Spring 2020

Parametric and Nonparametric Multiple Comparisons in Repeated Measures

Alexander Kuhn
Western Washington University

Parametric and Nonparametric Multiple Comparisons in Repeated Measures

Alexander Kuhn
Department of Mathematics, Western Washington University
June 16, 2020

Abstract: Many experiments in psychology, biology, medicine, etc., result in repeated measures data, i.e., multiple dependent observations over time. Researchers in these fields are often interested in reporting effect sizes; however, there currently is not a one-step procedure to deal with such a scenario. We achieve this through an application of the multivariate delta method, which enables us to derive an effect size generalization of the General Parametric Model (GPM) of Hothorn et al. (2008) which we refer to as the General Parametric Model with Effect Size (GPM-ES). We then utilize the GPM-ES framework to develop a one-step multiple contrast test procedure (MCTP). We demonstrate these methods by working out a real-world example with boys' dental growth data, and discuss how this framework can be applied to the nonparametric multiple comparisons – extending the work of Noguchi et al. (2020) to the case of repeated measures data.

1. Introduction

Repeated measures data typically arise in situations where each subject generates multiple observations under different conditions (e.g., over time). A typical data structure in the one-way setting with n subjects and d repeated measures is displayed below:

Subject\Time	1	2	...	d	Vector Notation
1	$X_{1,1}$	$X_{1,2}$	...	$X_{1,d}$	$\boldsymbol{X}_1$
2	$X_{2,1}$	$X_{2,2}$	...	$X_{2,d}$	$\boldsymbol{X}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$X_{n,1}$	$X_{n,2}$	...	$X_{n,d}$	$\boldsymbol{X}_n$
Parameters	θ_1	θ_2	...	θ_d	$\boldsymbol{\theta}$

Each subject is associated with a d-dimensional independent random vector, $\boldsymbol{X}_1, \ldots \boldsymbol{X}_n$, and there is one parameter of interest, θ_k, for each repeated measure $k = 1, \ldots, d$.

In such data, researchers are interested in comparing the effects of several different conditions,

known as multiple comparisons. Unfortunately, the current popular multiple comparison methods for repeated measures data can result in conflicting results between the overall effect of different conditions and specific pairwise comparisons. That is, two-step procedures such as an ANOVA F-test with an ad-hoc Bonferroni adjustment may provide results that are inconsistent between the two tests, making accurate interpretation of the analysis very difficult. Furthermore, in many of these cases, it is common for key assumptions such as normality of the data to be ignored in practice.

In this report, we set out to develop a one-step multiple contrast testing procedure (MCTP) which places minimal assumptions on the repeated measures data to avoid these issues. In developing an MCTP, we also work to develop a testing procedure with properties known as coherence and consonance. Taken together (informally for now), coherence and consonance are MCTP properties which guarantee that there is a statistically significant overall effect if and only if at least one specific pairwise comparison has a statistically significant effect.

In order to address the repeated measures setting in a way that can have practical significance in fields such as psychology, biology, and medicine, we also develop an MCTP that is built to accommodate various effect sizes. Here, effect size refers to any practically interpretable statistical quantity, i.e., some transformation of our parameters of interest. In particular, many professional organizations including the American Statistical Association (ASA) and American Psychological Association (APA) recommend reporting effect sizes and their confidence intervals instead of just reporting p-values (Wasserstein and Lazar, 2016). This idea is covered in more detail in Section 3.

Working up to the general effect size setting, we first consider specific linear transformations of the parameters of interest through the General Parametric Model (GPM) of Hothorn et al. (2008). Section 2.1 is devoted to the development of the asymptotic theory of the GPM. Next, we generalize

the asymptotic result via the multivariate delta method to accommodate various effect sizes via the GPM with Effect Size (GPM-ES) in Section 3.1. Then, in Section 4.1, we develop an MCTP with the GPM-ES by way of the maxT procedure, as well as the derivation of simultaneous confidence intervals (SCIs) for generalized effect sizes. Moreover, Section 4.3 explains how the GPM-ES can be applied to the nonparametric setting by using the relative effect, a measure of stochastic superiority between treatments.

Throughout the next three chapters (in particular, Sections 2.2, 3.2, and 4.2), we perform an analysis of a dental dataset from Potthoff and Roy (1964) to showcase an application of the GPM-ES to the multiple comparisons for repeated measures setting. The dataset is very well known as an example of growth curve data. Using the dataset, we investigate the mean distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure of 16 boys at ages 8, 10, 12, and 14.

2. General Parametric Model (GPM)

In this chapter, we derive some asymptotic results necessary in the subsequent chapters. First, we describe the General Parametric Model (GPM) of Hothorn et al. (2008) along with its asymptotic properties. Then, we discuss Theorem 1 which details the asymptotic distribution of our vector of test statistics. Lastly, we show a practical application of the GPM with the dental dataset.

2.1 GPM of Hothorn et al. (2008)

Recall our data structure under consideration:

Subject\Time	1	2	...	d	Vector Notation
1	$X_{1,1}$	$X_{1,2}$	...	$X_{1,d}$	$\boldsymbol{X}_1$
2	$X_{2,1}$	$X_{2,2}$	...	$X_{2,d}$	$\boldsymbol{X}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$X_{n,1}$	$X_{n,2}$	...	$X_{n,d}$	$\boldsymbol{X}_n$
Parameters	θ_1	θ_2	...	θ_d	$\boldsymbol{\theta}$

Let $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ be a collection of independent and identically distributed random vectors where each $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})$ for $i = 1, \ldots, n$ consists of random variables, $X_{i,k}$, for $k = 1, \ldots, d$. Here, n refers to the number of subjects (i.e., n could be the number of individuals participating in a study) and d is the number of observations per subject (such as the number of time points data collected). Note that we do not place any distributional assumptions on $\boldsymbol{X}_i$. However, we place some assumptions on the estimators for the asymptotic convergence.

Let $\mathcal{M}((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n), \boldsymbol{\theta}, \boldsymbol{\eta})$ denote a semi-parametric statistical model. The set of n subjects is described by $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. The model contains fixed but unknown elemental parameters $\boldsymbol{\theta} \in \mathbb{R}^d$, and other random or nuisance parameters $\boldsymbol{\eta}$. We are primarily interested in the linear functions $\boldsymbol{C\theta}$ of the parameter vector $\boldsymbol{\theta}$ as specified through the contrast matrix $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q]'$, where each $\boldsymbol{c}_\ell,\ \ell = 1, \ldots, q$, is a d-dimensional contrast vector satisfying

$$\sum_{i=1}^{d} \boldsymbol{c}_{\ell_i} = 0,$$

where $\boldsymbol{c}_{\ell_i}$ is the i-th element of $\boldsymbol{c}_\ell$. Note these are row vectors.

That is, $\boldsymbol{C}$ is a $q \times d$ contrast matrix whose row entries sum to zero. In what follows, we describe the underlying model assumptions, the asymptotic distribution of estimates of our parameters of interest $\boldsymbol{C\theta}$, as well as the corresponding test statistics for hypotheses about $\boldsymbol{C\theta}$ and their asymptotic joint distribution[1].

To describe the aforementioned properties, let $\hat{\boldsymbol{\theta}}_n := \hat{\boldsymbol{\theta}}$ denote the sample estimate of $\boldsymbol{\theta}$ for the case of n subjects. In the General Parametric Model (GPM) of Hothorn et al. (2008), which is applicable

[1]A brief note on notation: We use "$\xrightarrow{d}$" to denote convergence in distribution, "$\overset{\text{a}}{\sim}$" to denote the large-sample approximate or asymptotic distribution of a random variable/vector, and "$\xrightarrow{p}$" to denote convergence in probability.

to a large class of parameters, the sample covariance matrix of $\hat{\boldsymbol{\theta}}_n$, denoted by $\boldsymbol{S}_n := \widehat{\text{Cov}(\hat{\boldsymbol{\theta}}_n)} \in \mathbb{R}^{d,d}$, has the property

$$a_n \boldsymbol{S}_n \xrightarrow{p} \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ denotes an appropriate covariance matrix, for some positive nondecreasing sequence $\{a_n\}$. Furthermore, it has the asymptotic normality property given by

$$a_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ denotes the d-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. When these two aforementioned properties hold, we say

$$\hat{\boldsymbol{\theta}}_n \overset{\text{a}}{\sim} \mathcal{N}_d(\boldsymbol{\theta}, \boldsymbol{S}_n).$$

Let $\boldsymbol{S}_n^\star := \boldsymbol{C}\boldsymbol{S}_n\boldsymbol{C}' \in \mathbb{R}^{q,q}$. Then, by linearity, for the same positive nondecreasing sequence $\{a_n\}$,

$$a_n \boldsymbol{S}_n^\star = \boldsymbol{C}(a_n \boldsymbol{S}_n)\boldsymbol{C}' \xrightarrow{p} \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}' := \boldsymbol{\Sigma}^\star$$

and

$$a_n^{1/2}\boldsymbol{C}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}^\star)$$

hold. In other words,

$$\boldsymbol{C}\hat{\boldsymbol{\theta}}_n \overset{\text{a}}{\sim} \mathcal{N}_q(\boldsymbol{C}\boldsymbol{\theta}, \boldsymbol{S}_n^\star),$$

which nicely describes the asymptotic joint distribution of $\boldsymbol{C}\hat{\boldsymbol{\theta}}_n$.

Now, under the above assumptions, we can derive the asymptotic joint distribution of the vector of test statistics which we refer to as $\boldsymbol{T}_n$. We need a joint distribution for our test statistics in order to perform a proper multiple contrast testing procedure (MCTP). Hothorn et al. (2008) show the following theorem, which we have proven in detail (see Appendix A):

Theorem 1

Let $\hat{\boldsymbol{\theta}}_n \overset{a}{\sim} \mathcal{N}_d(\boldsymbol{\theta}, \boldsymbol{S}_n)$. *Then, the asymptotic joint distribution of the vector of test statistics* $\boldsymbol{T}_n := \boldsymbol{D}_n^{-1/2}\boldsymbol{C}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ *is given by*

$$\boldsymbol{T}_n \overset{\text{a}}{\sim} \mathcal{N}_q(\mathbf{0}, \boldsymbol{R}_n^\star),$$

where $\boldsymbol{D}_n = \text{diag}(\boldsymbol{S}_n^\star) \in \mathbb{R}^{q,q}$ *is a diagonal matrix which contains the sample variances associated with* $\boldsymbol{S}_n^\star$. *Moreover,* $\boldsymbol{R}_n^\star = \boldsymbol{D}_n^{-1/2}\boldsymbol{S}_n^\star\boldsymbol{D}_n^{-1/2} \in \mathbb{R}^{q,q}$ *is the sample correlation matrix with* 1*'s along the diagonal entries and sample correlation coefficients in the off-diagonals.*

The above theorem implies that $\boldsymbol{T}_n$ is approximately multivariate normal with mean vector $\mathbf{0}$ and sample covariance matrix $\boldsymbol{R}_n^\star$ for a sufficiently large n. Also, notice that these results imply that the asymptotic marginal distribution for each $T_n^\ell \in \boldsymbol{T}_n$, $\ell = 1, \ldots, q$, is the standard normal distribution, $N(0, 1)$, as the diagonal entries of $\boldsymbol{R}_n^\star$ (and thus the variances of each T_n^ℓ for $\ell = 1, \ldots, q$) contain only 1's. In other words, $\boldsymbol{R}_n^\star$ is also a sample correlation matrix.

2.2 GPM Example

In this section, we formally introduce our dataset under consideration. The dataset, which comes from a dental study, contains growth measurements of the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure of 11 girls and 16 boys at ages 8, 10, 12, and 14 (Potthoff and Roy, 1964). Since our method is designed for a homogeneous group of subjects, we consider the male data only. Furthermore, we use the square root of the dental growth measure to improve normality of the data and to stabilize variance. Although the square-root transformation is not relevant for the asymptotic arguments, approximate normality and homogeneity of variance tend to provide more reliable inference when performing small sample approximations with the multivariate t-distribution. Box plots of the original boys data and the square root of the boys data are shown in Figure 1 below.

Our parameters of interest are the means of the square-rooted distances (mm) from the center of the pituitary to the pteryomaxillary fissure at each time point, i.e., at ages 8, 10, 12, and 14. We refer to this measure as the root depth for brevity. That is, we have $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)' = (\mu_1, \mu_2, \mu_3, \mu_4)' := \boldsymbol{\mu}$. Suppose we are interested in comparing the mean root depth between ages 8 and 10, ages 10 and 12, and ages 12 and 14. Then, we can construct a contrast matrix

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix},$$

with corresponding contrast vectors

$$\boldsymbol{c}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{c}_2 = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}, \text{ and } \boldsymbol{c}_3 = \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix}.$$

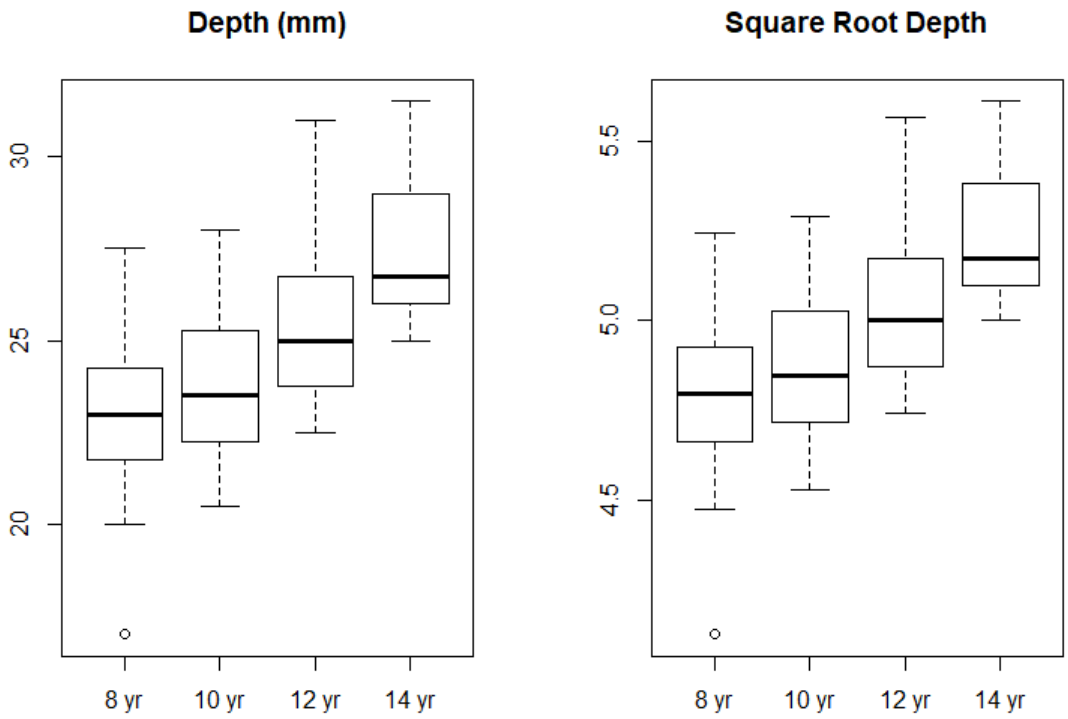

Figure 1: Box plots of the measurements of the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure for 16 boys at ages 8, 10, 12, and 14. Original data left, square root transform right.

Our choice of the contrast matrix above is motivated by a change-point analysis for detecting the age at which the root depth significantly differ by examining the adjacent time points sequentially.

To perform the analysis, we are interested in the linear transformation of $\boldsymbol{\mu}$ given by

$$\boldsymbol{C\mu} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \\ \mu_3 - \mu_4 \end{bmatrix}.$$

Of course, in practice, we do not know the true values of $\boldsymbol{\theta} = \boldsymbol{\mu}$ and must instead work with the sample mean vector $\overline{\boldsymbol{X}}_n = (\overline{X}_1, \overline{X}_2, \overline{X}_3, \overline{X}_4)'$ with $n = 16$. Note that $\overline{\boldsymbol{X}}_n$ is a special case of $\hat{\boldsymbol{\theta}}_n$ from the general setting. We are thus interested in the transformed observed values

$$\boldsymbol{C}\overline{\boldsymbol{x}}_n = \begin{bmatrix} \overline{x}_{n1} - \overline{x}_{n2} \\ \overline{x}_{n2} - \overline{x}_{n3} \\ \overline{x}_{n3} - \overline{x}_{n4} \end{bmatrix} = \begin{bmatrix} -0.09914503 \\ -0.19001352 \\ -0.17231448 \end{bmatrix}.$$

In order to use the GPM, we need $\boldsymbol{S}_n = \widehat{\text{Cov}(\overline{\boldsymbol{x}}_n)}$, the sample covariance matrix of the sample means. That is, we must calculate $s_{ij} = \widehat{\text{Cov}(\overline{x}_{ni}, \overline{x}_{nj})}$ for $i, j = 1, 2, 3, 4$. From R, we get

$$\boldsymbol{S}_n = \begin{bmatrix} 0.0042429 & 0.0014164 & 0.0022421 & 0.0008989 \\ 0.0014164 & 0.0029504 & 0.0013291 & 0.0016919 \\ 0.0022421 & 0.0013291 & 0.0041111 & 0.0018548 \\ 0.0008989 & 0.0016919 & 0.0018548 & 0.0024221 \end{bmatrix}.$$

While dependence of data is clear from the fact that multiple data points are taken from the same individual over time, we can tangibly see that each time point has a nonzero observed covariance with the others. Therefore, clearly, our vectors of observations for each time point are correlated

with one another. Furthermore,

$$\boldsymbol{S}_n^{\star} = \boldsymbol{C}\boldsymbol{S}_n\boldsymbol{C}' = \begin{bmatrix} 0.0043605 & -0.0024470 & 0.0017060 \\ -0.0024470 & 0.0044033 & -0.0026191 \\ 0.0017060 & -0.0026191 & 0.0028236 \end{bmatrix},$$

and

$$\boldsymbol{D_n} = \text{diag}(\boldsymbol{S}_n^{\star}) = \begin{bmatrix} 0.0043605 & 0 & 0 \\ 0 & 0.0044033 & 0 \\ 0 & 0 & 0.0028236 \end{bmatrix}.$$

To see how the computations above are related to Theorem 1, we expect the vector of test statistics to have the asymptotic distribution of $\boldsymbol{T}_n \overset{\text{a}}{\sim} \mathcal{N}_3(\boldsymbol{0}, \boldsymbol{R}_n^{\star})$, where

$$\boldsymbol{R}_n^{\star} = \boldsymbol{D}_n^{-1/2}\boldsymbol{S}_n^{\star}\boldsymbol{D}_n^{-1/2} = \begin{bmatrix} 1 & -0.5584342 & 0.4861965 \\ -0.5584342 & 1 & -0.7427891 \\ 0.4861965 & -0.7427891 & 1 \end{bmatrix}.$$

Notice that we have 1's along the diagonal entries of the correlation matrix $\boldsymbol{R}_n^{\star}$ as was expected.

Currently, there is not much to conclude from these calculations as these are a necessary mid-step in performing a hypothesis test later on. In Section 3.1, we consider a generalized setting where the test statistics and the associated SCIs are expressed in terms of appropriate effect sizes rather than conventional linear combinations of the parameter values associated with $\boldsymbol{C}$ used in the GPM. We then demonstrate its application using the dental study example of Potthoff and Roy (1964) in Section 3.2.

3. GPM with Effect Size (GPM-ES)

We now turn our attention to the use of effect sizes, which serve as measures of clinical or practical significance. Recall that effect size refers to any practically interpretable statistical quantity. In this chapter, we proceed in a similar fashion as the development of the GPM, working up to Theorem 2 which describes the asymptotic distribution of the vector of test statistics with an effect size transformation. We then revisit our dental dataset with the GPM-ES.

3.1 GPM-ES

Suppose we are interested in the arbitrary parameters θ_1 and θ_2, corresponding to two groups of interest. Then, meaningful effect sizes often take the form $g(\hat{\theta}_1) - g(\hat{\theta}_2)$, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of θ_1 and θ_2, and g is a strictly increasing and continuously differentiable function. Some examples for the two-sample case are as follows:

1. Cohen's d: $(\bar{X}_1 - \bar{X}_2)/\sigma = \bar{X}_1/\sigma - \bar{X}_2/\sigma$,
   where $\bar{X}_1$ and $\bar{X}_2$ are the sample means with known standard deviation, σ,

2. Log odds: $\log\left[\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)}\right] = \log[\hat{p}_1/(1-\hat{p}_1)] - \log[\hat{p}_2/(1-\hat{p}_2)]$,
   where $\hat{p}_1$ and $\hat{p}_2$ are the respective sample proportions, and

3. Log ratio: $\log(S_1^2/S_2^2) = \log S_1^2 - \log S_2^2$, where S_1^2 and S_2^2 are the respective sample variances.

However, researchers are often interested in hypotheses involving more than just two groups for comparison. That is, in general, hypotheses can be expressed by rewriting the contrast vector $\boldsymbol{c}_\ell$ as $\boldsymbol{c}_\ell = \boldsymbol{c}_{\ell,1} - \boldsymbol{c}_{\ell,2}$, where

$$(\boldsymbol{c}_{\ell,1})_j := c_{\ell,1,j} = \max\{(\boldsymbol{c}_\ell)_j, 0\} \text{ (positive elements), and}$$
$$(\boldsymbol{c}_{\ell,2})_j := c_{\ell,2,j} = -\min\{(\boldsymbol{c}_\ell)_j, 0\} \text{ (negative elements).}$$

Thus, we can express generalized effect sizes as $g_\ell(\boldsymbol{x}) := g(\boldsymbol{c}'_{\ell,1}\boldsymbol{x}) - g(\boldsymbol{c}'_{\ell,2}\boldsymbol{x}),\ \ell = 1, \ldots, q.$

Recall that, under the GPM, we have assumed $\sum_{i=1}^{d} \boldsymbol{c}_{\ell_i} = 0$ for each $\boldsymbol{c}_\ell$. Here, we assume further that, for each $\boldsymbol{c}_{\boldsymbol{\ell}}$, we have

$$\sum_{i=1}^{d} \boldsymbol{c}_{\ell_i} = 0 \text{ and } \sum_{i=1}^{d} |\boldsymbol{c}_{\ell_i}| = 2,$$

as is recommended in Noguchi et al. (2020). Under the above constraints, we can write

$$\sum_{i=1}^{d} c_{\ell,1,i} = \sum_{i=1}^{d} c_{\ell,2,i} = 1,$$

and thus we can conveniently interpret both $\boldsymbol{c}_{\ell,1}\boldsymbol{\theta}$ and $\boldsymbol{c}_{\ell,1}\boldsymbol{\theta}$ as weighted averages of the entries of $\boldsymbol{\theta}$. More importantly, assuming that $\theta_k \in (a, b)$, for $k = 1, .., d$, these assumptions ensure that $\boldsymbol{c}_{\ell,1}\boldsymbol{\theta} \in (a, b)$ and $\boldsymbol{c}_{\ell,2}\boldsymbol{\theta} \in (a, b)$, implying that the generalization works for any strictly increasing and continuously differentiable g whose domain is (a, b).

Now, let $\boldsymbol{\theta}^g = [g_1(\boldsymbol{\theta}), \ldots, g_q(\boldsymbol{\theta})]'$ and $\hat{\boldsymbol{\theta}}^g_n = [g_1(\hat{\boldsymbol{\theta}}_n), \ldots, g_q(\hat{\boldsymbol{\theta}}_n)]'$, where $g_\ell(\boldsymbol{x})$ is our generalized effect size, as defined above. To make meaningful statements about our new parameter of interest, $\boldsymbol{\theta}^g$, it is necessary to determine the asymptotic joint distribution of the vector of the g-transformed test statistics, $\boldsymbol{T}^g_n$. The effect size generalization can be achieved by use of the multivariate delta method. We arrive at the following result (see Appendix B for proof).

Theorem 2

Let $\hat{\boldsymbol{\theta}}_n \overset{a}{\sim} \mathcal{N}_d(\boldsymbol{\theta}, \boldsymbol{S}_n)$. *Then,* $\hat{\boldsymbol{\theta}}_n^g \overset{a}{\sim} \mathcal{N}_q(\boldsymbol{\theta}^g, \boldsymbol{S}_n^{\star g})$, *where* $\boldsymbol{S}_n^{\star g} = (\hat{\boldsymbol{\Phi}}^g)\boldsymbol{S}_n(\hat{\boldsymbol{\Phi}}^g)'$ *and* $\hat{\boldsymbol{\Phi}}^g = \left[\frac{\partial g_\ell}{\partial \theta_k}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}\right]_{q\times d}$. *Moreover the asymptotic joint distribution of the vector of g-transformed test statistics* $\boldsymbol{T}_n^g := (\boldsymbol{D}_n^g)^{-1/2}(\hat{\boldsymbol{\theta}}_n^g - \boldsymbol{\theta}^g)$ *is given by*

$$\boldsymbol{T}_n^g \overset{\text{a}}{\sim} \mathcal{N}_q(\mathbf{0}, \boldsymbol{R}_n^{\star g}),$$

where $\boldsymbol{D}_n^g = \text{diag}(\boldsymbol{S}_n^{\star g})$ *and* $\boldsymbol{R}_n^{\star g} = (\boldsymbol{D}_n^g)^{-1/2}\boldsymbol{S}_n^{\star g}(\boldsymbol{D}_n^g)^{-1/2}$.

Similar to Theorem 1, the above theorem implies that $\boldsymbol{T}_n^g$ is approximately multivariate normal with mean vector $\mathbf{0}$ and sample covariance matrix $\boldsymbol{R}_n^{\star g}$ for a sufficiently large n. Also, the asymptotic marginal distribution for each $T_n^{g,\ell} \in \boldsymbol{T}_n^g$ is $N(0,1)$, as the diagonal entries of $\boldsymbol{R}_n^{\star g}$ contain only 1's.

The above formulation, which we refer to as GPM-ES, enables us to design meaningful testing procedures and simultaneous confidence intervals (SCIs) around the g-transformed parameters of interest. For example, asymptotic SCIs of effect sizes can be constructed directly. Note that the GPM-ES with the function $g(x) = x$ is equivalent to the GPM.

3.2 GPM-ES Example

Continuing our dental example from Section 2.2, we now use the idea of generalized effect size. That is, we wish to obtain the log-ratio of the mean root depth of successive time points, for the sake of example, since our data is all positive-valued (we could use Cohen's d or the log-odds transformation instead, but Cohen's d would require assuming a known standard deviation and

log-odds is more appropriate when working with proportions).

Consider the function $g(x) = \log(x)$. The log transformation is a natural choice as it is a real-valued function defined on $\mathbb{R}$. To illustrate how the calculation of generalized effect sizes work, we can rewrite our contrast vectors as the difference of two positively valued row vectors:

$$\begin{aligned}
\boldsymbol{c}_1 &= \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \\
&:= \boldsymbol{c}_{1,1} - \boldsymbol{c}_{1,2}, \\
\boldsymbol{c}_2 &= \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \\
&:= \boldsymbol{c}_{2,1} - \boldsymbol{c}_{2,2}, \\
\boldsymbol{c}_3 &= \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \\
&:= \boldsymbol{c}_{3,1} - \boldsymbol{c}_{3,2}.
\end{aligned}$$

Next, we can write our generalized effect sizes for $\overline{\boldsymbol{x}}_n = (\overline{x}_{n1}, \overline{x}_{n2}, \overline{x}_{n3}, \overline{x}_{n4})'$ as

$$\begin{aligned}
g_1(\overline{\boldsymbol{x}}_n) &= g(\boldsymbol{c}_{1,1}\overline{\boldsymbol{x}}_n) - g(\boldsymbol{c}_{1,2}\overline{\boldsymbol{x}}_n) \\
&= \log(\overline{x}_{n1}) - \log(\overline{x}_{n2}), \\
&= -0.02054595
\end{aligned}$$

$$\begin{aligned}
g_2(\overline{\boldsymbol{x}}_n) &= g(\boldsymbol{c}_{2,1}\overline{\boldsymbol{x}}_n) - g(\boldsymbol{c}_{2,2}\overline{\boldsymbol{x}}_n) \\
&= \log(\overline{x}_{n2}) - \log(\overline{x}_{n3}), \\
&= -0.03823464
\end{aligned}$$

and

$$\begin{aligned} g_3(\overline{\boldsymbol{x}}_n) &= g(\boldsymbol{c}_{3,1}\overline{\boldsymbol{x}}_n) - g(\boldsymbol{c}_{3,2}\overline{\boldsymbol{x}}_n) \\ &= \log(\overline{x}_{n3}) - \log(\overline{x}_{n4}), \\ &= -0.03345289 \end{aligned}$$

so that

$$\hat{\boldsymbol{\theta}}_n^g = \overline{\boldsymbol{x}}_n^g = \begin{bmatrix} g_1(\overline{\boldsymbol{x}}_n) \\ g_2(\overline{\boldsymbol{x}}_n) \\ g_3(\overline{\boldsymbol{x}}_n) \end{bmatrix} = \begin{bmatrix} -0.02054595 \\ -0.03823464 \\ -0.03345289 \end{bmatrix}$$

is our consistent estimate of

$$\boldsymbol{\theta}^g = \boldsymbol{\mu}^g = \begin{bmatrix} g_1(\boldsymbol{\mu}) \\ g_2(\boldsymbol{\mu}) \\ g_3(\boldsymbol{\mu}) \end{bmatrix} = \begin{bmatrix} \log(\mu_1) - \log(\mu_2) \\ \log(\mu_2) - \log(\mu_3) \\ \log(\mu_3) - \log(\mu_4) \end{bmatrix} = \begin{bmatrix} \log(\mu_1/\mu_2) \\ \log(\mu_2/\mu_3) \\ \log(\mu_3/\mu_4) \end{bmatrix}.$$

Note that we have a vector of log-ratios between the mean root depths at each successive time point.

Next, in order to utilize Theorem 2, we need to calculate our transformation matrix

$$\begin{aligned} \hat{\Phi}^g = \left[\frac{\partial g_\ell}{\partial \mu_k} \Big|_{\boldsymbol{\mu}=\overline{\boldsymbol{x}}_n} \right]_{3\times 4} &= \begin{bmatrix} 1/\overline{x}_{n1} & -1/\overline{x}_{n2} & 0 & 0 \\ 0 & 1/\overline{x}_{n2} & -1/\overline{x}_{n3} & 0 \\ 0 & 0 & 1/\overline{x}_{n3} & -1/\overline{x}_{n4} \end{bmatrix} \\ &= \begin{bmatrix} 0.2093748 & -0.2051169 & 0 & 0 \\ 0 & 0.2051169 & -0.1974224 & 0 \\ 0 & 0 & 0.1974224 & -0.1909273 \end{bmatrix}. \end{aligned}$$

Then, our transformed observed covariance matrix is

$$\boldsymbol{S}_n^{\star g} = (\hat{\Phi}^g)\boldsymbol{S}_n^{\star}(\hat{\Phi}^g)' = \begin{bmatrix} 0.0001884728 & -0.0001021577 & 0.0000691815 \\ -0.0001021577 & 0.0001767203 & -0.0001027568 \\ 0.0000691815 & -0.0001027568 & 0.0001086982 \end{bmatrix},$$

and its diagonal matrix is

$$\boldsymbol{D}_n^g = \begin{bmatrix} 0.0001884728 & 0 & 0 \\ 0 & 0.0001767203 & 0 \\ 0 & 0 & 0.0001086982 \end{bmatrix}.$$

Finally, by Theorem 2, we have $\boldsymbol{T}_n^g \overset{\text{a}}{\sim} \mathcal{N}_3(\boldsymbol{0}, \boldsymbol{R}_n^{\star g})$, where

$$\boldsymbol{R}_n^{\star g} = (\boldsymbol{D}_n^g)^{-1/2} \boldsymbol{S}_n^{* g} (\boldsymbol{D}_n^g)^{-1/2} = \begin{bmatrix} 1 & -0.5597620 & 0.4833417 \\ -0.5597620 & 1 & -0.7414057 \\ 0.4833417 & -0.7414057 & 1 \end{bmatrix}$$

is the transformed sample correlation matrix.

Next, we develop an MCTP for the GPM-ES as well as its corresponding simultaneous confidence intervals (SCIs). In Section 4.2, we return to the dental study example to perform inference on the given data and report SCIs for the log-ratios of mean root depths.

4. MCTP with GPM-ES

This chapter details the development of a multiple contrast test procedure (MCTP) utilizing the GPM-ES. We begin by defining the global and component null hypotheses to examine their relationship. Then, we discuss the desirable testing properties the MCTP achieves by use of the so-called maxT procedure. Furthermore, we derive simultaneous confidence intervals (SCIs) for the generalized effect size in the GPM-ES. Next, we perform a hypothesis test on the dental dataset, completing our running example. Finally, we address the issue of using the MCTP in the nonparametric setting by defining the relative effect.

4.1 Multiple Contrast Testing Procedure (MCTP)

A typical setting in multiple comparisons is to have a null hypothesis of the form

$$H_0 : \boldsymbol{C\theta} = \mathbf{0},$$

which is usually associated with its one- or two-tailed alternative hypothesis. That is, each row of H_0 corresponds to some linear combination of the parameters of interest. Let us generalize H_0 to formulate a null hypothesis which takes into account the generalized effect size.

To have a null hypothesis which is compatible with the GPM-ES, we define a global hypothesis:

$$H_0^g : \boldsymbol{\theta}^g = \mathbf{0},$$

and the corresponding component hypotheses:

$$H_0^{g,\ell} : g_\ell(\boldsymbol{\theta}) = 0,\ \ell = 1, \ldots, q.$$

It is immediate to show that, when $g(x) = x$, we have $H_0^g : \boldsymbol{C\theta} = \mathbf{0}$, corresponding to the null hypothesis compatible with the GPM. Notice that, mathematically, $\{\boldsymbol{\theta}^g = \mathbf{0}\}$ is equivalent to $\cap_{\ell=1}^q \{g_\ell(\boldsymbol{\theta}) = 0\}$, so we can express H_0^g as an intersection of component hypotheses $H_0^{g,\ell}$, $\ell = 1, \ldots, q$. That is, we can write

$$H_0^g = \bigcap_{\ell=1}^{q} H_0^{g,\ell}.$$

Under these conditions, a multiple testing procedure ideally has two important properties known as coherence and consonance, as defined in Gabriel (1969):

- Coherence: The global null hypothesis, H_0^g, is rejected if at least one of the component hypotheses, $H_0^{g,\ell}$, is rejected.
- Consonance: If H_0^g is rejected, then at least one $H_0^{g,\ell}$ is rejected.

If a multiple testing procedure is both coherent and consonant, then H_0^g is rejected if, and only if, at least one $H_0^{g,\ell}$, $\ell = 1, \ldots, q$, is rejected.

For the GPM-ES, both coherence and consonance hold by use of the maxT procedure. Recall our vector of test statistics, $\boldsymbol{T}_n^g = (T_n^{g,1}, \ldots, T_n^{g,q})'$, and let $\alpha \in (0, 1)$ be some constant. Then, for the two-tailed hypothesis case (the results are quite similar for the one-sided cases), we can define a critical value, $z_{1-\alpha,2,R_n^{\star g}}$, where $z_{1-\alpha,2,R_n^{\star g}}$ is a quantile such that

$$P(\cap_{\ell=1}^q \{|Z_\ell| \leq z_{1-\alpha,2,R_n^{\star g}}\}) = 1 - \alpha,$$

where $(Z_1, \ldots, Z_q)' \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{R}_n^{\star g})$. We reject $H_0^{g,\ell}$ if and only if $|t_n^{g,\ell}| > z_{1-\alpha,2,R_n^{\star g}}$, where $t_n^{g,\ell}$ is the observed value of $T_n^{g,\ell}$.

In the maxT procedure, each $t_n^{g,\ell}$, $\ell = 1, \ldots, q$, is compared to the same critical value, and thus we can consider the event "at least one $H_0^{g,\ell}$ being rejected" as equivalent to

$$\max\{|t_n^{g,\ell}|, \ell = 1, \ldots, q\} > z_{1-\alpha,2,R_n^{\star g}}.$$

That is, by use of the largest test statistic (in magnitude)

$$t_n^g = \max\{|t_n^{g,\ell}|, \ell = 1, \ldots, q\},$$

the maxT procedure ensures

$$t_n^g = \max\{|t_n^{g,\ell}|, \ell = 1, \ldots, q\} > z_{1-\alpha,2,R_n^{*g}},$$

if, and only if, at least one of $|t_n^{g,\ell}| > z_{1-\alpha,2,R_n^{\star g}}$. Therefore, H_0^g is rejected if, and only if, at least one $H_0^{g,\ell}$ is rejected, and so the MCTP with GPM-ES is both coherent and consonant by using the maxT procedure.

We now discuss another important aspect of multiple testing procedures known as the familywise error rate (FWER), i.e., the probability of making at least one Type I error. Under the given set up, the FWER is defined as

$$\text{FWER}(\cap_{\ell \in I} H_0^{g,\ell}) = P(\text{"Reject at least one true } H_0^{g,\ell} \text{ "} \mid \cap_{\ell \in I} H_0^{g,\ell}),$$

where $I \subseteq \{1, \ldots, q\}$ is some non-empty index set, and $\cap_{\ell \in I} H_0^{g,\ell}$ denotes a set of true component null hypotheses, $H_0^{g,\ell}$. Here, it is desirable to have both weak and strong control of the FWER, which are defined as follows:

- Strong control of the FWER: A multiple comparison procedure has strong control of the FWER if

$$\text{FWER}(\cap_{\ell \in I} H_0^{g,\ell}) \leq \alpha,$$

for any given $\alpha \in (0,1)$ and a corresponding critical value ($z_{1-\alpha,2,R_n^{\star g}}$ in our case). That is, for any subset index I, corresponding to a configuration of true and false null hypotheses, the probability of making at least one Type I error is at most α (Pesarin & Salmaso, 2010; Noguchi et al., 2020).

- Weak control of the FWER: This is a less general case of strong control where $I = \{1, \ldots, q\}$. That is, a multiple comparison procedure has weak control of the FWER if

$$\text{FWER}(\cap_{\ell=1}^{q} H_0^{g,\ell}) = \text{FWER}(H_0^g) \leq \alpha,$$

for any $\alpha \in (0,1)$ and an appropriate critical value. Notice that $\cap_{\ell=1}^{q} H_0^{g,\ell} = H_0^g$ is the global null hypothesis.

Now, by Theorem 2 of Gabriel (1969), we have that the GPM-ES has strong control of the FWER asymptotically by using the maxT procedure. Furthermore, by the discussion above and a straightforward application of the complement rule, it follows that

$$P(\cup_{\ell=1}^{q}\{|T_n^{g,\ell}| > z_{1-\alpha,2,R_n^{\star g}} | H_0^g\}) = 1 - P(\cap_{\ell=1}^{q}\{|T_n^{g,\ell}| \leq z_{1-\alpha,2,R_n^{\star g}} | H_0^g\}) \to 1 - (1-\alpha) = \alpha,$$

as $n \to \infty$. Therefore, we conclude that the GPM-ES has exact weak control of the FWER in the asymptotic sense by using the maxT procedure.

Another important aspect of GPM-ES is the ability to construct $100(1-\alpha)\%$ SCIs for each generalized effect size

$$g_\ell(\boldsymbol{\theta}) = g(\boldsymbol{c}_{\ell,1}\boldsymbol{\theta}) - g(\boldsymbol{c}_{\ell,2}\boldsymbol{\theta}),\ \ell = 1, \ldots, q.$$

We can achieve construction of these SCIs by using the definition of $T_n^{g,\ell}$, $\ell = 1, \ldots, q$, and then solving for the generalized effect size within the probability given below. Recall each test statistic $T_n^{g,\ell}$ has an asymptotic $N(0,1)$ distribution. Furthermore, in Theorem 2, we define

$$T_n^{g,\ell} = \frac{g_\ell(\hat{\boldsymbol{\theta}}_n) - g_\ell(\boldsymbol{\theta})}{\sqrt{s_{\ell,\ell}^{\star g}}},$$

where $s_{\ell,\ell}^{\star g}$ refers to the ℓ-th diagonal entry of the estimated covariance matrix $\boldsymbol{S}_{\boldsymbol{n}}^{\star \boldsymbol{g}}$. Thus, we have,

$$\begin{aligned}
\mathrm{P}\left(\bigcap_{\ell=1}^{q}\{|T_n^{g,\ell}| \leq z_{1-\alpha,2,R_n^{\star g}}\}\right) &= \mathrm{P}\left(\bigcap_{\ell=1}^{q}\left\{\frac{|g_\ell(\hat{\boldsymbol{\theta}}_n) - g_\ell(\boldsymbol{\theta})|}{\sqrt{s_{\ell,\ell}^{\star g}}} \leq z_{1-\alpha,2,R_n^{\star g}}\right\}\right) \\
&= \mathrm{P}\left(\bigcap_{\ell=1}^{q}\left\{|g_\ell(\hat{\boldsymbol{\theta}}_n) - g_\ell(\boldsymbol{\theta})| \leq z_{1-\alpha,2,R_n^{\star g}}\sqrt{s_{\ell,\ell}^{\star g}}\right\}\right) \\
&= \mathrm{P}\left(\bigcap_{\ell=1}^{q}\left\{g_\ell(\boldsymbol{\theta}) \in \left[g_\ell(\hat{\boldsymbol{\theta}}_n) \pm z_{1-\alpha,2,R_n^{\star g}}\sqrt{s_{\ell,\ell}^{\star g}}\right]\right\}\right) \\
&\to 1-\alpha,
\end{aligned}$$

as $n \to \infty$, noting that $\boldsymbol{T}_n^g \overset{\text{a}}{\sim} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{R}_n^{\star g})$.

Therefore, large-sample $100(1-\alpha)\%$ SCIs for $g_\ell(\boldsymbol{\theta})$, $\ell = 1, \ldots, q$, are given by

$$\left[g_\ell(\hat{\boldsymbol{\theta}}_n) - z_{1-\alpha,2,R_n^{\star g}}\sqrt{s_{\ell,\ell}^{\star g}},\ g_\ell(\hat{\boldsymbol{\theta}}_n) + z_{1-\alpha,2,R_n^{\star g}}\sqrt{s_{\ell,\ell}^{\star g}}\right].$$

We have now shown that the GPM-ES asymptotically achieves the important multiple testing procedure properties of coherence and consonance, as well as strong control and exact weak control of the FWER. Furthermore, we have shown how to derive approximate $100(1-\alpha)\%$ SCIs for each effect size under consideration. Next, we complete our demonstration of the GPM-ES using the dental data example.

4.2 MCTP with GPM-ES Example

Returning to our mean root depth comparisons, we wish to consider the global null hypothesis

$$H_0^g : \boldsymbol{\mu}^g = \mathbf{0},$$

with corresponding component hypotheses $H_0^{g,\ell} : g_\ell(\boldsymbol{\mu}) = \log(\mu_i/\mu_{i+1}) = 0$ for $\ell = 1, 2, 3$. We perform a two-tailed hypothesis test, as both a decrease and an increase in mean root depth is meaningful for this dataset. That is, we are testing against the global alternative hypothesis

$$H_1^g : \boldsymbol{\mu}^g \neq \mathbf{0}.$$

We perform this test in two ways, both with significance level $\alpha = 0.05$. First, we simply use a multivariate normal approximation which is supported by the asymptotic results in the preceding chapter. Then, we compare these results to a small-sample approximation with the multivariate t-distribution with $n-1 = 15$ degrees of freedom. The computations are done using the "mvtnorm" package in R.

First, we calculate the maxT test statistic. To do so, we can use the formula under the null hypothesis

$$t_n^{g,\ell} = \frac{g_\ell(\overline{\boldsymbol{x}}_n) - g_\ell(\boldsymbol{\mu})}{\sqrt{s_{\ell,\ell}^{\star,g}}} = \frac{g_\ell(\overline{\boldsymbol{x}}_n)}{\sqrt{s_{\ell,\ell}^{\star,g}}},$$

for $\ell = 1, 2, 3$, to obtain

$$\begin{bmatrix} t_n^{g,1} \\ t_n^{g,2} \\ t_n^{g,3} \end{bmatrix} = \begin{bmatrix} -1.496587 \\ -2.876164 \\ -3.208651 \end{bmatrix},$$

and thus

$$t_n^g = \max\{|t_n^{g,\ell}|, \ell = 1, 2, 3\} = 3.208651.$$

Now, for the multivariate normal approximation, we use the critical value $z_{1-\alpha,2,\boldsymbol{R}_n^{\star,g}}$ corresponding to $\mathcal{N}_3(\mathbf{0}, \boldsymbol{R}_n^{\star,g})$ at $\alpha = 0.05$. That is, $z_{1-\alpha,2,\boldsymbol{R}_n^{\star,g}} = 2.322536$ and the corresponding adjusted p-value is 0.003320752. Since

$$t_n^g = 3.208651 > 2.322536 = z_{1-\alpha,2,\boldsymbol{R}_n^{\star,g}},$$

we reject the global null hypothesis. Notice that our adjusted p-value of 0.003320752, which is smaller than $\alpha = 0.05$, agrees with this conclusion. Furthermore, using the component test statistics $|t_n^{g,1}| = 1.496587, |t_n^{g,2}| = 2.876164$, and $|t_n^{g,3}| = 3.208651$ from above, we reject the second and third component hypotheses, and fail to reject the first by comparing to the same critical value as for the global hypothesis.

Additionally, using the formula

$$\left[g_\ell(\overline{\boldsymbol{x}}_n) - z_{1-\alpha,2,R_n^{\star g}}\sqrt{s_{\ell,\ell}^{\star g}},\ g_\ell(\overline{\boldsymbol{x}}_n) + z_{1-\alpha,2,R_n^{\star g}}\sqrt{s_{\ell,\ell}^{\star g}}\right],$$

for $\ell = 1, 2, 3$ we get the SCIs

$$[-0.0524419,\quad 0.0113500],$$
$$[-0.06912013, -0.00734914],$$
$$[-0.05767559, -0.01401194],$$

where each corresponds to its respective effect size, $g_\ell(\boldsymbol{\mu})$ for $\ell = 1, 2, 3$.

Next, for the multivariate-t approximation, we need to use the critical value $t_{15,\boldsymbol{R}_n^{\star,g}}$ corresponding to the multivariate t-distribution with 15 degrees of freedom. From R we get $t_{15,\boldsymbol{R}_n^{\star,g}} = 2.57343$. Again we have

$$t_n^g = 3.208651 > 2.57343 = t_{15,\boldsymbol{R}_n^{\star,g}}$$

and thus we reject the global null hypothesis. Furthermore, the adjusted p-value of 0.01449197 agrees with this conclusion. Again, we reject the second and third component hypotheses, and fail

to reject the first by similar reasoning as above.

We reject the global null hypothesis in both cases at $\alpha = 0.05$. That is, we conclude that the mean root depth changes significantly over time in young boys at $\alpha = 0.05$ because there is sufficient evidence to conclude that the log-ratio of the mean root depth between successive time points is non-zero. Namely, the log-ratio of the mean root depth between ages 10 and 12 and between 12 and 14 is non-zero since we rejected the second and third component hypotheses.

4.3 Nonparametric MCTP with GPM-ES

Even though our discussion so far has mainly focused on the mean comparisons (which we call "parametric" MCTP for convenience) as an application of GPM-ES, GPM-ES is applicable to a wide range of parameters which have consistent and asymptotically normal estimators. To demonstrate its applicability, in this section, we briefly introduce the idea of GPM-ES in the nonparametric framework (which we call "nonparametric" MCTP for convenience) by mainly discussing a relevant parameter known as relative effect. Note that, in the nonparametric MCTP, there is virtually no distributional assumption on the data, making comparisons of observations from different treatments (e.g., time points) possible even if the means do not exist. One popular example is when the data are coming from the Cauchy distribution.

In the nonparametric MCTP, instead of using the means as parameters of interest, we use relative effect, which is explained below by starting with the two-sample case. Suppose we have two random variables X_1 and X_2, representing observations from the first and second treatment, respectively. Then, the relative effect corresponding to X_1 is a measure of the stochastic superiority of the first

treatment compared to the second. That is, the relative effect, p_{12}, is defined as

$$p_{12} = \mathrm{P}(X_1 < X_2) + 0.5\mathrm{P}(X_1 = X_2).$$

If $p_{12} > 0.5$, then the second treatment is said to be stochastically larger than the first treatment. The terms "stochastically smaller", and "stochastically equal" are defined in a similar fashion.

For the one-way repeated measures case (one-way meaning one homogeneous group compared across repeated measures) with d observations per subject, we define the reference effect by following Konietschke et al. (2010). Let $X_{i,k} \sim F_k$, $i = 1, \ldots, n$, where F_k, $k = 1, \ldots, d$, are non-degenerate distribution functions. The mean distribution function, which is often referred to as reference distribution, is given by

$$G(x) = \frac{1}{d} \sum_{k=1}^{d} F_k(x), \quad -\infty < x < \infty.$$

By viewing G as a distribution function, we define the relative effect for each treatment as

$$p_k = \mathrm{P}(Y < X_{i,k}) + 0.5\mathrm{P}(Y = X_{i,k}),$$

where $Y \sim G$. If $p_k < p_j$, we say that the values from F_k (the k-th treatment) tend to be smaller than those from F_j (the j-th treatment). On the other hand, if $p_k = p_j$, neither distribution tends to be smaller or larger (Noguchi et al., 2012, 2020). The relative effects p_k are popular in nonparametric multiple comparisons as it avoids the problem of non-transitive paradox (Noguchi et al., 2020).

Let $\boldsymbol{p} := (p_1, \ldots, p_d)$ be the vector of relative effects. Konietschke et al. (2010) have shown an asymptotically unbiased, consistent, and asymptotically normal estimator of $\boldsymbol{p}$, which we call $\hat{\boldsymbol{p}}_n$, by estimating F_k using its empirical counterpart. Most importantly, the estimator of $\boldsymbol{p}$ satisfies the asymptotic properties that GPM-ES requires, allowing us to directly construct a nonparametric

MCTP using GPM-ES. That is, we formulate the global null hypothesis

$$H_0^g : \boldsymbol{p}^g = \mathbf{0},$$

where $\boldsymbol{p}^g$ denotes the g-transformed $\boldsymbol{p}$, and proceed as was done before with the estimator $\hat{\boldsymbol{p}}_n^g$. For a suitable g-transformation, Noguchi et al. (2020) suggest a constant multiple of log odds ratio as the resulting effect size resembles Cohen's d.

5. Conclusion

We have successfully constructed a one-step MCTP that accommodates a wide variety of effect sizes using a new framework called GPM-ES. The MCTP with GPM-ES is asymptotically distribution-free, and the resulting rejection criteria provide consistent results between the global and component hypotheses. Moreover, it is important to note that the mathematical interpretation of the hypothesis test matches the practical interpretation for the experimenter as the suggested MCTP is coherent and consonant. Thus, a researcher can utilize the suggested testing procedure to make broad statements about the data as a whole as well as fine-grain statements relating to individual component hypotheses, and these results do not contradict one another.

Addressing the so-called replication crisis in scientific reporting, the MCTP with GPM-ES provides a convenient and flexible framework to deal with effect sizes in multiple comparisons. Specifically, the proposed GPM-ES framework allows researchers to easily compute generalized effect sizes, test statistics, adjusted p-values, and corresponding SCIs simultaneously without any inconsistency. Such a framework greatly enhances the practice of reporting effect sizes as recommended by the American Statistical Association (ASA).

To demonstrate its wide applicability of the MCTP with GPM-ES, we have shown how the proposed framework can be used in both the parametric and nonparametric setting. Given the minimal

assumptions on the data, the nonparametric framework which utilizes relative effect can be applied in a wide array of problems, in addition to the traditional parametric framework. However, one major limitation of the MCTP with GPM-ES is its inability to make statements about a multi-way layout with heterogeneous groups. For example, in the dental dataset, we have used only the boys' data rather than the girls and boys as two distinct groups. Moving forward, it would be beneficial to extend the proposed testing procedure into the case of heterogeneous groups by following Gunawardana and Konietschke (2019), as data of this kind arise frequently in fields of biology, psychology, etc. Furthermore, more work can be done in the nonparametric setting to improve reliable small sample approximations, such as wild bootstrapping (Umlauft et al., 2019).

References

Gabriel K.R. (1969). Simultaneous test procedures – Some theory of multiple comparisons. *The Annals of Mathematical Statistics*, 40(1) , 224–250.

Gunawardana A. and Konietschke F. (2019). Nonparametric multiple contrast tests for general multivariate factorial designs. *Journal of Multivariate Analysis*, 173, 165–180.

Hothorn T., Bretz F., and Westfall P. (2008). Simultaneous inference in general parametric models. *Biometric Journal*, 50(3), 346–363.

Konietschke F., Bathke A.C., Hothorn L.A., and Brunner E. (2010). Testing and estimation of purely nonparametric effects in repeated measures designs. *Computational Statistics and Data Analysis*, 54 , 1895–1905.

Noguchi K., Gel Y.R., Brunner E., and Konietschke F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, 50(12).

Noguchi K., Abel R.S., Marmolejo-Ramos F., and Konietschke F. (2020). Nonparametric multiple comparisons. *Behavior Research Methods*, 52, 489–502.

Pesarin F. and Salmaso L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software*. Wiley.

Potthoff R.F. and Roy S.N. (1964). Generalized multivariate analysis of variance model useful

especially for growth curve problems. *Biometrika*, 51, 313–326.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Umlauft M., Placzek M., Konietschke F., and Pauly M. (2019). Wild bootstrapping rank-based procedures: Multiple testing in nonparametric factorial repeated measures designs. *Journal of Multivariate Analysis*, 171, 176–192.

Wasserstein R. and Lazar N. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133.

Appendices

A. Proof of Theorem 1

Theorem 1

Let $\hat{\boldsymbol{\theta}}_n \overset{a}{\sim} \mathcal{N}_d(\boldsymbol{\theta}, \boldsymbol{S}_n)$. *Then, the asymptotic joint distribution of the vector of test statistics* $\boldsymbol{T}_n := \boldsymbol{D}_n^{-1/2}\boldsymbol{C}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ *is given by*

$$\boldsymbol{T}_n \overset{a}{\sim} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{R}_n^\star),$$

where $\boldsymbol{D}_n = \text{diag}(\boldsymbol{S}_n^\star) \in \mathbb{R}^{q,q}$ *is a diagonal matrix which contains the sample variances associated with* $\boldsymbol{S}_n^\star$. *Moreover,* $\boldsymbol{R}_n^\star = \boldsymbol{D}_n^{-1/2}\boldsymbol{S}_n^\star\boldsymbol{D}_n^{-1/2} \in \mathbb{R}^{q,q}$ *is the sample correlation matrix with* 1*'s along the diagonal entries and sample correlation coefficients in the off-diagonals.*

Proof.

Suppose $\hat{\boldsymbol{\theta}}_n \in \mathbb{R}^d$ is an estimate of $\boldsymbol{\theta}$ and $\boldsymbol{S_n} \in \mathbb{R}^{d\times d}$ is an estimate of $\text{cov}(\hat{\boldsymbol{\theta}}_n)$ with

$$a_n\boldsymbol{S}_n \xrightarrow{p} \boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}, \tag{1}$$

for some positive, nondecreasing sequence a_n. Furthermore, we assume that a multivariate central limit theorem holds, i.e.,

$$a_n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma}). \tag{2}$$

If both (1) and (2) hold then we can write

$$\hat{\boldsymbol{\theta}}_n \overset{\text{a}}{\sim} \mathcal{N}_d(\boldsymbol{\theta}, \boldsymbol{S}_n).$$

Then, by Theorem 3.3 in Serfling (1980), the linear function $\boldsymbol{C}\hat{\boldsymbol{\theta}}_n$ also follows an approximate multivariate normal distribution

$$\boldsymbol{C}\hat{\boldsymbol{\theta}}_n \overset{\text{a}}{\sim} \mathcal{N}_q(\boldsymbol{C}\boldsymbol{\theta}, \boldsymbol{S}_n^\star),$$

with covariance matrix $\boldsymbol{S}_n^\star = \boldsymbol{C}\boldsymbol{S}_n\boldsymbol{C}'$ for any fixed $q \times d$ matrix $\boldsymbol{C}$. Now, simply assume (in analogy with (1) and (2))

$$\boldsymbol{C}\hat{\boldsymbol{\theta}}_n \overset{\text{a}}{\sim} \mathcal{N}_q(\boldsymbol{C}\boldsymbol{\theta}, \boldsymbol{S}_n^\star) \quad \text{with} \quad a_n\boldsymbol{S}_n^\star \xrightarrow{p} \boldsymbol{\Sigma}^\star := \boldsymbol{C}\boldsymbol{S}_n^\star\boldsymbol{C}'. \tag{3}$$

It is assumed that the diagonal elements of the covariance matrix are positive, i.e., $\boldsymbol{\Sigma}^{\star}_{\ell,\ell} > 0$ for $\ell = 1, \ldots, q$. Now, for $\boldsymbol{D}_n = \text{diag}(\boldsymbol{S}^{\star}_n)$ we have

$$a_n \boldsymbol{D}_n \xrightarrow{p} \text{diag}(\boldsymbol{\Sigma}^{\star}),$$

since $a_n \boldsymbol{S}^{\star}_n \xrightarrow{p} \boldsymbol{\Sigma}^{\star}$ by assumption.

Next, define the sequence $\tilde{a}_n$ needed to establish $\tilde{a}_n$-convergence by $\tilde{a}_n \equiv 1$. Then, from above, we have

$$\begin{aligned}
\tilde{a}_n \boldsymbol{R}^{\star}_n &= \boldsymbol{D}_n^{-1/2} \boldsymbol{S}^{\star}_n \boldsymbol{D}_n^{-1/2} \\
&= (a_n \boldsymbol{D}_n)^{-1/2} (a_n \boldsymbol{S}^{\star}_n)(a_n \boldsymbol{D}_n)^{-1/2} \\
&\xrightarrow{p} \text{diag}(\boldsymbol{\Sigma}^{\star})^{-1/2} \boldsymbol{\Sigma}^{\star} \text{diag}(\boldsymbol{\Sigma}^{\star})^{-1/2},
\end{aligned}$$

where convergence in probability to the constant $\boldsymbol{R}^{\star} := \text{diag}(\boldsymbol{\Sigma}^{\star})^{-1/2} \boldsymbol{\Sigma}^{\star} \text{diag}(\boldsymbol{\Sigma}^{\star})^{-1/2}$ follows from Slutsky's Theorem (Theorem 1.5.4, Serfling 1980). Thus,

$$\begin{aligned}
\boldsymbol{T}_n &:= \boldsymbol{D}_n^{-1/2} (\boldsymbol{C}\hat{\boldsymbol{\theta}}_n - \boldsymbol{C}\boldsymbol{\theta}) \\
&= (a_n \boldsymbol{D}_n)^{-1/2} a_n^{1/2} (\boldsymbol{C}\hat{\boldsymbol{\theta}}_n - \boldsymbol{C}\boldsymbol{\theta}) \\
&\xrightarrow{d} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{R}^{\star}),
\end{aligned}$$

or rather,

$$\boldsymbol{T}_n = \boldsymbol{D}_n^{-1/2} (\boldsymbol{C}\hat{\boldsymbol{\theta}}_n - \boldsymbol{C}\boldsymbol{\theta}) \overset{\text{a}}{\sim} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{R}^{\star}_n).$$

B. Proof of Theorem 2

Theorem 2

Let $\hat{\boldsymbol{\theta}}_n \overset{a}{\sim} \mathcal{N}_d(\boldsymbol{\theta}, \boldsymbol{S}_n)$. *Then,* $\hat{\boldsymbol{\theta}}_n^g \overset{a}{\sim} \mathcal{N}_q(\boldsymbol{\theta}^g, \boldsymbol{S}_n^{\star g})$, *where* $\boldsymbol{S}_n^{\star g} = (\hat{\boldsymbol{\Phi}}^g)\boldsymbol{S}_n(\hat{\boldsymbol{\Phi}}^g)'$ *and* $\hat{\boldsymbol{\Phi}}^g = \left[\left.\frac{\partial g_\ell}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}\right]_{q\times d}$. *Moreover the asymptotic joint distribution of the vector of g-transformed test statistics* $\boldsymbol{T}_n^g := (\boldsymbol{D}_n^g)^{-1/2}(\hat{\boldsymbol{\theta}}_n^g - \boldsymbol{\theta}^g)$ *is given by*

$$\boldsymbol{T}_n^g \overset{a}{\sim} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{R}_n^{\star g}),$$

where $\boldsymbol{D}_n^g = \operatorname{diag}(\boldsymbol{S}_n^{\star g})$ *and* $\boldsymbol{R}_n^{\star g} = (\boldsymbol{D}_n^g)^{-1/2}\boldsymbol{S}_n^{\star g}(\boldsymbol{D}_n^g)^{-1/2}$.

Proof.

First, by assumption, $g\colon \mathbb{R} \to \mathbb{R}$ is increasing and continuously differentiable in a neighborhood of $\theta_k \in (a, b)$ for each $k = 1, \ldots, d$, and thus g is differentiable in a neighborhood of $\boldsymbol{c}_{\ell,1}\boldsymbol{\theta} \in (a, b)$ and $\boldsymbol{c}_{\ell,2}\boldsymbol{\theta} \in (a, b)$ as well. Now, define $\boldsymbol{\Phi}^g = \left[\frac{\partial g_\ell}{\partial \theta_k}\right]_{q\times d}$ as our matrix of partial derivatives.

Let $\{a_n\}$ be some positive nondecreasing sequence. By definition of Hothorn et al. (2008), we have $\hat{\boldsymbol{\theta}}_n \overset{\text{a}}{\sim} \mathcal{N}_d(\boldsymbol{\theta}, \boldsymbol{S}_n)$ if and only if

$$a_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \boldsymbol{W} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma})$$

for some random vector $\boldsymbol{W}$, where $a_n\boldsymbol{S}_n \xrightarrow{p} \boldsymbol{\Sigma}$. By the multivariate delta-method,

$$a_n^{1/2}(\hat{\boldsymbol{\theta}}_n^g - \boldsymbol{\theta}^g) \xrightarrow{d} \boldsymbol{Y} \sim \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{\Sigma}^g),$$

where $\boldsymbol{\Sigma}^g := (\boldsymbol{\Phi}^g)\boldsymbol{\Sigma}(\boldsymbol{\Phi}^g)'$. Thus,

$$\hat{\boldsymbol{\theta}}_n^g \overset{\text{a}}{\sim} \mathcal{N}_q(\boldsymbol{\theta}^g, \boldsymbol{S}_n^{\star g}),$$

where $\boldsymbol{S}_n^{\star g} := (\hat{\boldsymbol{\Phi}}^g)\boldsymbol{S}_n(\hat{\boldsymbol{\Phi}}^g)'$, noting that

$$a_n\boldsymbol{S}_n^{*g} = (\hat{\boldsymbol{\Phi}}^g)(a_n\boldsymbol{S}_n)(\hat{\boldsymbol{\Phi}}^g)' \xrightarrow{p} (\boldsymbol{\Phi}^g)\boldsymbol{\Sigma}(\boldsymbol{\Phi}^g)' = \boldsymbol{\Sigma}^g.$$

Now, for $\boldsymbol{\Phi}^g$, the partial derivatives of g with respect to each θ_k, $k = 1, \ldots, d$, are given by

$$g_\ell(\boldsymbol{\theta}) = g(\boldsymbol{c}_{\ell,1}\boldsymbol{\theta}) - g(\boldsymbol{c}_{\ell,2}\boldsymbol{\theta}).$$

Thus,

$$\frac{\partial g_\ell(\boldsymbol{\theta})}{\partial \theta_k} = g'(\boldsymbol{c}_{\ell,1}\boldsymbol{\theta})c_{\ell,1,k} - g'(\boldsymbol{c}_{\ell,2}\boldsymbol{\theta})c_{\ell,2,k},$$

for $\ell = 1, \ldots, q$ and $k = 1, \ldots, d$.

Next, we wish to find the asymptotic joint distribution of the vector of test statistics

$$\boldsymbol{T}_n^g = (\boldsymbol{D}_n^g)^{-1/2}(\hat{\boldsymbol{\theta}}_n^g - \boldsymbol{\theta}^g),$$

where $\boldsymbol{D}_n^g = \text{diag}(\boldsymbol{S}_n^{\star g})$. First, we have

$$a_n\boldsymbol{D}_n^g = \text{diag}(a_n\boldsymbol{S}_n^{\star g}) \xrightarrow{p} \text{diag}(\boldsymbol{\Sigma}^g).$$

To show convergence of random vectors in GPM, we need to define some positive nondecreasing sequence $\{\tilde{a}_n\}$. For $\boldsymbol{T}_n^g$, we simply $\tilde{a}_n = 1$ for all n. Then,

$$\tilde{a}_n\boldsymbol{R}_n^{\star g} = \boldsymbol{R}_n^{\star g} = (a_n\boldsymbol{D}_n^g)^{-1/2}(a_n\boldsymbol{S}_n^{\star g})(a_n\boldsymbol{D}_n^g)^{-1/2} \xrightarrow{p} (\text{diag}(\boldsymbol{\Sigma}^g))^{-1/2}\boldsymbol{\Sigma}^g(\text{diag}(\boldsymbol{\Sigma}^g))^{-1/2}.$$

Then, by Slutsky's theorem,

$$\tilde{a}_n^{1/2}(\boldsymbol{D}_n^g)^{-1/2}(\hat{\boldsymbol{\theta}}_n^g - \boldsymbol{\theta}^g) = (a_n\boldsymbol{D}_n^g)^{-1/2}a_n^{1/2}(\hat{\boldsymbol{\theta}}_n^g - \boldsymbol{\theta}^g) \xrightarrow{d} (\text{diag}(\boldsymbol{\Sigma}^g))^{-1/2}\boldsymbol{Y}$$

where $(\text{diag}(\boldsymbol{\Sigma}^g))^{-1/2}\boldsymbol{Y} \sim \mathcal{N}_q(\boldsymbol{0}, (\text{diag}(\boldsymbol{\Sigma}^g))^{-1/2}\boldsymbol{\Sigma}^g(\text{diag}(\boldsymbol{\Sigma}^g))^{-1/2})$. Thus, we have shown that

$$\boldsymbol{T}_n^g \overset{\text{a}}{\sim} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{R}_n^{\star g}).$$

C. R Code

```
#################
### Section 2.2
### Using the GPM with Pairwise Comparisons
#################

dentaldata <- read.csv(file = "harvard_dental_data.csv")

# Observe data to make sure read in correctly
summary(dentaldata)

# Take square root to normalize data
root_dentaldata <- sqrt(dentaldata)

# Convert dataframes to vectors for histogram
dentaldata_vector <- c(dentaldata$yr8,dentaldata$yr10
                       ,dentaldata$yr12,dentaldata$yr14)
root_dentaldata_vector <- c(root_dentaldata$yr8,root_dentaldata$yr10
                            ,root_dentaldata$yr12,root_dentaldata$yr14)
hist(root_dentaldata_vector, 16)

hist(root_dentaldata$yr8, 16, main = "Mean Root Depth fo 8yr Old Boys"
     , xlab = "Mean Root Depth")
mean(root_dentaldata$yr8)
```

```
# Generate box plots for original data and square root of data
par(mfrow = c(1,2))
boxplot(dentaldata$yr8, dentaldata$yr10, dentaldata$yr12, dentaldata$yr14
        , names = c("8 yr", "10 yr", "12 yr", "14 yr")
        , main = "Depth (mm)")
boxplot(root_dentaldata$yr8, root_dentaldata$yr10
        , root_dentaldata$yr12, root_dentaldata$yr14
        , names = c("8 yr", "10 yr", "12 yr", "14 yr")
        , main = "Square Root Depth")


# Calculate sample mean for each repeated measure
xbar1 <- mean(root_dentaldata$yr8)
xbar2 <- mean(root_dentaldata$yr10)
xbar3 <- mean(root_dentaldata$yr12)
xbar4 <- mean(root_dentaldata$yr14)
n <- 16

# Create xbar vector
xbar <- c(xbar1,xbar2,xbar3,xbar4)
xbar_matrix <- as.matrix(xbar)

# Calculate covariance matrix estimator S_n
S_n <- cov(root_dentaldata)/n
```

```
# Create contrast matrix
C <- matrix(c(1,-1,0,0,0,1,-1,0,0,0,1,-1), nrow = 3, ncol=4, byrow=TRUE)

# Calculate transformed parameters of interest.
C_xbar <- C%*%xbar

# Calculate S_star and D_n
S_star <- C%*%S_n%*%t(C)

# Get diagonal matrix corresponding to S_star
D_n <- diag(diag(S_star))

# Get transformed version for test statistic
inv_root_D_n <- diag(1/sqrt(diag(S_star)))

# Calculate R_star_n
R_star_n <- inv_root_D_n%*%S_star%*%inv_root_D_n
```

```
#################
### Section 3.2
### Using the GPM-ES with g(x) = log(x)
#################

# Construct contrast vectors
c_11 <- c(1,0,0,0)
c_12 <- c(0,1,0,0)
c_21 <- c(0,1,0,0)
c_22 <- c(0,0,1,0)
c_31 <- c(0,0,1,0)
c_32 <- c(0,0,0,1)

c_1 <- c_11 - c_12 # First row of C matrix
c_2 <- c_21 - c_22 # Second row of C matrix
c_3 <- c_31 - c_32 # Third row of C matrix

# Calculate generalized effect size
g_1_xbar <- log(xbar1) - log(xbar2)
g_2_xbar <- log(xbar2) - log(xbar3)
g_3_xbar <- log(xbar3) - log(xbar4)

g_xbar <- c(g_1_xbar, g_2_xbar, g_3_xbar)
```

```
# Calculate phi matrix
grad_g_1_xbar <- c(1/xbar1, -1/xbar2, 0, 0)
grad_g_2_xbar <- c(0, 1/xbar2, -1/xbar3, 0)
grad_g_3_xbar <- c(0, 0, 1/xbar3, -1/xbar4)

phi <- matrix(c(grad_g_1_xbar, grad_g_2_xbar, grad_g_3_xbar)
              , nrow = 3, ncol = 4, byrow = TRUE)



# Calculate S_star_g
S_star_g <- phi%*%S_n%*%t(phi)

# Calculate D_n_g
D_n_g <- diag(diag(S_star_g))

# Negative root of D_n
inv_root_D_n_g <- diag(1/sqrt(diag(S_star_g)))

# Calculate R_star_g
R_star_g <- inv_root_D_n_g%*%S_star_g%*%inv_root_D_n_g
```

```
#################
### Section 4.2
### Example with GPM-ES MCTP
################

# Global Null is H_0: theta_g = 0 at alpha = 0.05

# Calculate vector of g-transformed test statistics
t_n_g <- inv_root_D_n_g%*%as.matrix(g_xbar)

max_t_g <- max(abs(t_n_g))
#[1] 3.208651

# Small-sample multivariate t-approximation for multivariate t-dist
t_quantile <- qmvt(p=.95, tail = "both.tails", df = 15, corr = R_star_g)
t_crit <- t_quantile$quantile
#[1] 2.574402

# Adjusted p-value for multivariate t approximation
t_prob <- pmvt(lower = rep(-max_t_g, 3), upper = rep(max_t_g, 3)
               , df = 15, corr = R_star_g)
adj_p_t <- 1-t_prob
#[1] 0.01449197 agrees with conclusion
```

```
# Multivariate normal approximation
z_quantile <- qmvnorm(p = 0.95, tail = "both.tails", corr = R_star_g)
z_crit <- z_quantile$quantile
#[1] 2.322684

# Adjusted p-value for multivariate normal approximation
z_prob <- pmvnorm(lower = rep(-max_t_g, 3)
                  , upper = rep(max_t_g, 3), corr = R_star_g)
adj_p_z <- 1-z_prob
#[1] 0.003320752 agrees with conclusion

# Calculate SCI with multivariate normal case
conf_int_1 <- c(g_1_xbar - z_crit*sqrt(S_star_g[1,1])
                , g_1_xbar + z_crit*sqrt(S_star_g[1,1]))
#[1] -0.0524419  0.0113500

conf_int_2 <- c(g_2_xbar - z_crit*sqrt(S_star_g[2,2])
                , g_2_xbar + z_crit*sqrt(S_star_g[2,2]))
#[1] -0.069120131 -0.007349141

conf_int_3 <- c(g_3_xbar - z_crit*sqrt(S_star_g[3,3])
                , g_2_xbar + z_crit*sqrt(S_star_g[3,3]))
#[1] -0.05767559 -0.01401194
```