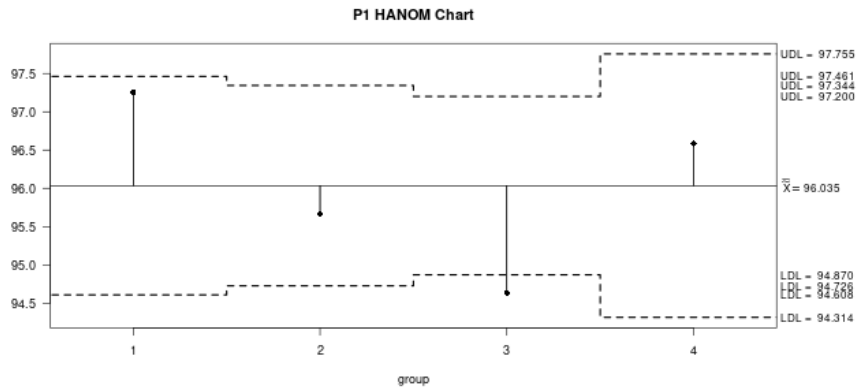


Table. Summary Statistics

	Treat.1	Treat.2	Treat.3	Treat.4
n_i	16.000	19.000	24.000	11.000
X_bar_i	97.103	95.614	94.566	97.160
S_i	1.482	2.109	2.301	0.878
U_i	0.082	0.058	0.042	0.161
V_i	-0.225	-0.045	0.042	-0.606
X_tilde_i	97.252	95.666	94.637	96.584

Critical Value

	Mean	S.E.
Critical Value	2.480	0.009

P-value

0.015

Determination of Power and Sample Size

- The power function for the HANOM

$$\begin{aligned}\text{power}(\boldsymbol{\mu}) &= 1-P\left(-h^* \le \frac{\widetilde{X}_i-\mu_i}{S_{[k]}/\sqrt{n_i}} - \frac{\overline{\widetilde{X}}-\mu_.}{S_{[k]}/\sqrt{n_i}} + \frac{\mu_i-\mu_.}{S_{[k]}/\sqrt{n_i}} \le h^*,\ i=1,\ldots,k \right) \\ &= 1-P\left(-h^* \le \widetilde{T}_i + \frac{\mu_i-\mu_.}{S_{[k]}/\sqrt{n_i}} \le h^*,\ i=1,\ldots,k \right) \qquad (4)\end{aligned}$$

- The power is determined using the least favorable configuration (LFC) of the means on the subspaces

$$M_\delta = \left\{ \boldsymbol{\mu} = (\mu_1,\ldots,\mu_k)' : \max_{i,j} |\mu_i - \mu_j| \ge \delta \right\}.$$

It is shown in Nelson and Dudewicz (2002) that on M_δ the LFC for HANOM is a configuration of the form

$$\boldsymbol{\mu} = (\delta/2, -\delta/2, 0, \ldots, 0)'.$$

Determination of Power

- The power function from Equation (4) will be rewritten by

$$\begin{aligned}\text{power}(\boldsymbol{\mu}) &= 1 - P\left(-h^* \leq \widetilde{T}_i^* \leq h^*,\ i = 1, \ldots, k\right) \\ &\geq 1 - P\left(\widetilde{T}_{[1]}^* < -h^*\right) - P\left(\widetilde{T}_{[k]}^* > h^*\right),\end{aligned}$$

  where $\widetilde{T}_i^* = \widetilde{T}_i + \frac{\mu_i - \mu_\cdot}{S_{[k]}/\sqrt{n_i}}, i = 1, \ldots, k$, while $\widetilde{T}_{[1]}^*$ is the minimum of $\left\{\widetilde{T}_1^*, \ldots, \widetilde{T}_k^*\right\}$ and $\widetilde{T}_{[k]}^*$ is the maximum of $\left\{\widetilde{T}_1^*, \ldots, \widetilde{T}_k^*\right\}$.
- The power value given the level of significance, α, number of treatments, k, degrees of freedom, $n_i - 2$ where $i = 1, \ldots, k$, the largest standard deviation, $S_{[k]}$, and the amount δ could be calculated in a similar method to the critical values.

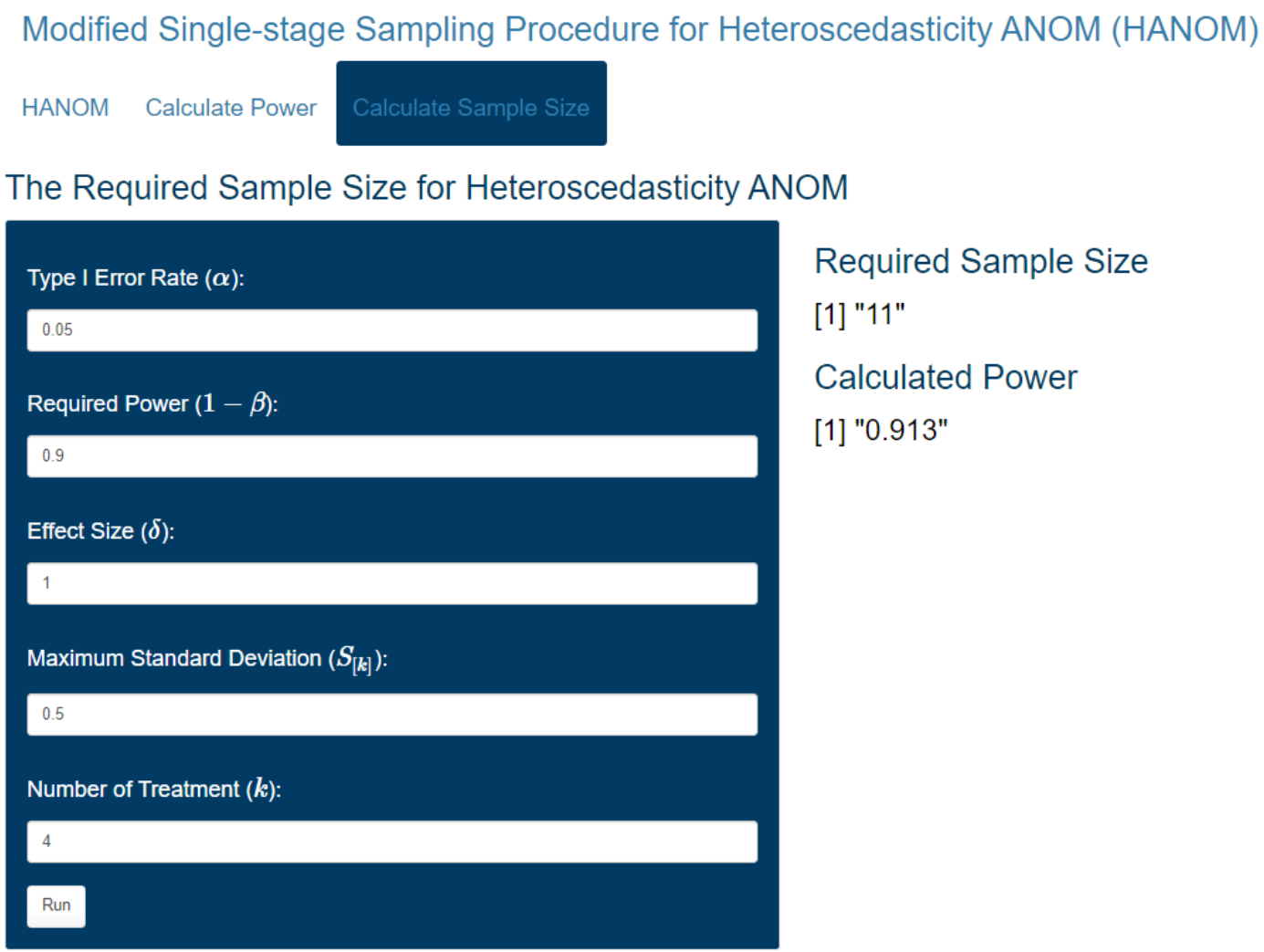
Determination of Sample Size

For a given level of significance, α, and a desired power of HANOM, $1-\beta$, the sample size n_i could be determined. We use a grid-searching method to obtain the required sample size as follows:

1. Take an initial sample size $n_{0i} = 3, i = 1, \ldots, k$. The critical value for HANOM can be obtained by a given level of significance, α.
2. We may calculate the power value for HANOM from Equation (4) under the initial sample size. If the power value meets the intended level, the required sample size is n_{0i}. If not, we raise n_{0i} by 1 until it reaches the desired power.
3. Iteratively repeat step ❶ and ❷ until the power reaches the desired $1-\beta$. The sample size n_i is currently the smallest size necessary to ensure the given power.

We can get these values from an interface of *R* Shiny. The URL of the interface can accessed at `https://weiming3524.shinyapps.io/hanom_ub/`.

The interface by *R* shiny



HANOM sample sizes when $\alpha = 0.05$

	$\delta/S_{[k]} = 2.25$ power					$\delta/S_{[k]} = 2.00$ power					$\delta/S_{[k]} = 1.75$ power				
k	0.70	0.80	0.85	0.9	0.95	0.70	0.80	0.85	0.9	0.95	0.70	0.80	0.85	0.9	0.95
3	7	8	8	9	10	8	9	9	10	12	9	10	11	12	14
4	8	9	9	10	11	9	10	10	11	13	10	11	12	13	15
5	8	9	10	10	11	9	10	11	12	14	11	12	13	14	16
6	9	10	10	11	12	10	11	12	13	14	12	13	14	15	17
7	9	10	11	11	13	11	11	12	13	15	12	14	15	16	18
8	10	10	11	12	13	11	12	13	14	15	13	14	15	16	19
9	10	11	11	12	13	11	12	13	14	16	13	15	15	17	19
10	10	11	12	12	14	12	13	13	14	16	14	15	16	17	20
15	11	12	13	14	15	13	14	15	16	17	15	17	18	19	22
20	12	13	14	14	16	14	15	16	17	18	16	18	19	20	23

	$\delta/S_{[k]} = 1.50$ power					$\delta/S_{[k]} = 1.25$ power					$\delta/S_{[k]} = 1.00$ power				
k	0.70	0.80	0.85	0.9	0.95	0.70	0.80	0.85	0.9	0.95	0.70	0.80	0.85	0.9	0.95
3	11	13	14	15	18	14	17	18	20	24	20	24	26	30	35
4	12	14	15	17	19	16	18	20	23	27	23	27	29	33	39
5	13	15	16	18	21	17	20	22	24	28	25	29	32	36	42
6	14	16	17	19	22	18	21	23	26	30	27	31	34	38	45
7	15	17	18	20	23	19	23	24	27	31	28	33	36	40	47
8	16	18	19	21	24	20	23	25	28	32	29	34	38	42	49
9	16	18	20	21	25	21	24	26	29	33	30	35	38	43	50
10	17	29	20	22	25	22	25	27	30	35	31	36	40	44	52
15	19	21	22	24	28	24	28	30	33	38	35	40	44	49	57
20	20	22	24	26	29	26	30	32	35	40	38	43	47	52	59

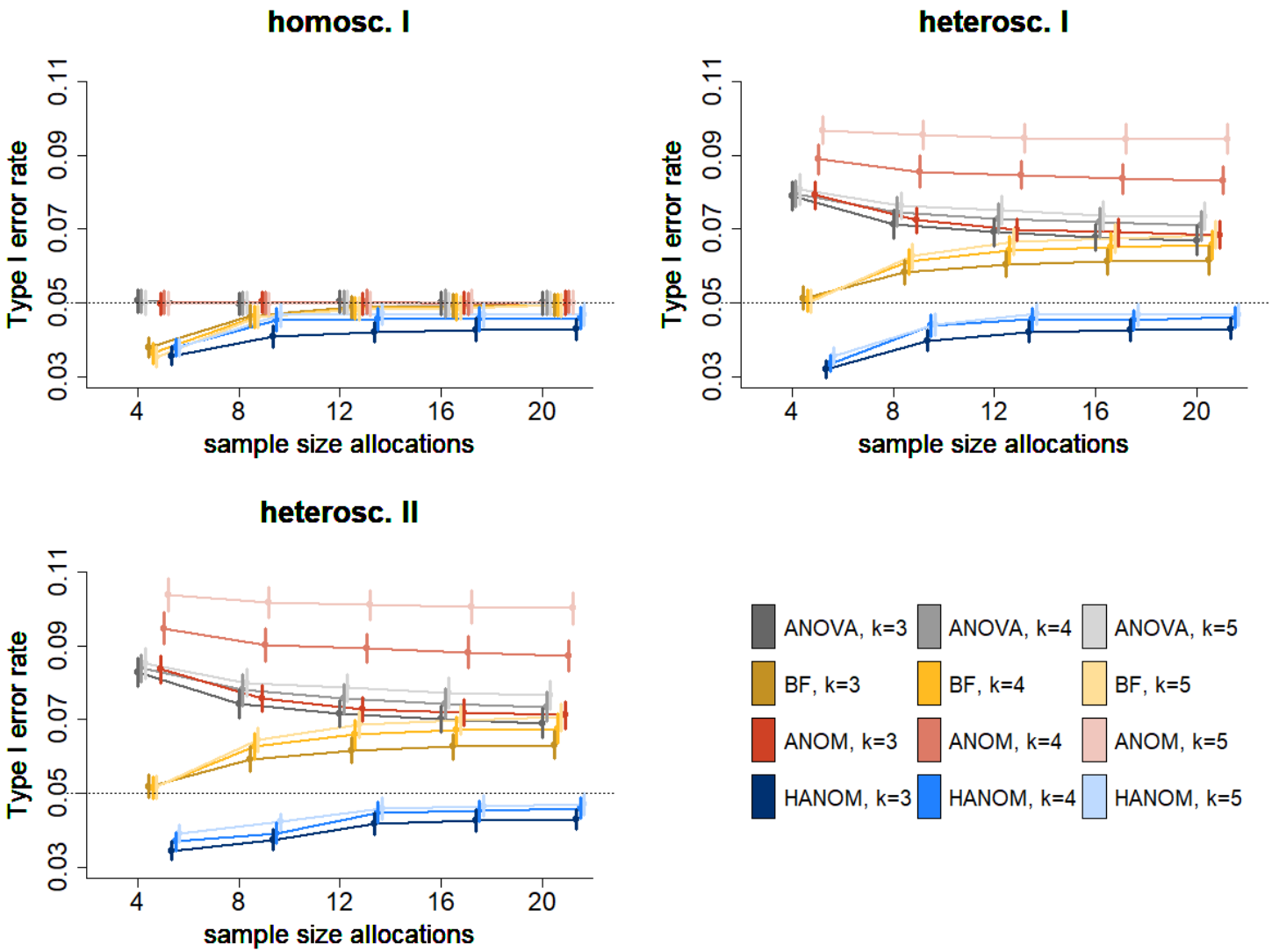
Simulations in Balanced Design

Similar to Hasler's (2016) scenarios, the expected value of all treatments was the same, $\mu_i = 100$ $(i = 1, \ldots, k)$. Five sample size allocations were chosen, i.e. $n = 4, 8, 12, 16, 20$.

- The homoscedastic case (case 1): $\sigma_i = 10, i = 1, \ldots, k$.
- The heteroscedastic case I (case 2):
  $\sigma_i = 10 + 40(i-1)/(k-1), i = 1, \ldots, k$.
- The heteroscedastic case II (case 3):
  $\sigma_i = 10 + 60(i-1)/(k-1), i = 1, \ldots, k$.

The empirical type I error rates are calculated by dividing the number of times we rejected the null hypothesis by the total number of 5000 trials. To obtain variability, we repeat this procedure 200 simulations.

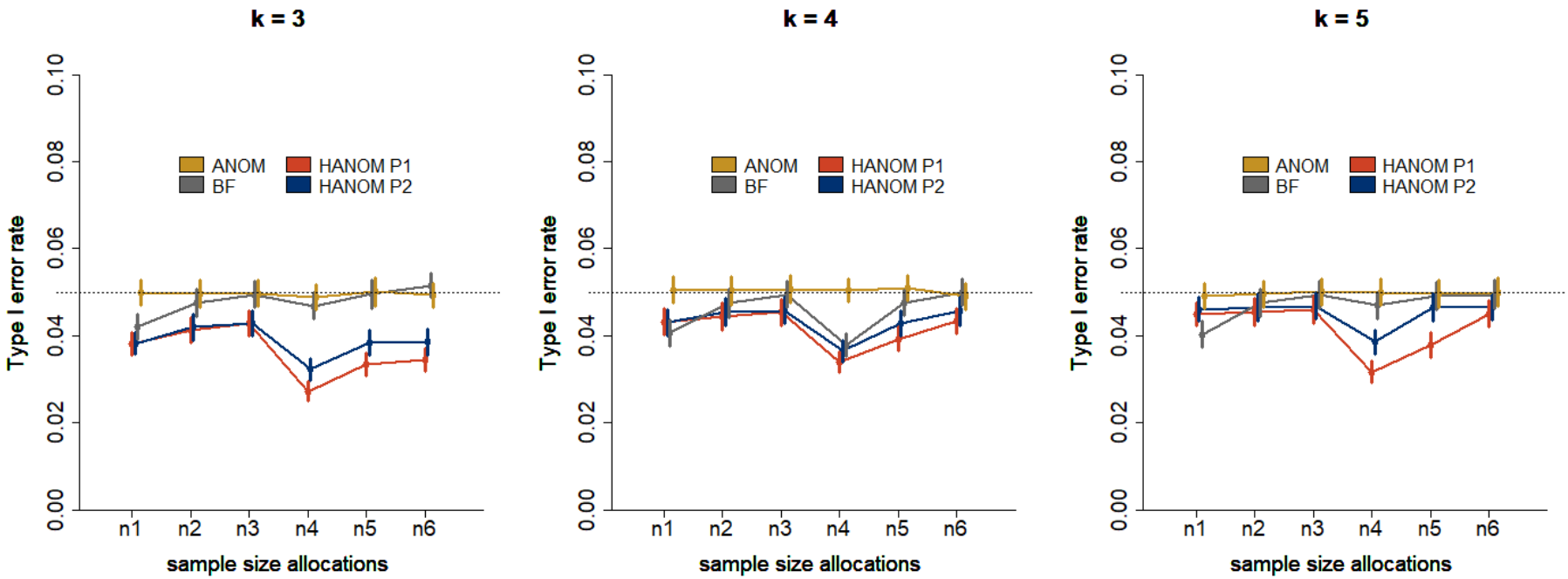
Simulations in Balanced Design



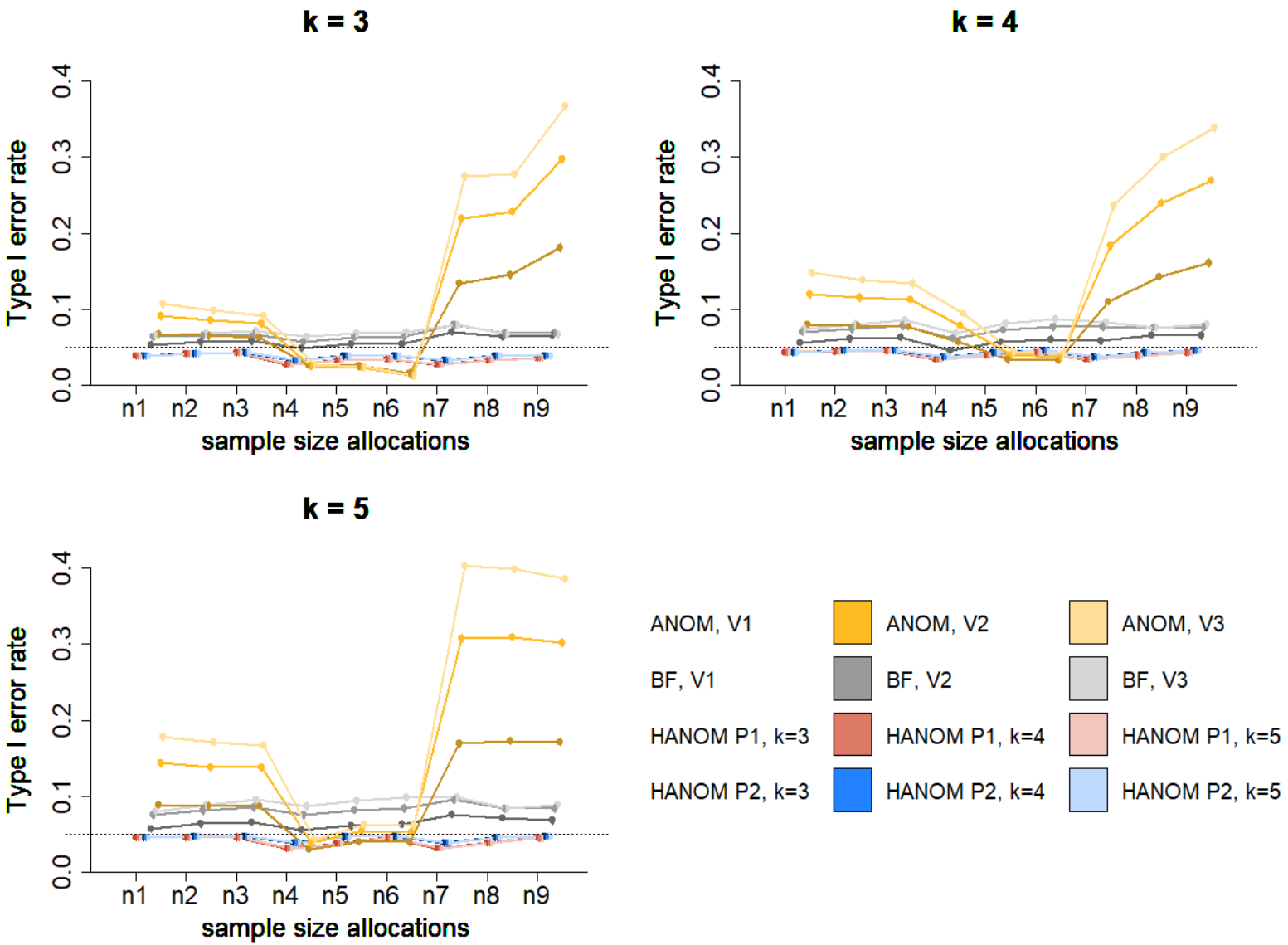
Simulations in Unbalanced Design

	Number of treatment groups		
	$k = 3$	$k = 4$	$k = 5$
Sample sizes			
n1	5:5:5	5:5:5:5	5:5:5:5:5
n2	10:10:10	10:10:10:10	10:10:10:10:10
n3	20:20:20	20:20:20:20	20:20:20:20:20
n4	3:5:8	3:3:5:5	3:5:7:10:15
n5	5:10:15	5:8:10:15	5:10:15:20:25
n6	5:15:25	10:20:30:40	10:20:30:40:50
n7	8:5:3	5:5:3:3	15:10:7:5:3
n8	15:10:5	15:10:8:5	25:20:15:10:5
n9	25:15:5	40:30:20:10	50:40:30:20:10
Variance ratio			
V0	1:1:1	1:1:1:1	1:1:1:1:1
V1	1:1:4	1:1:1:4	1:1:1:1:4
V2	1:1:10	1:1:1:10	1:1:1:1:10
V3	1:1:20	1:1:1:20	1:1:1:1:20

Results of homogeneous variances conditions



Results of heterogeneous variances conditions



Numerical Example 1

Table 1: Effect of Treatments Data from Juneau (2003)

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Responses	99.891	113.408	115.138	104.133
	99.651	111.645	108.581	116.034
	97.486	107.345	109.565	110.294
	101.120	108.763	110.846	93.026
Mean	99.537	110.290	111.033	105.872
SD	1.511	2.742	2.890	9.847

The p-value computed by the one-way ANOVA is 0.038 and the p-value computed by the Brown-Forsythe test is 0.107.

Numerical Example 1

Table 2: Pairwise comparisons test for Scheffé and Games-Howell.

Scheffé Test			
Contrast	Mean Difference	Std. Error	p-value
1 vs 2	–10.753	3.793	0.098
1 vs 3	–11.496	3.793	0.069
1 vs 4	–6.335	3.793	0.456
2 vs 3	–0.742	3.793	0.998
2 vs 4	4.419	3.793	0.772
3 vs 4	5.161	3.793	0.617
Games-Howell Test			
Contrast	Mean Difference	Std. Error	p-value
1 vs 2	–10.753	1.566	**0.005***
1 vs 3	–11.496	1.631	**0.005***
1 vs 4	–6.335	4.981	0.631
2 vs 3	–0.742	1.992	0.981
2 vs 4	4.419	5.111	0.823
3 vs 4	5.161	5.131	0.757

The results are contradicted between the test equality of means and the pairwise comparison differences.

Numerical Example 1

Table 3: Summary Statistics for HANOM

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
$\overline{X}_i$	99.099	110.799	111.095	110.154
S_i	1.325	3.119	3.536	5.952
U_i	0.882	0.485	0.445	0.250
V_i	–1.647	–0.454	–0.336	0.250
$\widetilde{X}_i$	95.534	111.723	111.178	105.872

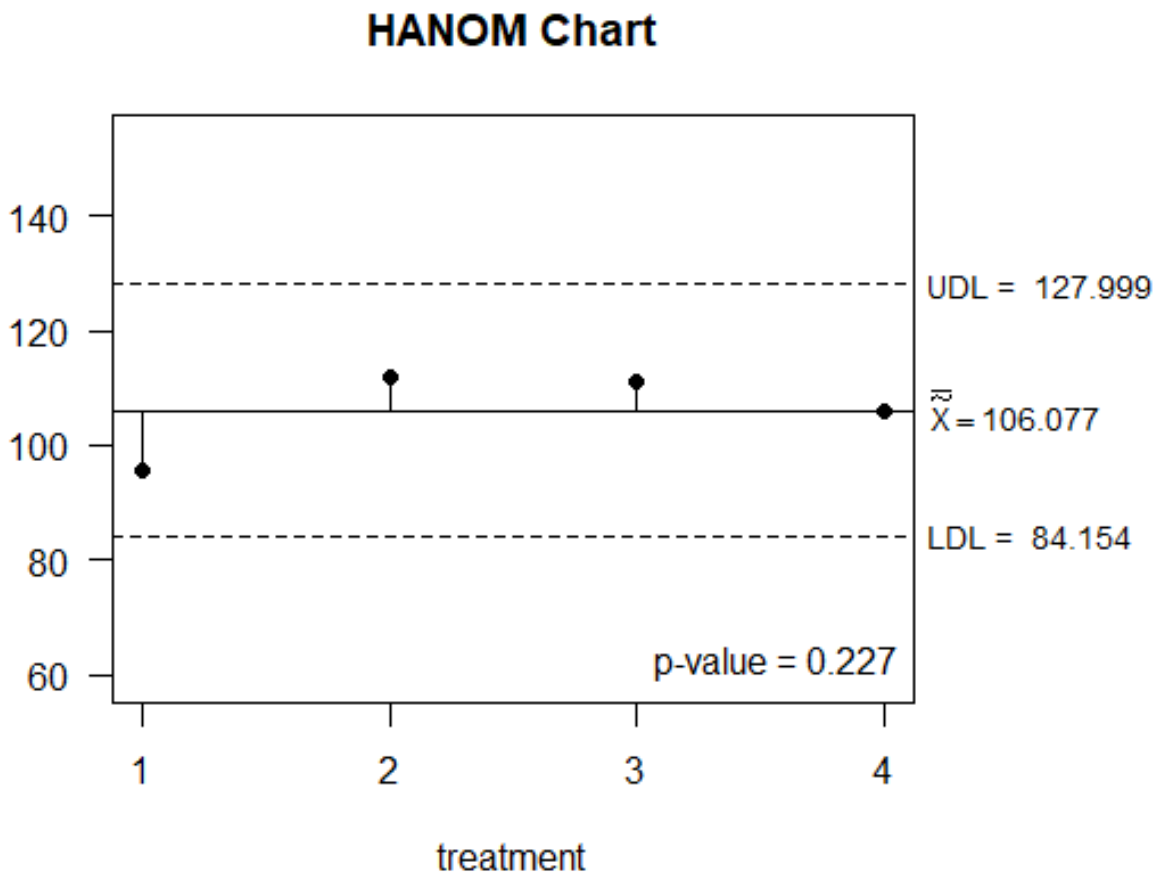
$$\overline{\widetilde{X}} = \frac{95.534 + 111.723 + 111.178 + 105.872}{4} = 106.077$$

$$LDL = 106.077 - 7.367 \times \frac{5.952}{\sqrt{4}} = 84.154$$

$$UDL = 106.077 + 7.367 \times \frac{5.952}{\sqrt{4}} = 127.999$$

The critical value $d^* = d^*(\frac{\alpha}{2} = 0.025, k = 4, df = 2) = 7.367$ obtained from Table A2 or Shiny interface.

The HANOM Chart for Effect of Treatments Data



Numerical Example 2

Table 4: Bacterial Killing Ability Data from Bishop and Dudewicz (1978)

Solvent 1	98.29	95.58	97.39	98.43	95.41	96.87	98.66	95.29	98.3	98.59
	95.28	98.89	97.41	97.52	94.63	96.44				
Solvent 2	98.47	94.67	95.68	97.52	97.52	92.43	96.86	97.57	94.03	97.21
	98.06	98.2	92.09	93.63	94.61	95.28	94.14	93.09	94.47	
Solvent 3	92.29	91.57	94.13	98.05	91.43	93.38	97.55	98.93	94.96	93.61
	92.52	94.2	94.21	92.65	92.68	98.79	92.42	97.48	94.34	96.89
	93.86	95.31	93.76	96.27						
Solvent 4	96.06	96.33	95.9	96.71	98.38	97.97	97.13	97.42	97.65	98.05
	98.11									

Numerical Example 2

Table 5: Summary statistics of procedure P_1 and P_2 for HANOM

$\mathcal{P}_1$	Solvent 1	Solvent 2	Solvent 3	Solvent 4
n_i	16	19	24	11
$\overline{X}_i$	97.103	95.614	94.566	97.160
S_i	1.482	2.109	2.301	0.878
U_i	0.082	0.058	0.042	0.161
V_i	-0.225	-0.045	0.042	-0.606
$\widetilde{X}_i$	97.252	95.666	94.637	96.584
$\mathcal{P}_2$	Solvent 1	Solvent 2	Solvent 3	Solvent 4
n_0	10	10	10	10
$\overline{X}_{i,n_0}$	97.281	96.196	94.590	97.160
S_{i,n_0}	1.396	1.929	2.723	0.878
U_i	0.123	0.091	0.042	0.144
V_i	-0.038	0.010	0.042	-0.440
$\widehat{X}_i$	97.413	96.068	94.637	96.742

The HANOM Chart of procedure P_1 and P_2 for HANOM

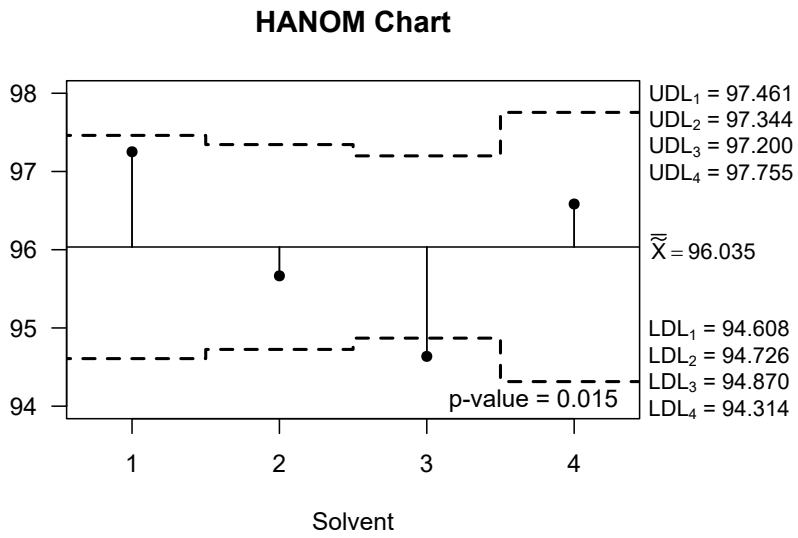


Figure 1: The HANOM Chart of procedure P_1.

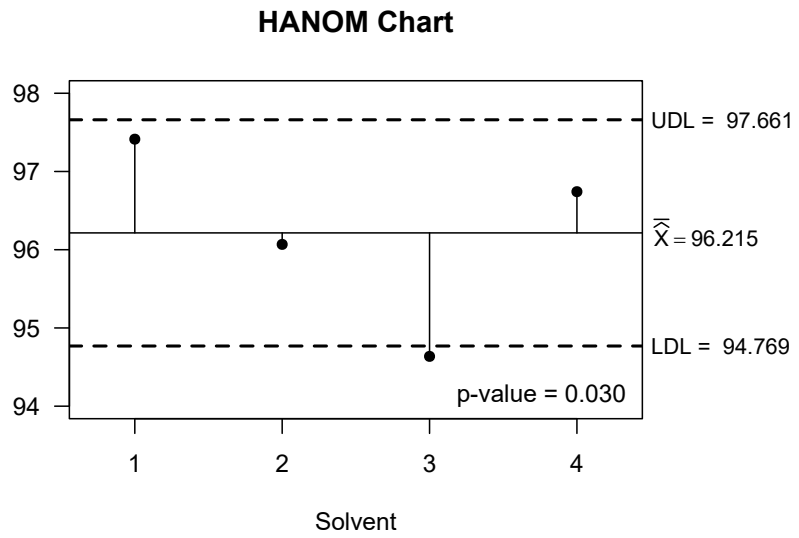


Figure 2: The HANOM Chart of procedure P_2.

Multiple Comparisons with a Control (MCC)

MCC is a multiple comparison procedure for comparing several treatments simultaneously with a control. This research objective is to find simultaneous confidence interval on the $k-1$ paired differences means, $\mu_i - \mu_1, i = 2, \ldots, k$.

- Paulson (1952) and Dunnett (1955) considered the case with equal but unknown variances.
- Under unequal variances, Tamhane (1977) proposed approximated approaches for the multiple comparisons with all-pairwise and a control comparison.
- Hasler and Hothorn (2008) proposed a plug-in approach for the multiple contrast test that provided approximated simultaneous confidence intervals for all-contrast comparisons.
- Li and Ning (2012) proposed three approximate simultaneous confidence intervals procedures for MCC, which were derived from Bonferroni approximation, Slepian's inequality, and multivariate t distribution.

MCC under Heteroscedasticity

- When variances are unknown and unequal, Dudewicz and Dalal (1983) considered a two-stage procedure which was originally proposed by Stein (1945) for comparing all μ_i's with μ_1 to get exact distributions of the test statistics.

- Wen and Chen (1994) used a single-stage sampling procedure to multiple comparison with the best and a control when the sample sizes are equal.

Single-Stage Sampling Procedure for MCC

- For the single-stage sampling procedure for MCC, we randomly select $n_i - 1$ observations and the remaining observation to calculate the weighted sample mean, $\widehat{X}_i = U_i \sum_{j=1}^{n_i-1} X_{ij} + V_i X_{in_i}$, among each population.
- The weights of U_i and V_i of the n_i observations satisfy $(n_i - 1)U_i + V_i = 1$ and $(n_i - 1)U_i^2 + V_i^2 = \frac{z^*}{S_i^2}$, where z^* is the maximum of $\left\{ \frac{S_1^2}{n_1}, \ldots, \frac{S_k^2}{n_k} \right\}$. Thus, we can derive the weights of U_i and V_i as

$$U_i = \frac{1}{n_i} + \frac{1}{n_i} \sqrt{\frac{1}{n_i - 1}[n_i z^* / S_i^2 - 1]}$$
$$V_i = \frac{1}{n_i} - \frac{1}{n_i} \sqrt{(n_i - 1)[n_i z^* / S_i^2 - 1]}.$$

Single-Stage Sampling Procedure for MCC

- The sampling distribution of the weighted sample mean, $\frac{\widehat{X}_i - \mu_i}{\sqrt{z^*}}$, is also an unconditional t distribution with degrees of freedom $n_i - 2$.
- The simultaneous lower and upper confidence intervals for $\mu_i - \mu_1, i = 2, \ldots, k$ are considered as

$$I_L = (\widehat{X}_i - \widehat{X}_1 - a_i\sqrt{z^*}, \infty) \text{ and}$$
$$I_U = (-\infty, \widehat{X}_i - \widehat{X}_1 + b_i\sqrt{z^*})$$

where a_i 's are constants chosen to satisfy the equation

$$\int_{-\infty}^{\infty} \prod_{i=2}^{k} [F_{v_i}(t_1 + a_i)]\, f_{v_1}(t_1) dt_1 = 1 - \alpha.$$

where $F_{v_i}(\cdot)$ and $f_{v_i}(\cdot)$ are the distribution and density functions of a t random variables with $n_i - 2$ degrees of freedom, respectively.

Single-Stage Sampling Procedure for MCC

- The two-sided confidence intervals for $\mu_i - \mu_1, i = 2, \ldots, k$ are as follow

$$I_s = (\widehat{X}_i - \widehat{X}_1 - a_i\sqrt{z^*}, \widehat{X}_i - \widehat{X}_1 + b_i\sqrt{z^*})$$

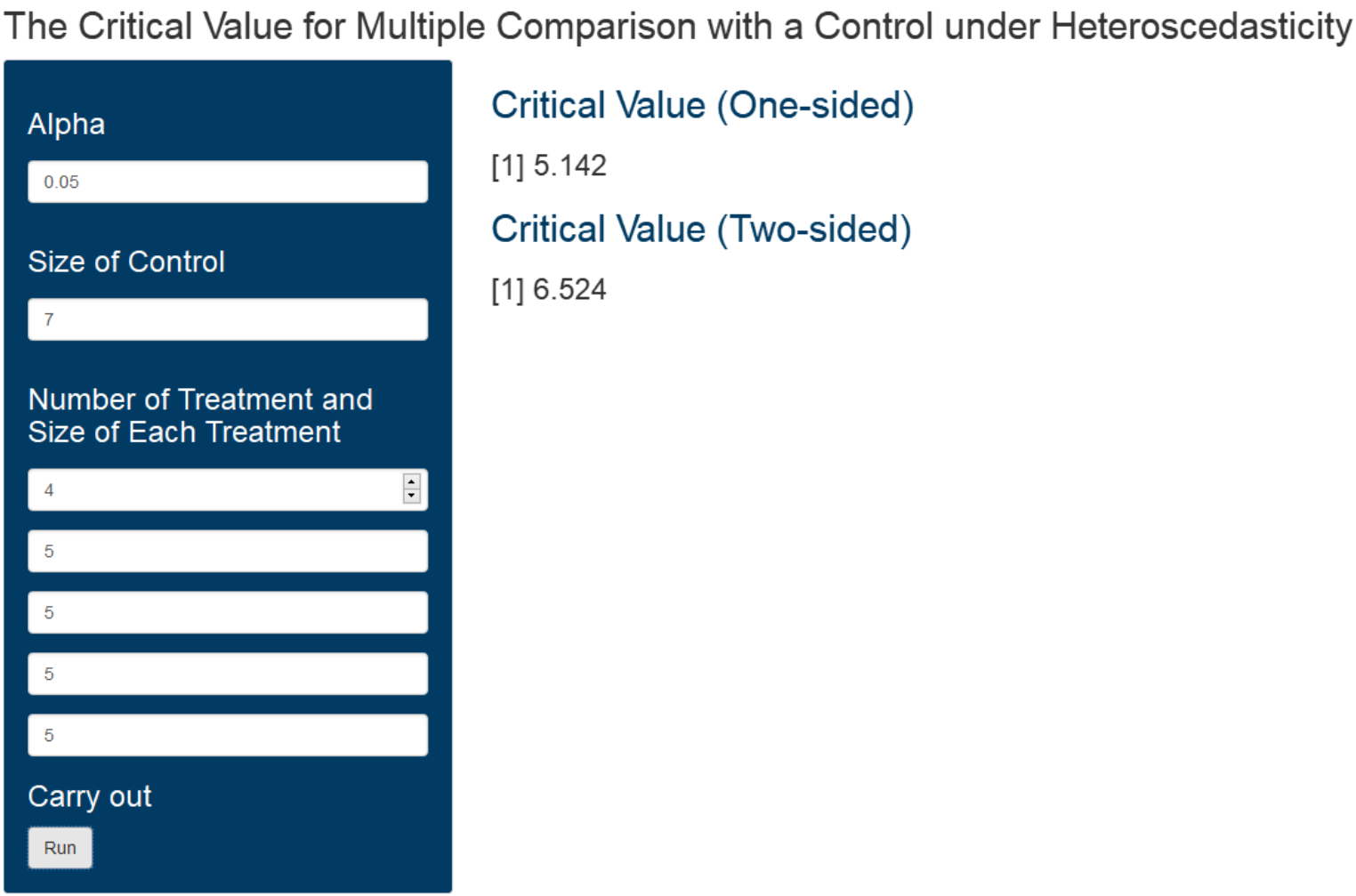
- The critical values a_i's and b_i's are the solution of

$$\int_{-\infty}^{\infty} \prod_{i=2}^{k} \left[F_{v_i}(t_1 + a_i) - F_{v_i}(t_1 - b_i)\right] f_{v_1}(t_1)\, dt_1 = P^*.$$

  where $F_{v_i}(\cdot)$ and $f_{v_i}(\cdot)$ are the distribution and density functions of a t random variables with $n_i - 2$ degrees of freedom, respectively.

- Guenther (1969) and Casella and Berger (2002) have both shown that the shortest length of confidence interval occurs when the critical values $a = -b$ if the distribution is symmetric.

The interface by *R* shiny



The URL of the interface can be accessed at

`https://weiming3524.shinyapps.io/mc-control/`.

Modified Single-Stage Sampling Procedure for MCC

- The weights of U_i and V_i for the modified single-stage sampling procedure for MCC satisfy Equation (2) and (3). As a result, the sampling distribution of the weighted sample mean, $\frac{\widetilde{X}_i - \mu_i}{S_{[k]}/\sqrt{n_i}}$, follows a t distribution with degrees of freedom $n_i - 2$.
- The modified single-stage sampling procedure to the one-sided simultaneous confidence intervals for $\mu_i - \mu_1, i = 2, \ldots, k$ are considered as

$$I_L = (\widetilde{X}_i - \widetilde{X}_1 - c^* \frac{S_{[k]}}{\sqrt{n_i}}, \infty)$$

$$I_U = (-\infty, \widetilde{X}_i - \widetilde{X}_1 + d^* \frac{S_{[k]}}{\sqrt{n_i}})$$

The critical value c^* needs to satisfy

$$\int_{-\infty}^{\infty} \left[\prod_{i=2}^{k} F_{v_i}(t_1 \sqrt{\frac{n_i}{n_1}} + c^*) \right] f_{v_1}(t_1)\, dt_1 = 1 - \alpha.$$

Determination of Sample Size

- The determination of the total sample size for MCC problem depends on the definition of power, such as any-pair power, all-pairs power and per-pair power. We use an any-pair power which can be considered as the minimum required sample size to meet a pre-specified significance and power level for the modified single-stage sampling procedure for MCC.
- The power function for the modified single-stage sampling procedure for MCC is as follows.

$$
\begin{aligned}
& P\{\mu_i - \mu_1 \notin I_L \text{ and } \exists i, i = 2, \ldots, k, \mu_i - \mu_1 > \delta\} \\
=& 1 - P\left(T_i < T_1 \sqrt{\frac{n_i}{n_1}} + \frac{\delta}{S_{[k]}/\sqrt{n_i}}, i = 2, \ldots, k \right) \\
=& 1 - \int_{-\infty}^{\infty} \left(\prod_{i=2}^{k} \left[F_{v_i}(t_1 \sqrt{\frac{n_i}{n_1}} + \frac{\delta}{S_{[k]}/\sqrt{n_i}}) \right] \right) f_{v_1}(t_1)\, dt_1 \geq 1 - \beta
\end{aligned}
$$

Determination of Sample Size

For a given level of significance, α, and a desired power of HANOM, $1-\beta$, the sample size n_i could be determined. We use a grid-searching method to obtain the required sample size as follows:

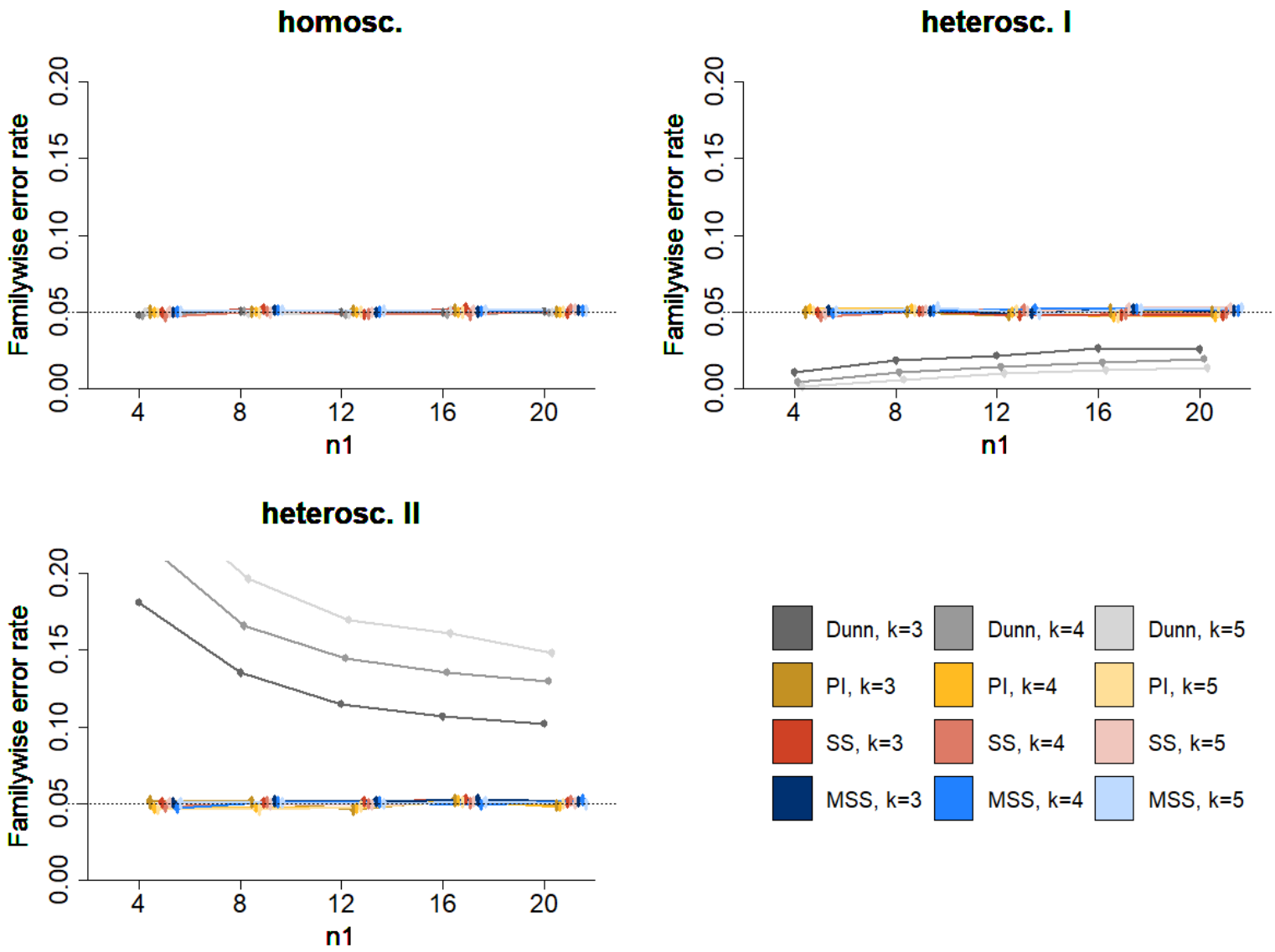
1. Consider an initial sample size $n_{01} = 3$ and $n_{0i} = 3, i = 2, \ldots, k$. A power value for MCC may be determined from the aforementioned power equation by a given largest standard deviation $S_{[k]}$ between treatments, and a difference amount δ between any two comparisons that will lead to rejection of the null hypothesis.
2. If the $1-\beta$ and α for MCC under the initial sample size meet the desired power and the level of significance, α, the required sample size is n_{0i}. If not, to calculate the power equation under $n_i = n_{0i} + 1$ and to check if it meets the desired power and level of significance.
3. Iteratively this process until the power and α values meets the desired level. The sample size n_i is currently the smallest size necessary to ensure the given power and significance level.

Simulations for MCC

Similar to Hasler (2016) scenario setting, the expected value of all treatments was identical, $\mu_i = 100$ $(i = 1, \ldots, k)$. Five sample size allocations were chosen, i.e. $n_1 = 4, 8, 12, 16, 20$, and also, unbalanced sample sizes per allocations, i.e. $n_i = n_{i-1} + 2$ $(i = 2, \ldots, k)$.

- The homoscedastic case: $\sigma_i = 10, i = 1, \ldots, k$.
- The heteroscedastic case I: $\sigma_i = 10 + 40(i-1)/(k-1), i = 1, \ldots, k$.
- The heteroscedastic case II: $\sigma_i = 50 - 40(i-1)/(k-1), i = 1, \ldots, k$.

Simulations for MCC



Numerical Example for MCC

Table 6: Summary statistics of the mutagenicity data set (Adler and Kliesch (1990))

Treatment group	Mean	Standard deviation	Sample size
Vehicle control	2.571	1.272	7
30 mg/kg	3.800	1.095	5
50 mg/kg	6.200	1.483	5
75 mg/kg	14.000	3.937	5
100 mg/kg	20.000	4.062	5

Numerical Example for MCC

Table 7: Summary statistics with MCC for the mutagenicity data

	Treatment groups				
	Vehicle control	30 mg/kg	50 mg/kg	75 mg/kg	100 mg/kg
SS procedure	$I_L = (\widetilde{X}_i - \widetilde{X}_1 - a\sqrt{z^*}, \infty)$, where $a = 4.834$				
n_i	7	5	5	5	5
$\overline{X}_i$	2.167	3.500	5.750	13.000	19.250
S_i	0.753	1.000	1.258	3.742	4.272
U_i	0.530	0.615	0.524	0.255	0.200
V_i	−2.181	−1.461	−1.098	−0.020	0.200
$\widehat{X}_i$	−4.012	1.308	3.280	12.898	20.000
Lower conf. intervals		$(-3.915, \infty)$	$(-1.943, \infty)$	$(7.675, \infty)^*$	$(14.776, \infty)^*$
MSS procedure	$I_L = (\widetilde{X}_i - \widetilde{X}_1 - c^* S_{[k]}/\sqrt{n_i}, \infty)$, where $c^* = 4.635$				
U_i	0.469	0.615	0.524	0.255	0.200
V_i	−1.812	1.461	−1.098	−0.020	0.200
$\widetilde{X}_i$	−2.967	1.308	3.280	12.090	20.000
Lower conf. intervals		$(-4.580, \infty)$	$(-2.608, \infty)$	$(7.010, \infty)^*$	$(14.112, \infty)^*$
PI procedure					
Lower conf. intervals		$(-0.557, \infty)$	$(1.391, \infty)^*$	$(5.495, \infty)^*$	$(11.319, \infty)^*$

Conclusion and Discussion

- Analysis of means under heteroscedasticity (HANOM) can easily identify the mean having a significant difference from the overall mean. The single-stage sampling procedure for HANOM does not require the assumption of equal variances.
- Similarly, the proposed single-stage sampling procedures for multiple comparisons with a control (MCC) do not need to assume homoscedastic variances.
- These procedures can provide an exact distribution for its statistics employed in HANOM and MCC under heteroscedasticity.
- The single-stage sampling procedure is advantageous since it is data-analysis oriented, and works for any given data set whether the variances are equal or unequal.

Conclusion and Discussion

- The simulation results showed that the single-stage sampling procedures for HANOM can maintain robust regardless of homoscedasticity or heteroscedasticity. Under the homogeneity situations, the single-stage sampling procedures for HANOM are conservative in a fewer number of groups and smaller sample sizes situations.
- The single-stage sampling procedures also can maintains the FWE for the MCC.
- We suggested that the sample size n_i for each treatment in the single-stage sampling procedure for HANOM should consist of at least 10 observations as suggested by Bishop and Dudewicz (1978). In fact, we can achieve reasonably decent results in our simulations as long as the number of groups is more than 4 and the size of each treatment is greater than 5 observations.

Future Work

- More efficient for determining sample sizes under different least favorable configurations and power definitions.
- To use and modify the single stage sampling procedures to the correlated populations.
- HANOM in a two-way layout or higher-way model may be developed.
- Consider the arithmetic mean instead of weighted sample mean to modify a two-stage sampling procedure for HANOM and MCC problems.

Thank You

Two-Stage Sampling Procedure for HANOM

The classical two-stage sampling procedure for HANOM that proposed by Nelson and Dudewicz (2002) is briefly described as follows:

1. Take an initial sample n_0 (≥ 2) from each i $(i = 1, \ldots, k)$ population, and calculate the sample means $\overline{X}_{i,n_0}$ and sample variances S^2_{i,n_0}.
2. Specify the significance α, power γ, and the difference δ between any two treatment means that will lead to rejection of the all of treatment means are equal hypothesis. From Figure A1-A36 in Dudewicz and Nelson (2003), given k, α, γ, and degree of freedom $df = n_0 - 1$ combination, find the corresponding value of ω.
3. Calculate

$$n_i = \max\left\{n_0 + 1, [\![\left(\frac{\omega}{\delta}\right)^2 S^2_{i,n_0}]\!] + 1\right\}$$

for i $(i = 1, \ldots, k)$ population, where $[\![x]\!]$ denoted the greatest integer in x.

Two-Stage Sampling Procedure for HANOM

4. Take the remaining $n_i - n_0$ observation to calculate the sample means

$$\overline{X}_{i,n_i-n_0} = \frac{X_{i,n0+1} + \ldots + X_{i,n_i}}{n_i - n_0}$$

5. For each i $(i = 1, \ldots, k)$ population, compute

$$b_i = \frac{n_i - n_0}{n_i}\left[1 + \sqrt{\frac{n_0}{n_i - n_0}\left(\left[\frac{\delta}{\omega}\right]^2 \frac{n_i}{S^2_{i,n_0}} - 1\right)}\right]$$

$$\widetilde{\overline{X}}_i = (1 - b_i)\overline{X}_{i,n_0} + b_i\overline{X}_{i,n_i-n_0}$$

$$\widetilde{\overline{X}} = \frac{\sum_{i=1}^{k} \widetilde{\overline{X}}_i}{k}$$

6. Calculate decision lines of a HANOM:

$$\widetilde{\overline{X}} \pm H(\alpha; k; n_0 - 1)\frac{\delta}{\omega}$$

where $H(\alpha; k; n_0 - 1)$ is found in Table A.1 in Nelson and Dudewicz (2002).

Two-Stage Sampling Procedure for MCC

The classical two-stage sampling procedure for MCC that proposed by Dudewicz and Dalal (1983) is a design-oriented method whose purpose is to obtain simultaneous confidence intervals with a overall confidence coefficient $1-\alpha$. The detailed is as follows:

1. Take an initial sample n_0 (≥ 2) from each i $(i = 1, \ldots, k)$ population, and calculate the sample means $\overline{X}_{i,n_{0i}}$ and sample variances $S^2_{i,n_{0i}}$.
2. Calculate

$$n_i = \max\left\{ n_{0i}+1, [\![\frac{S^2_{i,n_{0i}}}{d^2}]\!] + 1 \right\}$$

for i-th population. Let $d = d(n_{0i}, k, a, 1-\alpha)$ be the solution of the equation

$$\int_{-\infty}^{\infty} \prod_{i=2}^{k} \left[F_{n_{0i}}\left(t_1 + \frac{l}{d} \right) \right] f_{n_{0i}}(t_1)\, dt_1 = 1-\alpha$$

where $F_v(\cdot)$ and $f_v(\cdot)$ are the distribution and density functions respectively of a student's t random variables with $n_{0i}-1$ degrees of freedom.

Two-Stage Sampling Procedure for MCC

3. Take the remaining $n_i - n_{0i}$ $(i = 1, \ldots, k)$ observation to calculate the sample means

$$\overline{X}_{i,n_i-n_{0i}} = \frac{X_{i,n0i+1} + \ldots + X_{i,n_i}}{n_i - n_{0i}}$$

4. For each i $(i = 1, \ldots, k)$ population, compute

$$c_i = \frac{n_{0i}}{n_i}\left[1 + \sqrt{1 - \frac{n_i}{n_{0i}}\left(1 - \frac{n_i - n_{0i}}{(S_{i,n_{0i}}/d)^2}\right)}\right]$$

and

$$\widetilde{\overline{X}}_i = c_i\overline{X}_{i,n_{0i}} + (1 - c_i)\overline{X}_{i,n_i-n_{0i}}$$

5. Calculate the one-sided confidence intervals:

$$\left(\widetilde{\overline{X}}_i - \widetilde{\overline{X}}_1\right) - I \leq \mu_i - \mu_1$$

or

$$\mu_i - \mu_1 \leq \left(\widetilde{\overline{X}}_i - \widetilde{\overline{X}}_1\right) + I$$

Chen (2001), Chen and Wen (2006) as well as Wen et al. (2017) addressed the relative advantages of the single-stage and two-stage sampling procedures. The comparison is concluded as follows:

1. When $S_i^2/n = S_j^2/n$ for all i, j, the single-stage and two-stage sampling procedures have the same decision lines. Under this condition, $\left(\frac{\omega}{\delta}\right)^2 S_{i,n_0}^2$ except for a rounding error in the sample size determined by a classical two-stage sampling procedure is equal to $S_{[k]}^2/n$ in the single-stage sampling procedure, which gives the same sample sizes in addition to the same decision lines.
2. When $\underset{1\le j\le k}{max}\, S_i^2/n < \left(\frac{\omega}{\delta}\right)^2 S_{i,n_0}^2$, the single-stage sampling procedure has a smaller decision line than that of two-stage sampling procedure.
3. When $\underset{1\le j\le k}{min}\, S_i^2/n > \left(\frac{\omega}{\delta}\right)^2 S_{i,n_0}^2$, the single-stage sampling procedure has a larger decision line than that of two-stage sampling procedure.
4. In other situations, the decision lines of the single-stage sampling procedure could be larger than, smaller than or equal to those of the two-stage sampling procedure depending on actual sample data.