5-2024

IANOVA: Multi-Sample Means Comparisons for Imprecise Interval Data

Zachary Rios
Utah State University, a02282356@usu.edu

IANOVA: MULTI-SAMPLE MEANS COMPARISONS FOR IMPRECISE INTERVAL DATA

by

Zachary Rios

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTERS OF SCIENCE

in

Statistics

Approved:

Yan Sun, Ph.D. Major Professor	Daniel Coster, Ph.D. Committee Member
John Stevens, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan,Utah

2024

ABSTRACT

IANOVA: Multi-Sample Means Comparisons for Imprecise Interval Data

by

Zac Rios, Master of Science

Utah State University, 2024

Major Professor: Yan Sun, Ph.D.
Department: Mathematics and Statistics

In recent years, interval data has become an increasingly popular tool to solve modern data problems. Intervals are now often used for dimensionality reduction, data aggregation, privacy censorship, and quantifying awareness of various uncertainties. Among many statistical methods that are being studied and developed for interval data, the significance test is particularly of importance due to its fundamental value both in theory and practice. The difficulty in developing such tests mainly lies in the fact that the concept of normality does not extend naturally to interval data, causing the exact tests to be hard to formulate. As a result, most existing works have relied on bootstrap methods to approximate null distributions. However, this is not always feasible given limited sample sizes or other intrinsic characteristics of the data. Asymptotic tests, on the other hand, are good alternatives. In the literature, asymptotic tests for comparing means of one or two sample data have been developed, which motivates the exploration of the multi-sample case. In this thesis, we propose a novel asymptotic method for comparing multi-sample means with interval data which turns out to be analogous to classical ANOVA. This procedure builds a test statistic based on a ratio of between-group interval variance and within-group interval variance. We derive the limiting distribution of this test-statistic under usual assumptions and mild regularity conditions.

Simulation results with both discrete and continuous data further validate this result, and show promising small sample performances. Finally, we apply our method to ground snow load interval data, where we are able to detect interval mean differences across regions in Canada.

(55 pages)

PUBLIC ABSTRACT

IANOVA: Multi-Sample Means Comparisons for Imprecise Interval Data

Zac Rios

In recent years, interval data has become an increasingly popular tool to solve modern data problems. Intervals are now often used for dimensionality reduction, data aggregation, privacy censorship, and quantifying awareness of various uncertainties. Among many statistical methods that are being studied and developed for interval data, the significance test is particularly of importance due to its fundamental value both in theory and practice. The difficulty in developing such tests mainly lies in the fact that the concept of normality does not extend naturally to interval data (due the range of an interval being necessarily non-negative), causing the exact tests to be hard to formulate. In the literature, tests for comparing means of one or two sample interval data have been developed, which motivates the exploration of the multi-sample case. In this thesis, we propose a novel asymptotic (as the sample size goes to infinity) method for comparing multi-sample means with interval data. This procedure builds a test statistic based on a ratio of between-group interval variance and within-group interval variance. The theoretical results for this procedure are derived. Simulation results with both discrete and continuous data validate our procedure, and show promising small sample performances. Finally, we apply our method to ground snow load interval data, where we are able to detect interval mean differences across regions in Canada.

ACKNOWLEDGMENTS

I would like to extend my deepest thanks to my advisor, Dr. Sun. There were times I did not feel that I was cut out for this, but I could not have asked for a better mentor. Your encouragement, advice, and assistance made me want to create something I could be proud of, and I will always be grateful for that.

Zac Rios

CONTENTS

LIST OF TABLES

Table | Page

LIST OF FIGURES

CHAPTER 1
INTRODUCTION AND PRELIMINARIES

1.1
Introduction

Interval-valued data has become increasingly popular over the past decades, owing largely to its ability to help with modern data problems. These problems range from a need for data aggregation or dimension reduction, to a greater awareness about uncertainty from various resources, to necessary censoring of private information. Interval-valued data are commonly represented by the associated lower and upper bounds, or centers and ranges alternatively, as opposed to single numbers. This thesis aims to extend classical ANOVA tests for multi-sample means comparisons to interval-valued data.The proposed tests are expected to provide meaningful inferences regarding the means that aligns well with the nature of interval-valued data.

The probabilistic foundation for set-valued data analysis (with interval-valued data being a special case) has been well-established for many years [1]. Many important properties of random sets, such as mean, variance and covariance, have been rigorously studied [2]. These results paved the way for many classical statistical methods to be extended to the domain of random sets. In the past decades, numerous models and methods have been proposed for interval-valued data analysis, often extensions of linear regression [3, 4, 5], but other methods like checking dependency [6, 7] and kriging [8, 9] have been extended to the interval-valued case. This thesis particularly concerns the testing of multiple interval-valued means.

Analogues to certain statistical tests for interval-valued data have already been developed. In particular, analogues to one and two-sample t-tests were developed in Sun [10], where distance metrics on random intervals were used to derive a closed form for comparing means of one and two groups of intervals. There have also been fruitful results in the literature on the comparisons of multi-sample means for intervals, typically done in the more abstract framework of fuzzy random sets. In Montenegro et. al [11], hypothesis tests were constructed for normal fuzzy random variables, using both asymptotic and bootstrap methods. The theoretical results have thus far been fairly limited, with most real problems relying on bootstrapping. In Gil et. al [12], a direct analog to ANOVA was suggested using the same kind of "between-group variance" and "within group variance" as in classical ANOVA. This paper utilized bootstrapping on simulated fuzzy random data, constructing a limiting distribution from the data and finding those associated test statistics. In Colubi [13], the authors worked specifically with intervals, due to their prevelance in real data. This paper also used bootstrap methods, but worked with only the one sample case. This thesis focuses on random intervals and constructs the ANOVA test by deriving the asymptotic null distribution, which allows for direct inference on the observed data. This is particularly useful for small sample designs, where bootstrapping may not be feasible.

To develop the test statistics and the associated null distribution, an essential task is to define an appropriate distance for intervals. Several distance functions have been studied in the literature for random sets (with random intervals being the one-dimensional special case) including the Hausdorff distance. Recent advances in the random sets theory shows that the L_2 type distances derived from embedding the space of compact convex sets $\mathcal{K}_\mathcal{C}(\mathbb{R}^\mathrm{d})$ into the Banach space of continuous functions is more preferred for statistical inferences (see Section 1.2 for details). As such, our ANOVA test statistic is developed based on a generalized L_2 metric in $\mathcal{K}_\mathcal{C}(\mathbb{R})$, referred to as the W-distance. It has an analogous form to the standard ANOVA test statistic that follows an F distribution under the null hypothesis with normal assumptions. For our test statistic, under mild assumptions, we show that the asymptotic null distribution is represented as a linear combination of chi-squared random variables with a certain correlation structure.This result is further validated by a systematic simulation that considers a wide range of distributions for (both the center and range of) the intervals. Through simulation, we are also able to evaluate its small sample performance, which becomes fairly accurate for group sizes of as small as 8. Finally, our real data results show similar promise. We utilize a data set with intervals containing the maximum ground ratio (ratio between maximal roof snow load and maximal ground snow load), and minimum ground ratio (ratio between minimal roof snow load and maximal ground snow load). We use this snow load ground ratio (GR) interval data to analyze the potential differences in GR between regions of Canada. For meaningful partitions of the data, we are able to show a significant effect for the eco-region of the observation on the GR interval. This interval analysis is compared to a classical ANOVA on this data, and the benefits/drawbacks of each approach are discussed. In particular, the ability of interval-valued ANOVA to detect different kinds of results shows merit in this analysis. These results provide justification for further investigation into the regional factors affecting GR, which have thus far been only sparingly considered.

The rest of the paper is organized as follows. Section 1.2 gives a brief review of the basics of random sets theory. Section 2.1 extends the sums of squares decomposition result commonly used in ANOVA to the interval-valued context, which is an important basis for developing a testing procedure. This is followed by Section 2.2, which formulates the test under a standard hypothesis testing framework, i.e. assumptions, hypotheses, statistics. Section 3.1 summarizes our simulation results. Section 3.2 utilizes the proposed method to analyze a real data set, which builds on a previously established result [9] to compare interval GR snow loads across different regions of Canada. Finally, Section 4 makes a few final notes on a similar method (MANOVA) and briefly discusses particularly relevant ideas for future research in interval data analysis.

1.2

Random Sets Preliminaries

Let $\mathcal{K}\left(\mathbb{R}^d\right)$ or $\mathcal{K}$ denote the collection of all non-empty compact subsets of $\mathbb{R}^d$. The Hausdorff metric

$$\rho_H\left(A, B\right) = \max\left(\sup_{a\in A}\rho\left(a, B\right), \sup_{b\in B}\rho\left(b, A\right)\right),\ \forall A, B \in \mathcal{K},$$

where ρ denotes the Euclidean metric, defines a metric in $\mathcal{K}\left(\mathbb{R}^d\right)$. $\mathcal{K}\left(\mathbb{R}^d\right)$ is complete and separable [14] as a metric space. In the space $\mathcal{K}$, Minkowski addition and scalar multiplication can define a linear structure:

$$A + B = \{a + b : a \in A, b \in B\}, \qquad \lambda A = \{\lambda a : a \in A\}, \tag{1.1}$$

$\forall A, B \in \mathcal{K}$ and $\lambda \in \mathbb{R}$. $\mathcal{K}\left(\mathbb{R}^d\right)$, however, is neither a linear space or a vector space, as there is no inverse element of addition.

Consider a probability space given by $(\Omega, \mathcal{L}, P)$. A random compact set is a Borel measurable function $A : \Omega \to \mathcal{K}$, $\mathcal{K}$ being equipped with the Borel σ-algebra induced by the Hausdorff metric. If $A(\omega)$ is convex almost surely, then A is called a random compact convex set [15]. Denote the collection of all compact convex subsets of $\mathbb{R}^d$ as $\mathcal{K}_\mathcal{C}\left(\mathbb{R}^d\right)$ or $\mathcal{K}_\mathcal{C}$. Historically, random sets theory has often focused on compact convex sets [16, 17, 18, 19]. When $d = 1$, $\mathcal{K}_\mathcal{C}(\mathbb{R})$ contains all the non-empty bounded closed intervals in $\mathbb{R}$. A Borel measurable function $[X] : \Omega \to \mathcal{K}_\mathcal{C}\left(\mathbb{R}\right)$ is called a random interval. The expectation of a random compact convex random set A is defined by the

Aumann integral of set-valued function [17, 16] as

$$E(A) = \{E\xi : \xi \in A \text{ almost surely}\}.$$

In the interval case, the Aumann expectation of [X] is given by

$$E([X]) = [E(X^L), E(X^U)]. \tag{1.2}$$

For each $X \in \mathcal{K}_{\mathcal{C}}(\mathbb{R}^d)$, the function defined on the unit sphere S^{d-1}:

$$s_X(u) = \sup_{x \in X} \langle u, x \rangle, \quad \forall u \in S^{d-1}$$

is called the support function of X. Let $\mathcal{S}$ be the space of support functions of all non-empty compact convex subsets in $\mathcal{K}_{\mathcal{C}}$. Then, $\mathcal{S}$ is a Banach space equipped with the L_2 metric

$$\|s_X(u)\|_2 = \left[\int_{S^{d-1}} |s_X(u)|^2 \mu(du)\right]^{\frac{1}{2}},$$

where μ is the normalized Lebesgue measure on S^{d-1}. Various embedding theorems [20, 21] show that $\mathcal{K}_{\mathcal{C}}$ can be embedded isometrically into the Banach space $C(S)$ of continuous functions on S^{d-1}, and $\mathcal{S}$ is the image of $\mathcal{K}_{\mathcal{C}}$ into $C(S)$. Therefore, $\rho_2(X, Y) := \|s_X - s_Y\|_2$, $\forall X, Y \in \mathcal{K}_{\mathcal{C}}$, defines an L_2 metric on $\mathcal{K}_{\mathcal{C}}$. ρ_H and ρ_2 are known to be equivalent metrics, but ρ_H is less preferred for statistical analysis for several reasons. Of note, $E\rho_H^2(X, h(X))$ is not minimized at $h(X) = E(X)$. Particularly the ρ_2-metric for an interval $[x] \in \mathcal{K}_{\mathcal{C}}(\mathbb{R})$ is

$$\|[x]\|_2 = \|s_{[x]}(u)\|_2 = \frac{1}{2}(x^L)^2 + \frac{1}{2}(x^U)^2 = (x^C)^2 + (x^R)^2,$$

and the ρ_2-distance between two intervals $[x], [y] \in \mathcal{K}_{\mathcal{C}}(\mathbb{R})$ is

$$\begin{aligned}
\rho_2([x], [y]) &= \left[\frac{1}{2}(x^L - y^L)^2 + \frac{1}{2}(x^U - y^U)^2\right]^{\frac{1}{2}} \\
&= \left[(x^C - y^C)^2 + (x^R - y^R)^2\right]^{\frac{1}{2}}.
\end{aligned}$$

The ρ_2-distance can be generalized to the ρ_W-distance [22] as

$$\rho_W([x],[y]) = \left\{ \int_{[0,1]} \left[f_{[x]}(\lambda) - f_{[y]}(\lambda) \right]^2 dW(\lambda) \right\}^{\frac{1}{2}}, \tag{1.3}$$

where $f_{[x]}(\lambda) = \lambda x^U + (1-\lambda)x^L$, $\forall\lambda \in [0,1]$, and W is some non-degenerate symmetric measure on $[0,1]$. The ρ_W-distance has an advantage over ρ_2 with its flexibility to assign weights to the points in the interval. In particular, this can be interpreted as a probability distribution for a random point inside the interval. On the other hand, it can be shown that

$$\rho_W^2([x],[y]) = \left(x^C - y^C\right)^2 + \left(x^R - y^R\right)^2 \int_{[0,1]} (2\lambda - 1)^2 \, dW(\lambda).$$

Notice that $\omega = \int_{[0,1]} (2\lambda-1)^2 \, dW(\lambda) \in [0,1]$ is a constant determined by W. Thus, the ρ_W-distance can also be interpreted as choosing a weight for $\left(X^R - Y^R\right)^2$ in calculating the L_2 distance.

CHAPTER 2
SUM OF SQUARES AND TEST FORMULATION

2.1
Sum of Squares for Intervals

Throughout this thesis, I will denote by $[x] = [x^L, x^U]$ a bounded closed interval, where x^L and x^U are the associated lower and upper bounds, respectively. Alternatively, $[x]$ can also be represented by its center and radius (half-range), denoted by x^C and x^R, respectively. Capital letters denote random elements, so a random interval which takes values in $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$ is denoted by $[X]$. Bolded letters denote vectors. For example, $[\mathbf{x}] = [[x_1], \cdots, [x_p]]^T$ denotes a p-dimensional hyper interval. Additionally, the ith group mean and overall means for both the center and radius will be denoted by $\bar{x}_{i.}$ and $\bar{x}_{..}$, respectively. The jth individual observation for the ith group is denoted x_{ij}.

We would like to show that an error decomposition similar to that of ANOVA can be reasonably applied to interval data. Consider defining the error using the w-distance for a given interval as follows:

$$(x^C_{ij} - \bar{x}^C_{..})^2 + \omega(x^R_{ij} - \bar{x}^R_{..})^2 \tag{2.1}$$

This lends itself well to a similar formulation for sum of squares total, as seen in classical ANOVA

$$\begin{aligned} \text{SST} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x^C_{ij} - \bar{x}^C_{..})^2 + \omega(x^R_{ij} - \bar{x}^R_{..})^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x^C_{ij} - \bar{x}^C_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\omega(x^R_{ij} - \bar{x}^R_{..})^2 \end{aligned} \tag{2.2}$$

We want to show that this total error can be similarly decomposed into orthogonal components associated with treatment/error effects. For ease of formulation, the following two terms are further defined:

$$SST^C = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}^C - \bar{x}_{..}^C)^2$$
$$SST^R = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\omega(x_{ij}^R - \bar{x}_{..}^R)^2$$
$$SST = SST^C + SST^R \tag{2.3}$$

Taking a look at SST^C first:

$$\begin{aligned} SST^C &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}^C - \bar{x}_{..}^C)^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(x_{ij}^C - \bar{x}_{i.}^C) + (\bar{x}_{i.}^C - \bar{x}_{..}^C)]^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(x_{ij}^C - \bar{x}_{i.}^C)^2 + 2(x_{ij}^C - \bar{x}_{i.}^C)(\bar{x}_{i.}^C - \bar{x}_{..}^C) + (\bar{x}_{i.}^C - \bar{x}_{..}^C)^2] \end{aligned}$$

We are most interested in the middle term here, which can be rewritten after distributing the sum through as follows:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}2(x_{ij}^C - \bar{x}_{i.}^C)(\bar{x}_{i.}^C - \bar{x}_{..}^C) = 2\sum_{i=1}^{k}(\bar{x}_{i.}^C - \bar{x}_{..}^C)\sum_{j=1}^{n_i}(x_{ij}^C - \bar{x}_{i.}^C)$$

The inner summation is 0, since it calculate the sum of the deviations from the group mean:

$$\sum_{j=1}^{n_i}(x_{ij}^C - \bar{x}_{i.}^C) = 0$$

This implies that the entire middle term is also 0, leaving us with:

$$\begin{aligned} SST^C &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(x_{ij}^C - \bar{x}_{i.}^C)^2 + (\bar{x}_{i.}^C - \bar{x}_{..}^C)^2] \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}^C - \bar{x}_{i.}^C)^2 + \sum_{i=1}^{k} n_i(\bar{x}_{i.}^C - \bar{x}_{..}^C)^2 \end{aligned}$$

This form of the equation decomposes all of the error we see with regard to the center of the interval into a treatment component and a random error component. Staying consistent with notation, we could say that:

$$\mathrm{SST}^{C} = \mathrm{SSE}^{C} + \mathrm{SSTr}^{C} \tag{2.4}$$

The process for the "range" error component is similar, and shown quickly below:

$$\begin{aligned}
\mathrm{SST}^{R} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \omega(x_{ij}^{R} - \bar{x}_{..}^{R})^2 \\
&= \omega \sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij}^{R} - \bar{x}_{i.}^{R})^2 + \omega \sum_{i=1}^{k} n_i(\bar{x}_{i.}^{R} - \bar{x}_{..}^{R})^2 \\
\mathrm{SST}^{R} &= \omega\mathrm{SSE}^{R} + \omega\mathrm{SSTr}^{R} \qquad (2.5)
\end{aligned}$$

Finally, we can show that:

$$\begin{aligned}
\mathrm{SST} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij}^{C} - \bar{x}_{..}^{C})^2 + \omega(x_{ij}^{R} - \bar{x}_{..}^{R})^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij}^{C} - \bar{x}_{..}^{C})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i} \omega(x_{ij}^{R} - \bar{x}_{..}^{R})^2 \\
&= \mathrm{SST}^{C} + \mathrm{SST}^{R} \\
SST &= \mathrm{SSE}^{C} + \mathrm{SSTr}^{C} + \omega\mathrm{SSE}^{R} + \omega\mathrm{SSTr}^{R} \qquad (2.6)
\end{aligned}$$

Thus completes the error decomposition for interval valued data, using the w-distance. This provides solid justification to approach IANOVA similarly to classical ANOVA.

2.2

Test Formulation

2.2.1

Assumptions

Our assumptions are consistent with those of point valued ANOVA, with two notable exceptions. The assumptions are, namely:

- All observations are independent of one another
- Equal variances for both the center and radius in each group
- Equal correlation between the center and radius in each group

Our first two assumptions are functionally the same as ANOVA, with a natural extension to interval-valued data. If the observations are not independent of each other, we are naturally unable to distinguish between error, dependence, and treatment effect. Our intervals must also have equal variances across groups. This is necessary to have a manageable limiting distribution for our test statistic under the null hypothesis. While this assumption does not provide a convenient cancellation as in standard ANOVA, it allows us to meaningfully analyze the convergence of our test statistic. Our final assumption is similarly important for our limiting distribution. An equal correlation across groups lets us preserve the relationship between our chi-square random variables, and simplify our final form. Without the previous two assumptions, we run into the problem of having far too many parameters, and an intractable limiting distribution. Of note, the common assumption of normality in ANOVA is missing from our list. It is not reasonable to assume normality for the interval-valued case, as the radius is necessarily not normal by definition. This is also the purpose of constructing a limiting distribution, rather than an exact one. The structure of intervals lends itself better to utilizing the central limit theorem, rather than assuming an exact distribution. With these assumptions in mind, we are able to construct hypotheses for our test.

2.2.2 Hypotheses

Assuming observing g independent samples of i.i.d. random intervals:

$$\begin{aligned}&[X_{ij}] = [X_{ij}^L, X_{ij}^U] = [X_{ij}^C - X_{ij}^R, X_{ij}^C + X_{ij}^R],\\ &i = 1, \cdots, g,\\ &j = 1, \cdots, n_i,\end{aligned}$$

where n_i is the size of the i^{th} sample and $\sum_{i=1}^{g} n_i = N$. The interval-valued observations within each sample are assumed to be independent from an underlying distribution with a unique mean and common variance-covariance. It can be conveniently described by the joint distribution of the center and radius as, $\forall i$:

$$\begin{bmatrix} X_{ij}^C \\ X_{ij}^R \end{bmatrix} \overset{i.i.d.}{\sim} F\left(\begin{bmatrix} \mu_i^C \\ \mu_i^R \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_C^2 & \sigma_{CR} \\ \sigma_{CR} & \sigma_R^2 \end{bmatrix} \right), \quad j = 1, \cdots, n_i, \tag{2.7}$$

where $\mu_i^C, \sigma_{CR} \in \mathbb{R}$, $\mu_i^R, \sigma_C^2, \sigma_R^2 > 0$. It follows that the population mean for the i^{th} group is

$$E([X_{ij}]) = [\mu_i^C - \mu_i^R, \mu_i^C + \mu_i^R] := [\mu_i].$$

Our null hypothesis is that the group means are uniformly the same, which is stated as

$$H_0 : [\mu_1] = [\mu_2] = \cdots = [\mu_g]. \tag{2.8}$$

The corresponding alternative hypothesis is

$$H_A : [\mu_i] \neq [\mu_j] \quad \text{for some i, j.} \tag{2.9}$$

2.2.3

Test Statistic

Our test statistic will compare quantities similar to how ANOVA does.

$$\begin{aligned} F &= \frac{MSTr}{MSE} = \frac{SSTr/df_1}{SSE/df_2} = \frac{(SSTr^C + SSTr^R)/df_1}{(SSE^C + SSE^R)/df_2} \\ &= \frac{\sum_{i=1}^{g} n_i d_w(\bar{X}_{i\cdot}, \bar{X}_{\cdot\cdot})^2/(g-1)}{\sum_{i=1}^{g}\sum_{j=1}^{n} d_w(X_{ij}, \bar{X}_i)^2/(N-g)} \\ &= \frac{\sum_{i=1}^{g} n_i((\bar{X}_{i\cdot}^C - \bar{X}_{\cdot\cdot}^C)^2 + \omega(\bar{X}_{i\cdot}^R - \bar{X}_{\cdot\cdot}^R)^2)/(g-1)}{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}^C - \bar{X}_{i\cdot}^C)^2 + \omega(X_{ij}^R - \bar{X}_{i\cdot}^R)^2(N-g)} \end{aligned}$$

Intuitively, this compares how different the averages of our groups are to the average of our entire sample, over how different individual intervals are from their group average. Large values of our test statistic imply that most of our variance is attributed to how different groups are from other groups, while small values imply that most of our variance is attributed to how different our groups are within themselves. We will first show the weak convergence of the numerator in Theorem 2.2.1, then the convergence in probability of the denominator in Theorem 2.2.2.

Proofs of Convergence

Theorem 2.2.1. *Assume the interval-valued ANOVA model (2.7). Under the null hypothesis (2.8), the numerator of our test statistic converges in distribution to a weighted sum of correlated chi-*

squared random variables:

$$\sum_{i=1}^{g} n_i \left[(\bar{X}_{i\cdot}^{C} - \bar{X}_{\cdot\cdot}^{C})^2 + \omega(\bar{X}_{i\cdot}^{R} - \bar{X}_{\cdot\cdot}^{R})^2\right] \xrightarrow{\mathcal{D}} \sigma_C^2\chi_C^2(g-1) + \omega\sigma_R^2\chi_R^2(g-1). \tag{2.10}$$

Proof. Denote

$$S^C = \sum_{i=1}^{g} n_i \left(\bar{X}_{i\cdot}^{C} - \bar{X}_{\cdot\cdot}^{C}\right)^2, \quad S^R = \sum_{i=1}^{g} n_i \left(\bar{X}_{i\cdot}^{R} - \bar{X}_{\cdot\cdot}^{R}\right)^2. \tag{2.11}$$

Consider separately the two statistics:

$$W^C = \sum_{i=1}^{g} n_i \left(\bar{X}_{i\cdot}^{C} - \mu_i^C\right)^2, \quad W^R = \sum_{i=1}^{g} n_i \left(\bar{X}_{i\cdot}^{R} - \mu_i^R\right)^2. \tag{2.12}$$

Under the null hypothesis, we have

$$\begin{aligned}
W^C &= \sum_{i=1}^{g} n_i \left(\bar{X}_{i\cdot}^{C} - \bar{X}_{\cdot\cdot}^{C} + \bar{X}_{\cdot\cdot}^{C} - \mu_i^C\right)^2 \\
&= \sum_{i=1}^{g} n_i \left[\left(\bar{X}_{i\cdot}^{C} - \bar{X}_{\cdot\cdot}^{C}\right)^2 + \left(\bar{X}_{\cdot\cdot}^{C} - \mu_i^C\right)^2 + 2\left(\bar{X}_{i\cdot}^{C} - \bar{X}_{\cdot\cdot}^{C}\right)\left(\bar{X}_{\cdot\cdot} - \mu_i^C\right)\right] \\
&= \sum_{i=1}^{g} n_i \left(\bar{X}_{i\cdot}^{C} - \bar{X}_{\cdot\cdot}^{C}\right)^2 + \sum_{i=1}^{g} n_i \left(\bar{X}_{\cdot\cdot}^{C} - \mu^C\right)^2 + 2\left(\bar{X}_{\cdot\cdot} - \mu^C\right)\sum_{i=1}^{g} n_i \left(\bar{X}_{i\cdot}^{C} - \bar{X}_{\cdot\cdot}^{C}\right) \\
&= S^C + N\left(\bar{X}_{\cdot\cdot}^{C} - \mu^C\right)^2,
\end{aligned}$$

Since the cross-term is 0. Similarly,

$$W^R = S^R + N\left(\bar{X}_{\cdot\cdot}^{R} - \mu^R\right)^2.$$

Thus, we obtain

$$W^C + \omega W^R = \left(S^C + \omega S^R\right) + N\left[\left(\bar{X}_{\cdot\cdot}^{C} - \mu^C\right)^2 + \omega\left(\bar{X}_{\cdot\cdot}^{R} - \mu^R\right)^2\right]. \tag{2.13}$$

Now, under the null hypothesis that $\mu_i^C \equiv \mu^C$, $\mu_i^R \equiv \mu^R$, $i = 1, 2, \ldots, g$, we have

$$\sqrt{N}\begin{bmatrix} \bar{X}_{\cdot\cdot}^{C} - \mu^C \\ \bar{X}_{\cdot\cdot}^{R} - \mu^R \end{bmatrix} = \frac{1}{\sqrt{N}}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\begin{bmatrix} X_{ij}^{C} - \mu^C \\ X_{ij}^{R} - \mu^R \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(0, \Sigma\right), \text{ as } N \to \infty$$

By the continuous mapping theorem:

$$N\left[\left(\bar{X}_{..}^{C}-\mu^{C}\right)^{2}+\omega\left(\bar{X}_{..}^{R}-\mu^{R}\right)^{2}\right]\xrightarrow{\mathcal{D}} Z_{0,C}^{2}+\omega Z_{0,R}^{2}\sim\sigma_{C}^{2}\chi_{C}^{2}(1)+\omega\sigma_{R}^{2}\chi_{R}^{2}(1),$$

where $[Z_{0,C} Z_{0,R}]^{T}\sim N(0,\Sigma)$, and $\chi_{C}^{2}(1)$ and $\chi_{R}^{2}(1)$ are correlated chi-squared distributions with one degree of freedom.

Then, define the $2g$-dimensional statistic: $i=1,2,\ldots,g$, we have

$$\boldsymbol{w}=\begin{bmatrix}\sqrt{n_1}(\bar{X}_{1.}^{C}-\mu_{1}^{C})\\ \sqrt{n_1}(\bar{X}_{1.}^{R}-\mu_{1}^{R})\\ \vdots\\ \sqrt{n_g}(\bar{X}_{g.}^{C}-\mu_{g}^{C})\\ \sqrt{n_g}(\bar{X}_{g.}^{R}-\mu_{g}^{R})\end{bmatrix}$$

Assume W.L.O.G., we notice $\forall i$,

$$\begin{aligned}w_{2i-1} &= \sqrt{n_i}(\bar{X}_{i.}^{C}-\mu_{i}^{C})\\ &= \sqrt{n_i}\frac{\sum_{j=1}^{n_i}(X_{ij}^{C}-\mu_{i}^{C})}{n_i}=\frac{1}{\sqrt{n_i}}\sum_{j=1}^{n_i}(X_{ij}^{C}-\mu_{i}^{C})\\ &= \sqrt{n_1/n_i}\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1}(X_{ij}^{C}-\mu_{i}^{C})+\frac{1}{\sqrt{n_i}}\sum_{j=n_1+1}^{n_i}(X_{ij}^{C}-\mu_{i}^{C})\end{aligned}$$

and similarly,

$$w_{2i} = \sqrt{n_i}(\bar{X}_{i.}^{R}-\mu_{i}^{R})=\sqrt{n_1/n_i}\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1}(X_{ij}^{R}-\mu_{i}^{R})+\frac{1}{\sqrt{n_i}}\sum_{j=n_1+1}^{n_i}(X_{ij}^{R}-\mu_{i}^{R})$$

Thus, we have

$$\begin{aligned} \boldsymbol{w} &= \mathrm{diag}\{\begin{bmatrix} 1 \\ 1 \\ \sqrt{n_1/n_2} \\ \sqrt{n_1/n_2} \\ \vdots \\ \sqrt{n_1/n_g} \\ \sqrt{n_1/n_g} \end{bmatrix}\} \cdot \frac{1}{\sqrt{n_1}} \sum_{j=1}^{n_1} \begin{bmatrix} X_{1j}^C - \mu_1^C \\ X_{1j}^R - \mu_1^R \\ X_{2j}^C - \mu_2^C \\ X_{2j}^R - \mu_2^R \\ \vdots \\ X_{gj}^C - \mu_g^C \\ X_{gj}^R - \mu_g^R \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\sqrt{n_2}} \sum_{j=n_1+1}^{n_2} (X_{2j}^C - \mu_2^C) \\ \frac{1}{\sqrt{n_2}} \sum_{j=n_1+1}^{n_2} (X_{2j}^R - \mu_2^R) \\ \vdots \\ \frac{1}{\sqrt{n_g}} \sum_{j=n_1+1}^{n_g} (X_{gj}^C - \mu_g^C) \\ \frac{1}{\sqrt{n_g}} \sum_{j=n_1+1}^{n_g} (X_{gj}^R - \mu_g^R) \end{bmatrix} \\ &:= \mathrm{I} \cdot \mathrm{II} + \mathrm{III} \end{aligned}$$

By the multivariate CLT,

$$\mathrm{II} \xrightarrow{\mathcal{D}} N(0, \Sigma_{\boldsymbol{w}}),$$

where

$$\Sigma_{\boldsymbol{w}} = \begin{bmatrix} \Sigma & & & \\ & \Sigma & & \\ & & \ddots & \\ & & & \Sigma \end{bmatrix}$$

is the 2g-dimensional block diagonal matrix. Additionally, by the assumption that $n_i/n_1 \to 1, i = 1, \ldots, g$, $\mathrm{I} \to \mathbf{I}_{2g}$ and $\mathrm{III} \xrightarrow{\mathcal{P}} 0$. Therefore,

$$\boldsymbol{w} \xrightarrow{\mathcal{D}} N(0, \Sigma_{\boldsymbol{w}}), \text{ as } n_i \to \infty,\ i = 1, \ldots, g$$

Then, by the continuous mapping theorem, we obtain

$$\begin{aligned} W^c + \omega W^R &= \sum_{i=1}^{g} n_i(\bar{X}_{i\cdot}^C - \mu_i^C) + \omega \sum_{i=1}^{g} n_i(\bar{X}_{i\cdot}^R - \mu_i^R) \\ &\xrightarrow{\mathcal{D}} \sum_{i=1}^{g} Z_{i,C}^2 + \omega \sum_{i=1}^{g} Z_{i,R}^2, \text{ as } n_i \to \infty,\ i = 1, \ldots, g \end{aligned}$$

where $[Z_{i,C}, Z_{i,R}]^T \sim N(0, \Sigma)$. It follows immediately that

$$W^c + \omega W^R \xrightarrow{\mathcal{D}} \sigma_C^2 \chi_C^2(g) + \omega \sigma_R^2 \chi_R^2(g), \text{ as } n_i \to \infty,\ i = 1, \ldots, g$$

Now, notice that under the null hypothesis

$$\begin{aligned}\sqrt{N}(\bar{X}_{..}^{C} - \mu^{C}) &= (\frac{1}{\sqrt{N}})\sum_{i=1}^{g} n_i(\bar{X}_{i.}^{C} - \mu^{C}) \\ &= (\frac{1}{\sqrt{N}})\sum_{i=1}^{g} n_i(\bar{X}_{i.}^{C} - \mu_i^{C}) \\ &= \sqrt{\frac{n}{N}}\sum_{i=1}^{g} n_i(\bar{X}_{i.}^{C} - \mu_i^{C})\sqrt{\frac{n_i}{n}} \\ &= \sqrt{\frac{1}{g}}\sum_{i=1}^{g} n_i(\bar{X}_{i.}^{C} - \mu_i^{C})\sqrt{\frac{n_i}{n}}\end{aligned}$$

and similarly,

$$\sqrt{N}(\bar{X}_{..}^{R} - \mu^{R}) = \sqrt{\frac{1}{g}}\sum_{i=1}^{g} n_i(\bar{X}_{i.}^{R} - \mu_i^{R})\sqrt{\frac{n_i}{n}}$$

where $n = \sum_{i=1}^{g} \frac{n_i}{g}$ denotes the average sample size across groups. This shows that

$$\sqrt{N}[\bar{X}_{..}^{C} - \mu^{C}, \bar{X}_{..}^{R} - \mu^{R}]^{T}$$

converges jointly with $\boldsymbol{w}$ assuming n_i's tend to infinity at the same rate. It follows by the continuous mapping theorem that

$$S^{C} + \omega S^{R} = W^{C} + \omega W^{R} - N[(\bar{X}_{..}^{C} - \mu^{C}) + \omega(\bar{X}_{..}^{R} - \mu^{R})] \tag{2.14}$$

converges.

The last step is to show that $[S^{C}, S^{R}]$ and $[\bar{X}_{..}^{C}, \bar{X}_{..}^{R}]$ are asymptotically independent. Notice obviously that

$$\bar{X}_{i.}^{C} - \bar{X}_{..}^{C},\ \bar{X}_{i.}^{R} - \bar{X}_{..}^{R},\ \bar{X}_{..}^{C},\ \bar{X}_{..}^{R}$$

are all asymptotically normal. It is therefore sufficient to show that $\forall i = 1, \ldots, g$:

$$\begin{aligned}Cov(\bar{X}_{i.}^{C} - \bar{X}_{..}^{C}, \bar{X}_{..}^{C}) &= Cov(\bar{X}_{i.}^{C} - \bar{X}_{..}^{C}, \bar{X}_{..}^{R}) \\ = Cov(\bar{X}_{i.}^{R} - \bar{X}_{..}^{R}, \bar{X}_{..}^{C}) &= Cov(\bar{X}_{i.}^{R} - \bar{X}_{..}^{R}, \bar{X}_{..}^{R}) = 0\end{aligned}$$

The first covariance is computed to be

$$\begin{aligned}
Cov(\bar{X}^C_{i\cdot} - \bar{X}^C_{\cdot\cdot}, \bar{X}^C_{\cdot\cdot}) &= Cov(\bar{X}^C_{i\cdot}, \bar{X}^C_{\cdot\cdot}) - Var(\bar{X}^C_{\cdot\cdot}) \\
&= Cov(\frac{\sum_{j=1}^{n_i} X^C_{ij}}{n_i}, \frac{\sum_{k=1}^{g}\sum_{l=1}^{n_i} X^C_{kl}}{N}) - \frac{\sigma^2_C}{N} \\
&= \frac{1}{n_i N}\sum_{j=1}^{n_i}\sum_{k=1}^{g}\sum_{l=1}^{n_i} Cov(X^C_{ij}, X^C_{kl}) - \frac{\sigma^2_C}{N} \\
&= \frac{1}{n_i N}\sum_{j=1}^{n_i} Cov(X^C_{ij}, X^C_{ij}) - \frac{\sigma^2_C}{N} \\
&= \frac{n_i\sigma^2_C}{n_i N} - \frac{\sigma^2_C}{N} \\
&= 0
\end{aligned}$$

where we have assumed $k = i, l = j$ from the third to the fourth line. Similarly

$$\begin{aligned}
Cov(\bar{X}^R_{i\cdot} - \bar{X}^R_{\cdot\cdot}, \bar{X}^R_{\cdot\cdot}) &= \frac{\sigma^2_R}{N} - \frac{\sigma^2_R}{N} = 0 \\
Cov(\bar{X}^C_{i\cdot} - \bar{X}^C_{\cdot\cdot}, \bar{X}^R_{\cdot\cdot}) = Cov(\bar{X}^R_{i\cdot} - \bar{X}^R_{\cdot\cdot}, \bar{X}^C_{\cdot\cdot}) &= \frac{\sigma_{CR}}{N} - \frac{\sigma_{CR}}{N} = 0
\end{aligned}$$

Thus, $[S^C, S^R]$ as a function of $\bar{X}^C_{i\cdot} - \bar{X}^C_{\cdot\cdot}$ and $\bar{X}^R_{i\cdot} - \bar{X}^R_{\cdot\cdot}$, $i = 1, \ldots, g$ must be independent of $[\bar{X}^C_{\cdot\cdot}, \bar{X}^R_{\cdot\cdot}]$ asymptotically. By the uniqueness of the limit, we must have

$$S^C + \omega S^R \xrightarrow{\mathcal{D}} \sigma^2_C \chi^2_C(g-1) + \omega\sigma^2_R\chi^2_R(g-1). \tag{2.15}$$

as $n_i \to \infty$, $i = 1, \ldots, g$. This completes the proof.

□

Theorem 2.2.2. *Under the null hypothesis, the denominator of our test statistic converges in probability to a linear combination of the population variances for the range and center.*

$$\frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X^C_{ij} - \bar{X}^C_{i\cdot})^2 + \omega\frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X^R_{ij} - \bar{X}^R_{i\cdot})^2 \xrightarrow{\mathcal{P}} \sigma^2_C + \omega\sigma^2_R. \tag{2.16}$$

Proof. Consider the denominator of our test statistic:

$$\frac{1}{N-g}(\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X^C_{ij} - \bar{X}^C_{i\cdot})^2 + \omega\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X^R_{ij} - \bar{X}^R_{i\cdot})^2) \tag{2.17}$$

First, we'll distribute the $\frac{1}{N-g}$. Worth noting is that $N = \sum_{i=1}^{g} n_i$

$$\frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}^{C} - \bar{X}_{i\cdot}^{C})^2 + \omega\frac{1}{N}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}^{R} - \bar{X}_{i\cdot}^{R})^2$$

The two relevant assumptions for convergence are

- Equal variances between the groups (common assumption of ANOVA)
- Under the null hypothesis, all groups have the same mean center and radii

With these assumptions, and using the law of large numbers, each $\bar{X}_{i\cdot}$ converges in probability to the same mean μ. This holds for both the center term and the radius term. We'll work with just the center term to demonstrate briefly. Working with what is inside the summation:

$$(X_{ij}^{C} - \mu^{C} + (\mu^{C} - \bar{X}_{i\cdot}^{C}))^2 = (X_{ij}^{C} - \mu^{C})^2 + 2(X_{ij}^{C} - \mu)\cdot(\mu^{C} - \bar{X}_{i\cdot}^{C}) + (\mu^{C} - \bar{X}_{i\cdot}^{C})^2$$

After distributing the j summation through, and doing some algebra we get:

$$(X_{ij}^{C} - \mu^{C} + (\mu^{C} - \bar{X}_{i\cdot}^{C}))^2 = (X_{ij}^{C} - \mu^{C})^2 + 2(X_{ij}^{C} - \mu)\cdot(\mu^{C} - \bar{X}_{i\cdot}^{C}) + (\mu^{C} - \bar{X}_{i\cdot}^{C})^2$$

$$\frac{1}{N-g}\sum_{i=1}^{g}\sum_{g=1}^{n}(X_{ij}^{C} - \mu^{C})^2 - (\bar{X}_{i\cdot}^{C} - \mu^{C})^2$$

For the second term, the limit as n goes to infinity of X_i is μ, so that term goes to 0. The first term is the definition of variance for X_{ij}, which is identical for all i (equal variance assumption). More explicitly, we can rewrite and simplify our form as follows:

$$\frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}^{C} - \mu^{C})^2 = \frac{1}{\sum_{i=1}^{g} n_i - 1}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}^{C} - \mu^{C})^2 \xrightarrow{\mathcal{P}} \sigma_C^2 \tag{2.18}$$

A process that is essentially identical is used on the radius term, with the only exception being making sure that ω is factored out. Given these proofs, we have:

$$\frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}^{C} - \bar{X_i}^{C})^2 \xrightarrow{\mathcal{P}} \sigma_C^2$$

$$\frac{1}{N-g}\omega\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}^{R} - \bar{X_i}^{R})^2 \xrightarrow{\mathcal{P}} \omega\sigma_R^2$$

Now an application of Slutsky's theorem shows that:

$$\frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}^{C}-\bar{X}_{i\cdot}^{C})^2+\omega\frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}^{R}-\bar{X}_{i\cdot}^{R})^2 \xrightarrow{\mathcal{P}} \sigma_C^2+\omega\sigma_R^2 \tag{2.19}$$

Hence the denominator, under certain conditions, converges to a linear combination of the population variances for the range and center. □

Final Results and Closed Form

With the convergence of both the numerator and the denominator, a straightforward application of Slutsky's theorem shows that:

$$\frac{\sum_{i=1}^{g}(n_i((\bar{X}_{i\cdot}^{C}-\bar{X}_{\cdot\cdot}^{C})^2+\omega(\bar{X}_{i\cdot}^{R}-\bar{X}_{\cdot\cdot}^{R})^2))/(g-1)}{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}^{C}-\bar{X}_{i\cdot}^{C})^2+\omega(X_{ij}^{R}-\bar{X}_{i\cdot}^{R})^2/(N-g)} \xrightarrow{\mathcal{D}} \frac{\sigma_C^2\chi_C^2(g-1)+\omega\sigma_R^2\chi_R^2(g-1)/(g-1)}{\sigma_C^2+\omega\sigma_R^2}$$

or equivalently

$$\frac{\sum_{i=1}^{g}(n_i((\bar{X}_{i\cdot}^{C}-\bar{X}_{\cdot\cdot}^{C})^2+\omega(\bar{X}_{i\cdot}^{R}-\bar{X}_{\cdot\cdot}^{R})^2))/(g-1)}{\sum_{i=1}^{g}\sum_{j=1}^{n}(X_{ij}^{C}-\bar{X}_{\cdot\cdot}^{C})^2+\omega(X_{ij}^{R}-\bar{X}_{\cdot\cdot}^{R})^2/(N-g)} \xrightarrow{\mathcal{D}} \frac{\sum_{i=1}^{g-1}\sigma_C^2\chi_C^2(1)+\omega\sigma_R^2\chi_R^2(1)/(g-1)}{\sigma_C^2+\omega\sigma_R^2}$$

This is the limiting distribution that we will use for our hypothesis tests. Of note, there are other asymptotically equivalent choices for degrees of freedom, but the classical ANOVA choices avoid bias and naturally improve small sample performance. Additionally, each interval is still considered an independent piece of information. We lose 1 piece of information to calculate the grand interval mean, and g pieces of information to calculate each groups interval mean, further justifying the use of the same degrees of freedom from classical ANOVA. This distribution is fundamentally just a linear combination of correlated chi-square distributions. χ_C^2 and χ_R^2 have correlation ρ^2, where ρ is the correlation between $\bar{X}_{i\cdot}^{C}-\bar{X}_{\cdot\cdot}^{C}$ and $\bar{X}_{i\cdot}^{R}-\bar{X}_{\cdot\cdot}^{R}$ (Theorem 4.0.1). The derivation of a closed form for this kind of distribution has been done before [23], using the gamma and hypergeometric functions. Due to the complicated nature of this closed form, we will instead simulate this distribution, and further validate our theoretical results with various simulated examples.

CHAPTER 3
SIMULATION AND REAL DATA

3.1
Simulation

To strengthen our theoretical results, we've also simulated data-sets and their theoretical limiting null distributions. The simulated distributions below are constructed using 10000 simulated data sets. From left to right, the approximate individual group size is increased to show how the simulated distribution tends to its limiting distribution. The limiting distribution is constructed using 100000 samples and overlaid onto the graphs in dark red.

3.1.1
Normal Data

Our first simulation is the simplest, where we use interval data that is simulated with a normally distributed center and range (3 groups). Explicitly, the center is distributed N(50, 10), while the radius is distributed N(20, 2). There is a weak positive correlation between the center and range in our simulation ($\rho = 0.2$), and we will use an omega parameter of 1, essentially weighting the center and radius term the same. The distribution for the numerator of our test-statistic is shown first.

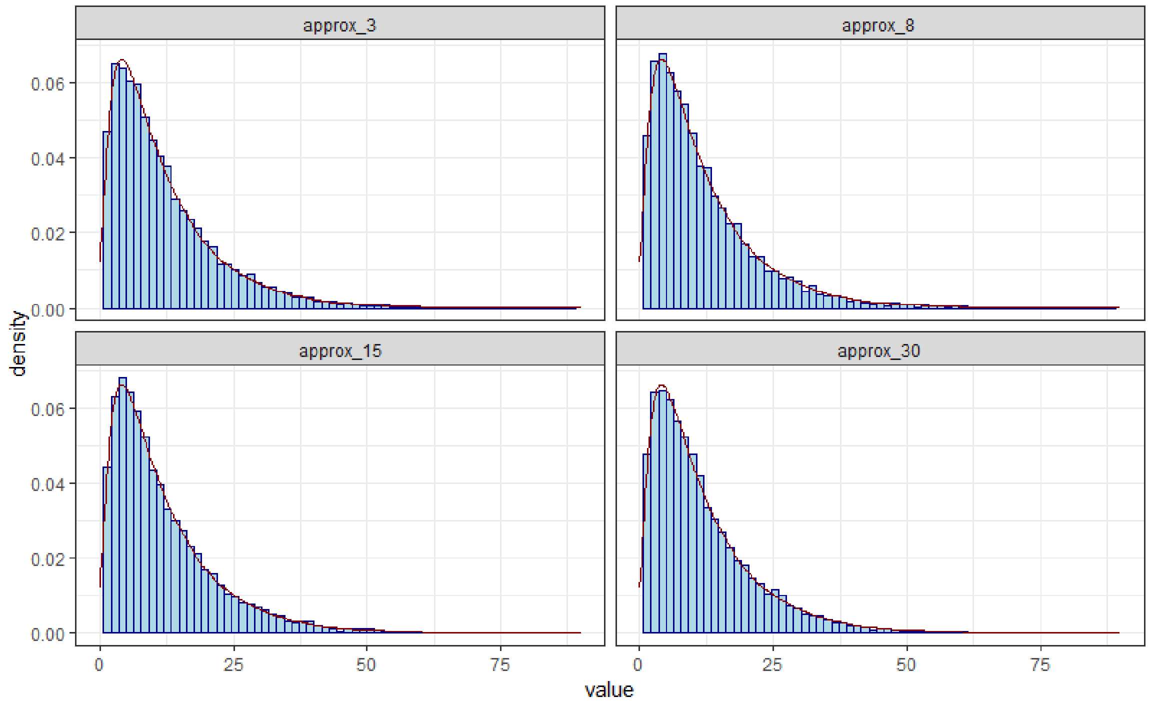

Figure 3.1: Normal Numerator

The distribution of our numerator is essentially identical to it's limiting distribution even for smaller individual group sizes. This behavior is in line with our expectation, as the pieces of our numerator are precisely normal regardless of sample size. A table for the results of the denominator is shown below.

$n_i \approx$	3	8	15	30	∞
Denominator Average	11.993	11.975	11.954	11.993	12
Denominator Standard Deviation	5.492	3.161	2.278	1.495	0

Table 3.1: Denominator Values - Normal

Here, we see the behavior of our denominator as N grows larger. We get less and less variance in our simulations. This is again in line with our expectations. Finally, the distribution for the full test statistic is shown below

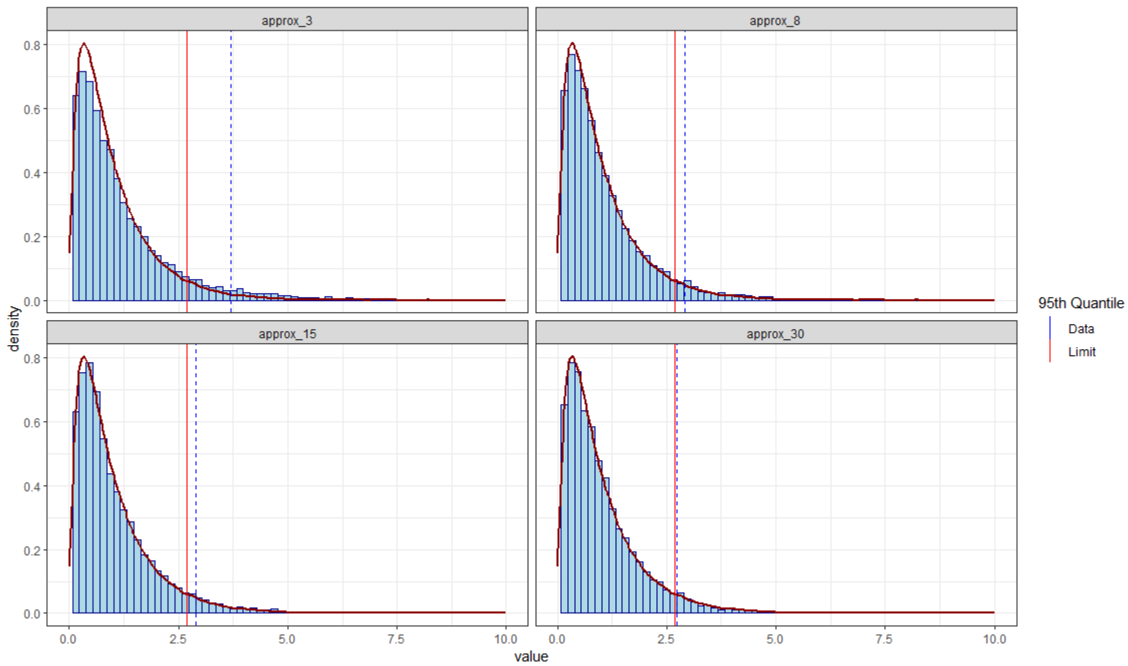

Figure 3.2: Normal Full Statistic

These distributions are quite close to each other, but the consequence of such a individual group sample size can be seen in the top left. Additionally, different simulated critical values are shown in the table below.

Critical Values	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	2.671	3.689	6.238
$n_i \approx 8$	2.270	2.923	4.664
$n_i \approx 15$	2.243	2.898	4.553
$n_i \approx 30$	2.145	2.730	4.208
Theory	2.123	2.703	4.125

Table 3.2: Critical Values - Normal

Here we see the critical values for every sample size behaving relatively similarly. The $\alpha = 0.01$ critical value for $n_i = 3$ is the furthest away from it's theoretical value, but this is a fairly natural result of the lower sample size and high α for that case. This deviation may lessen by simulating more data. A similar table, where # of rejections out of 10000 samples, is shown below

# Rejections	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1548	972	375
$n_i \approx 8$	1163	625	167
$n_i \approx 15$	1132	619	158
$n_i \approx 30$	1018	516	108
Limit	1000	500	100

Table 3.3: Rejection Table - Normal

The rejections show the somewhat unreliable nature of our test at very small sample sizes. We are generally not conservative enough, rejecting the null hypothesis more than we should for what the percentile of the data suggests. By the time we get to 15 though, we are quite close to what we should be.

3.1.2
Uniform Data

Moving to a slightly more complicated example, our data for the next simulation comes from a uniform distribution, where we are comparing four groups. Our center follows a uniform distribution with minimum value 5 and maximum values 15. Our range also follows a uniform distribution with minimum 1 and maximum 2. Our center and range for this sample are heavily correlated ($\rho = 0.8$). Additionally, we are using an omega of 0.5, essentially underweighting the radius relative to the center. Again, first the distribution of the numerator

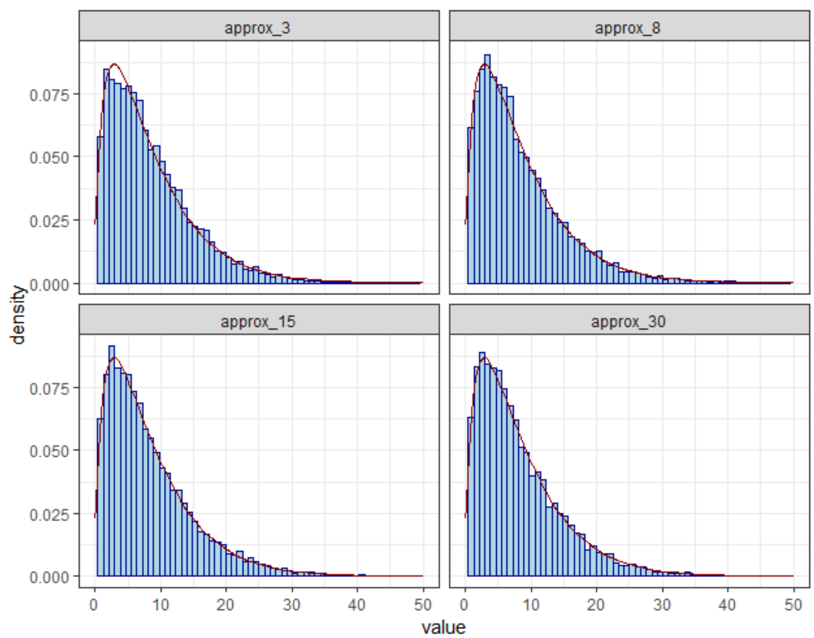

Figure 3.3: Uniform Numerator

Though it's somewhat hard to see, we are more prone to outliers in the smaller sample cases, which does stand to reason. By the time we reach a sample size of 50, however, the limiting distribution and the simulated distribution behave almost identically.

$n_i \approx$	3	8	15	30	∞
Denominator Average	8.400	8.337	8.376	8.370	8.375
Denominator Standard Deviation	3.015	1.542	1.057	0.679	0

Denominator Values - Uniforml

The behavior is exactly the same as the normal case, as expected. Finally, the distribution for the full test statistic is shown below

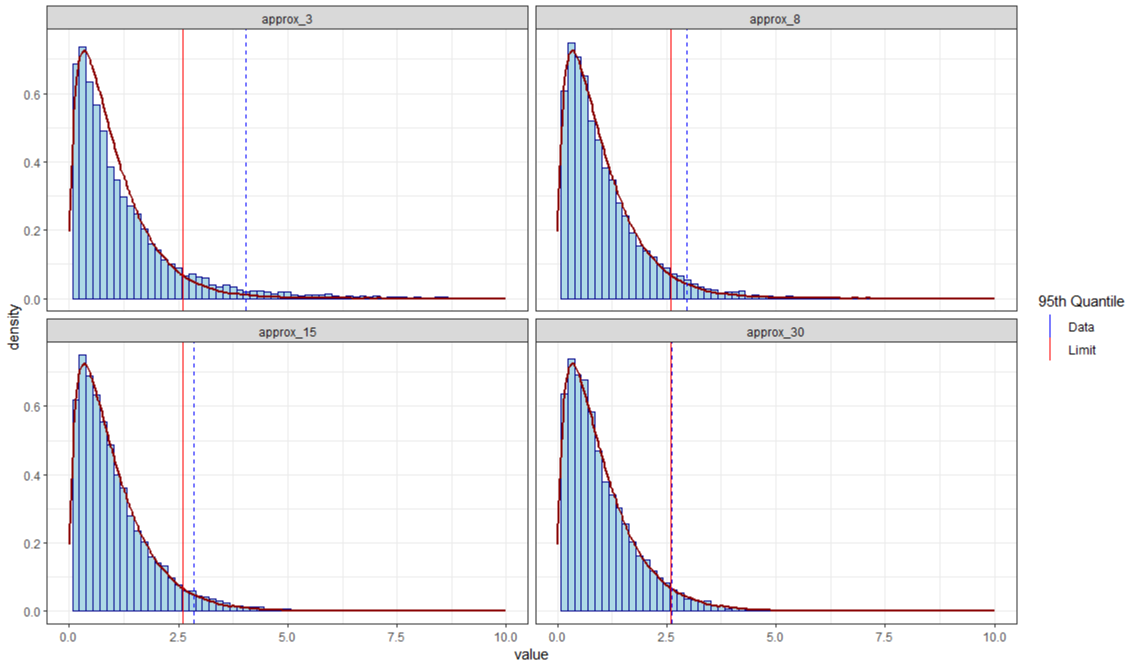

Figure 3.4: Uniform Full Statistic

This is a similar case to the numerator, but the outliers are now even more apparent, and we again have a heavier right tails. These things have a fairly dramatic effect on the critical values, as shown in the table below.

Critical Values	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	2.851	4.041	8.106
$n_i \approx 8$	2.324	2.969	4.722
$n_i \approx 15$	2.215	2.848	4.290
$n_i \approx 30$	2.106	2.609	3.775
Theory	2.081	2.595	3.794

Table 3.4: Critical Values - Uniform

There are fairly major deviations from the critical values given in the limiting distribution, although only really in the $n_i \approx 3$ case.

# Rejections	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1713	1177	564
$n_i \approx 8$	1292	746	241
$n_i \approx 15$	1189	646	166
$n_i \approx 30$	1039	514	97
Limit	1000	500	100

Table 3.5: Rejection Table - Uniform

Same thing here as the normal case, we only really get reliable starting around $n_i \approx 15$. This example in particular is highly prone to outliers in the smaller sample cases.

3.1.3
Gamma Data

For another more complicated example, we will use gamma distributed data. There are now 5 groups, a correlation of 0.5, and an omega of 1. The center is gamma distributed with mean parameter 5 and dispersion parameter 2, while the radius is gamma distributed with mean parameter 10 and dispersion parameter 1. Numerator distribution shown below.

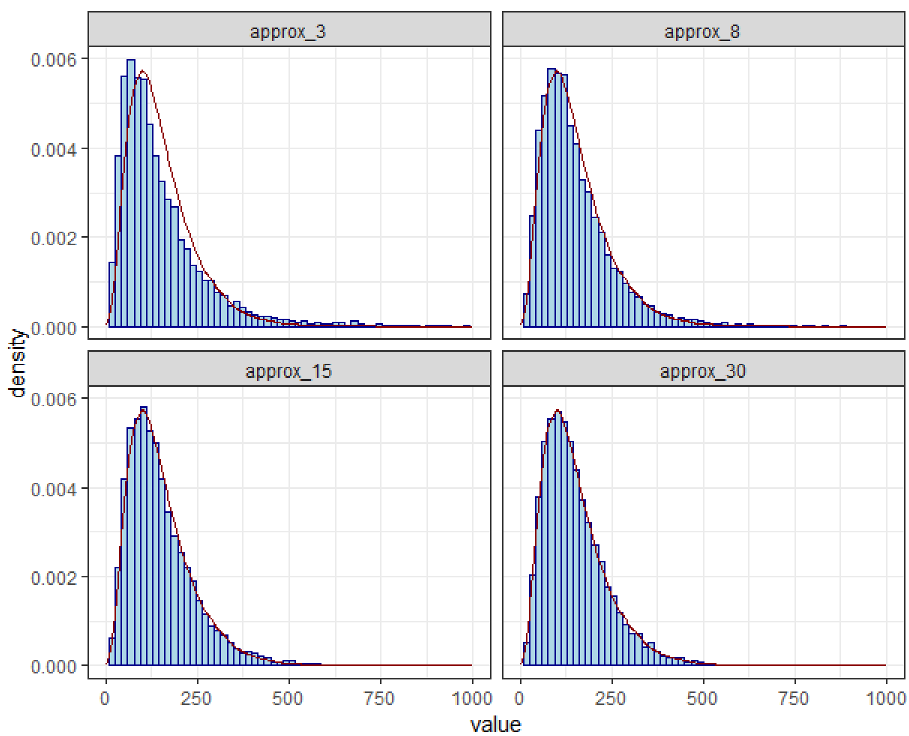

Figure 3.5: Gamma Numerator

Here, the numerators are pretty distinct from one another, with the small sample case having an earlier mode and a slower decay than the large sample case and the limiting distribution. However, it only takes a sample size of about 15 to be very close to the limiting distribution. Denominator results:

$n_i \approx$	3	8	15	30	∞
Denominator Average	150.121	149.161	150.061	150.332	150
Denominator Standard Deviation	95.445	59.371	44.066	30.301	0

Table 3.6: Denominator Values - Gamma

This is a much higher variance example than the previous ones, but the same general idea holds.

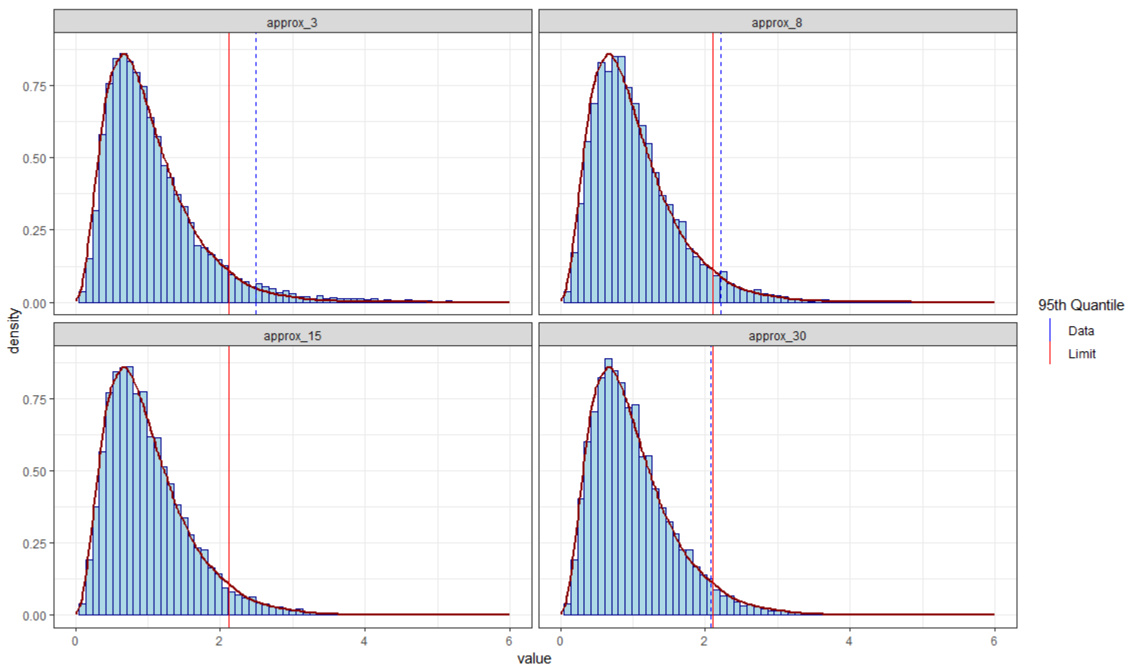

Figure 3.6: Gamma Full Statistic

The distributions on top are somewhat distinct from the ones on bottom. They have heavier right tails and are more prone to outliers. Critical values are listed below.

Critical Values	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1.945	2.487	4.082
$n_i \approx 8$	1.816	2.211	3.097
$n_i \approx 15$	1.776	2.124	2.940
$n_i \approx 30$	1.773	2.080	2.880
Theory	1.776	2.111	2.855

Table 3.7: Critical Values - Gamma

These critical values are in line with what we would expect from looking at the graphs. Generally the critical values for the simulated data are still larger than that of the theoretical critical values. They do get much closer as the sample size increases though.

# Rejections	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1299	787	317
$n_i \approx 8$	1080	594	158
$n_i \approx 15$	1001	508	119
$n_i \approx 30$	993	469	106
Limit	1000	500	100

Table 3.8: Rejection Table - Gamma

The rejections are similar, but are even more accurate than the previous cases by the time we get to $n_i \approx 15$

3.1.4

Poisson Data

Our next example will extend these results to discrete data. Our center will be Poisson distributed with $\lambda = 1$, while our radius will be Poisson distributed with $\lambda = 15$. Here, we'll use a strong negative correlation of -0.75, an omega of 0.9, and 5 groups.

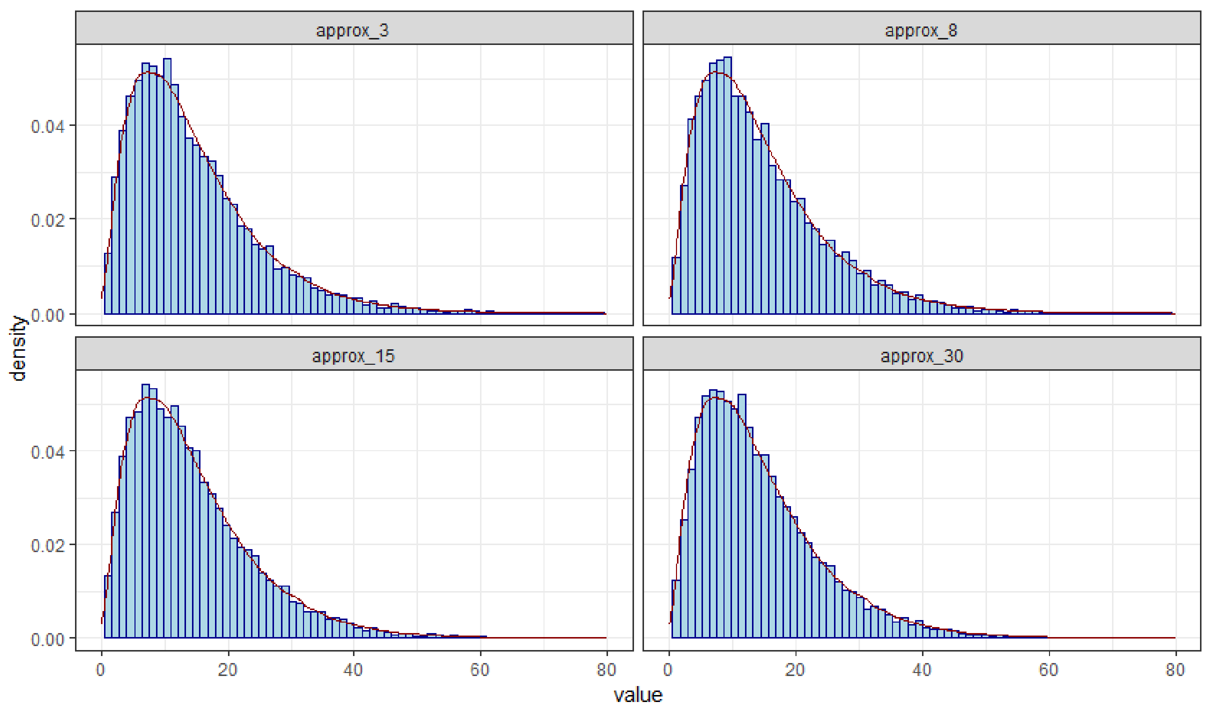

Figure 3.7: Poisson Numerator

The distribution of the numerators look fairly similar, with the main difference between them being the somewhat smoother nature of the larger sample cases, and fewer extreme values.

Denominator results:

$n_i \approx$	3	8	15	30	∞
Denominator Average	14.492	14.519	14.478	14.476	14.5
Denominator Standard Deviation	5.852	3.383	2.428	1.630	0

Table 3.9: Denominator Values - Poisson

The same general idea continues to hold here.

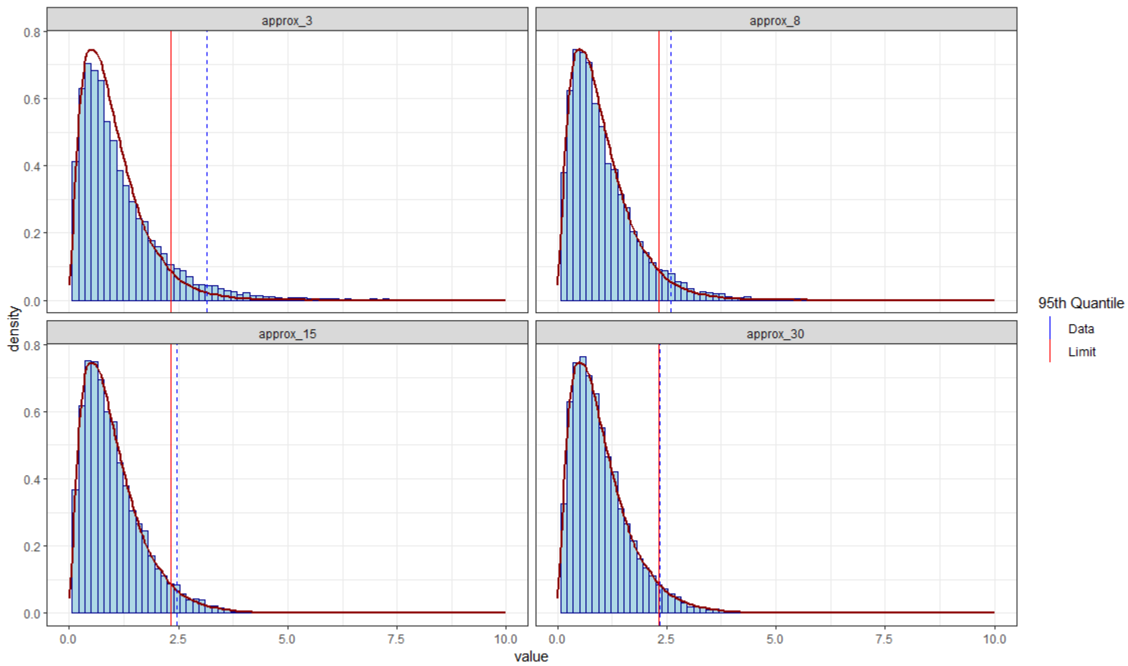

Figure 3.8: Poisson Full Statistic

For our full statistic, the small sample cases are distinctly heavier tailed than our "approx_30" and our limiting case. This is also clear in the critical values table, shown below:

Critical Values	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	2.409	3.145	5.216
$n_i \approx 8$	2.086	2.603	3.739
$n_i \approx 15$	1.977	2.455	3.425
$n_i \approx 30$	1.926	2.353	3.295
Theory	1.917	2.329	3.252

Table 3.10: Critical Values - Poisson

These heavy tails are apparent in the consistent overestimation of our critical values. They do all remain relatively close to each other, but there would be cause for some hesitancy in evaluating the $\alpha = 0.01$ critical value for small sample sizes.

# Rejections	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1659	1083	451
$n_i \approx 8$	1247	736	209
$n_i \approx 15$	1096	600	137
$n_i \approx 30$	1010	521	107
Limit	1000	500	100

Table 3.11: Rejection Table - Poisson

Once again, it takes until about an individual group sample size of about 15 to behave well enough.

3.1.5

Bernoulli Center, Gamma Radius

Our second to last example will continue to extend these results to discrete data. Our center will be Bernoulli distributed with $p = 0.6$, while our radius will be Gamma distributed with shape parameter 0.5 and rate parameter 1. Here, we'll use a positive empirical correlation of 0.25, an omega of 0.8, and 5 groups.

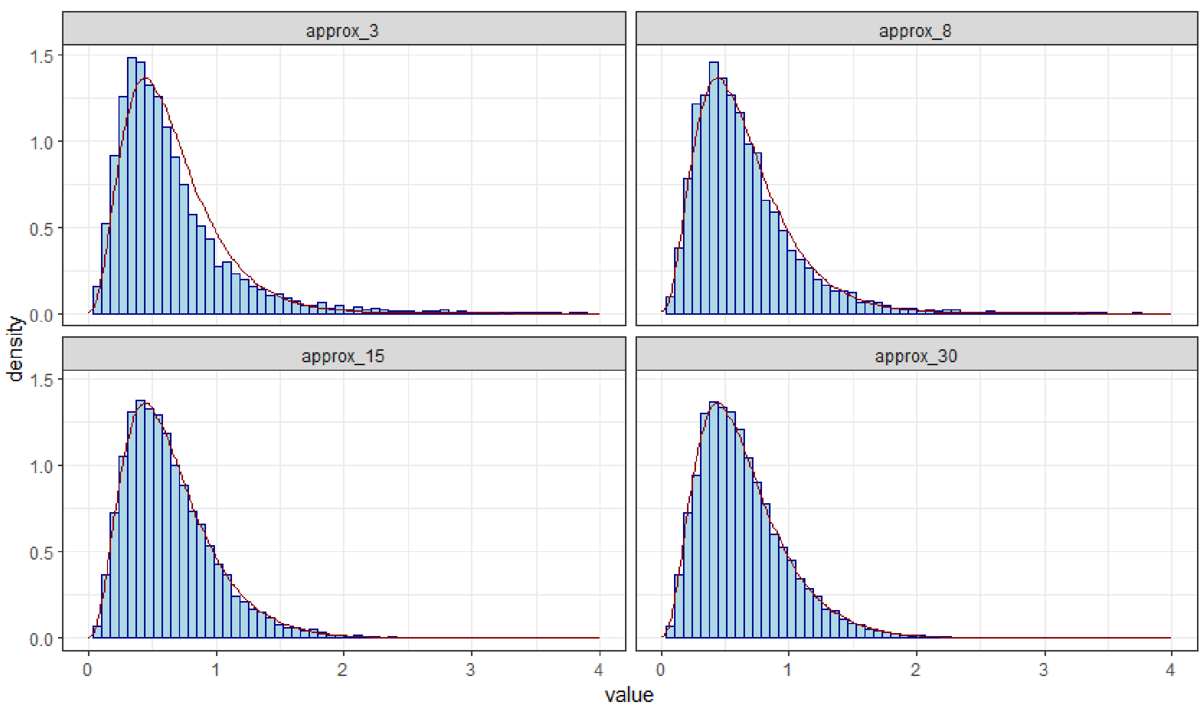

Figure 3.9: Bernoulli Numerator

The main difference between the numerators is an earlier peak and slightly slower drop off in the small sample case.

Denominator results:

$n_i \approx$	3	8	15	30	∞
Denominator Average	0.638	0.637	0.643	0.641	0.65
Denominator Standard Deviation	0.380	0.234	0.176	0.119	0

Table 3.12: Denominator Values - Bernoulli, Gamma

We get fairly accurate with our variance estimation quite quickly.

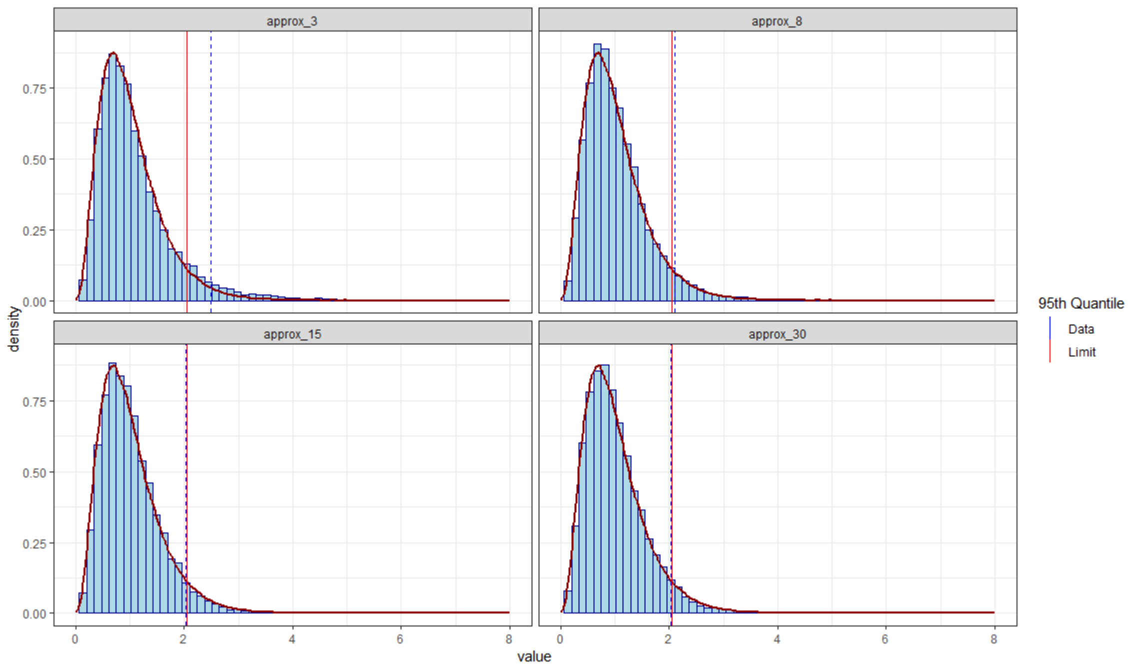

Figure 3.10: Bernoulli Full Statistic

The full statistic dies down slower for small samples. The critical values and rejection tables are shown below:

Critical Values	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1.994	2.484	3.892
$n_i \approx 8$	1.757	2.109	2.944
$n_i \approx 15$	1.726	2.023	2.776
$n_i \approx 30$	1.726	2.027	2.747
Theory	1.738	2.058	2.762

Table 3.13: Critical Values - Bernoulli, Gamma

# Rejections	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1424	920	358
$n_i \approx 8$	1040	556	140
$n_i \approx 15$	976	462	106
$n_i \approx 30$	976	460	100
Limit	1000	500	100

Table 3.14: Rejection Table - Bernoulli, Gamma

In this case, our values and rejections are actually generally slightly more conservative for the $n_i \approx 15, 30$ cases, while the same pattern as before holds for the other sizes.

3.1.6

A Counterexample of Cauchy Data

Our final simulation provides an example of how our theorem breaks when assumptions are not met. The Cauchy distribution lacks a mean, a finite variance, and even a well defined notion of correlation. All of these things pose significant problems for our theoretical results, and thus are a great way to check its ability to fail. Our simulation uses the Cauchy for both the center and radius. Our center uses location parameter 5 and scale parameter 10, while the radius uses a location parameter of 100, and a scale parameter of 2. We also simulate these using an intended ρ of 0.5. Our limiting distribution cannot be simulated with a theoretical mean, variance, and correlation, so we instead use the sample results for these parameters to simulate our limiting distribution. This obviously isn't perfectly reliable, but is as close as we can get to trying to capture the limiting behavior with our theory. For this example, we also use $\omega = 1$ and $g = 5$. The results for our full statistic are shown in Figure 3.11.

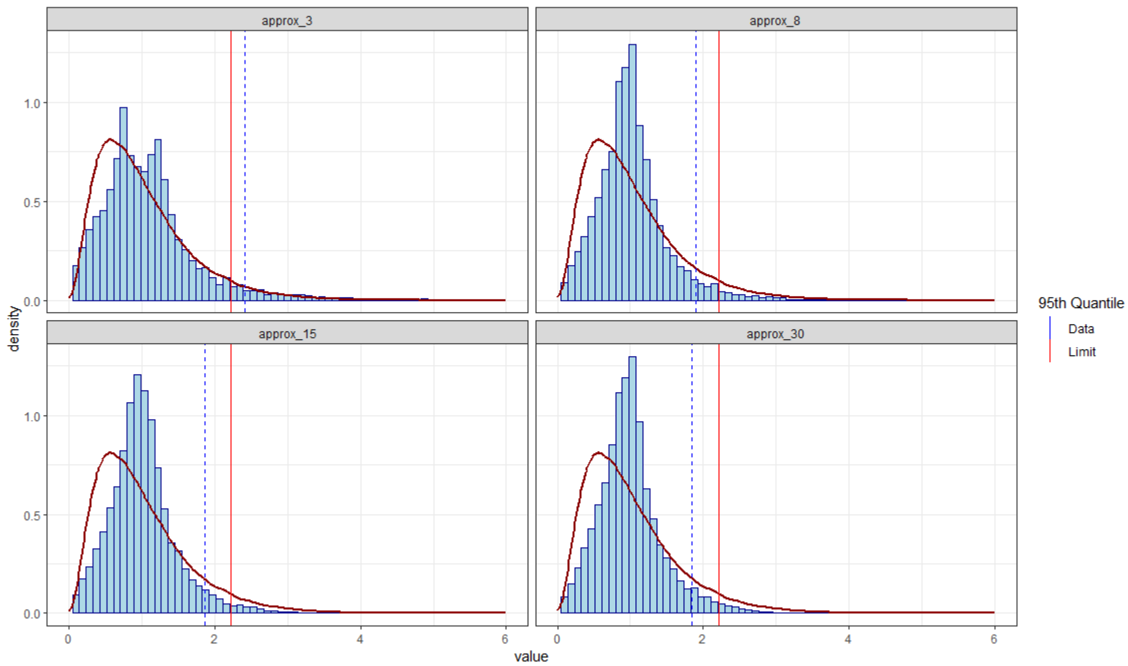

Figure 3.11: Cauchy Full Statistic

The behavior of our simulated datasets/statistics is clearly not in line with the theoretical results, and there is no sign of convergence to our theoretical results as N increases, unlike our other examples. While there are convergence results for the Cauchy distribution, the large sample behavior is totally unable to be captured by the central limit theorem, and therefore our results. Shown below are the results for critical values and number of rejections.

Critical Values	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	1.748	2.194	3.610
$n_i \approx 8$	1.583	1.905	2.696
$n_i \approx 15$	1.560	1.905	2.537
$n_i \approx 30$	1.550	1.851	2.426
Theory	1.843	2.221	3.045

Table 3.15: Critical Values - Cauchy

# Rejections	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_i \approx 3$	847	486	171
$n_i \approx 8$	565	244	47
$n_i \approx 15$	514	216	35
$n_i \approx 30$	510	191	16
Limit	1000	500	100

Table 3.16: Rejection Table - Cauchy

Our simulated results are not consistent with our theoretical results, which is again in line with what we would expect.

3.1.7

Simulation Summary

The results from simulation validate our test. Even with distributions that are fairly different from the normal distribution, we appear to converge quickly to the theoretical limiting distribution. Explicitly, it appears to take until an individual group sample size of about 15 to have results that are quite reliable, but we often have consistent results even around a sample size of 8. 3 is generally too small for the central limit theorem to take over, which is in line with what we would expect. These small sample results demonstrate the capacity for IANOVA to be used on smaller data sets. The primary issue to be aware of is that the mistakes made in small sample cases are generally not being conservative enough, i.e. the test finds significant results more often than the test should find significant results. Ideally, we'd want to be making the opposite mistake, but practically, it just means that we should be cautious when analyzing data with group sample sizes below 8, or an overall sample size of less than 30. The limiting behavior is consistent with what was theorized, with our large sample cases being very close to the theory. In cases where our results should fail, they do, which further strengthens the validity of our theory.

3.1.8

Package

The simulation work above, and more generally the IANOVA procedure, have been documented and organized into an R package currently available on Github. All figures and results are entirely reproducible with the scripts provided. Functions are provided within the package for ease of use. `sim_data` allows the user to simulate interval datasets from a few distributions, and returns them in a way organized such that `observe_num` and `observe_den` can return the numerator and denominator of the test statistic. Additionally `sim_theory_num` and `sim_theory_den` allow the user to get the theoretical cutoffs for their data, which is useful when analyzing real data. The general procedure for running IANOVA on real data is as follows:

1. Format the data to be a list, where each entry contains the interval data for one group (column 2 is centers, column 3 is radii).

2. Get the test statistic from `observe_num` divided by `observe_den`.

3. Find the variance of the center and range (may be previously known, assumed, or empirical), as well as the correlation.

4. Use these to get the theoretical limiting distribution from `sim_theory_num` divided by `sim_theory_den`.

5. Compare your test statistic to the quantile of the limiting distribution appropriate for your significance level.

A test statistic greater than your selected quantile means statistical significance at the $1-$quantile level. This would provide evidence that your interval means are different across groups. Alternatively, a test statistic less than your selected quantile means a failure to reject the null hypothesis. In that case, there would not be sufficient evidence of a difference in interval means across groups.

3.2

Real Data Analysis

3.2.1

Data

One key building design criteria is the weight, or load, of settled snow on the roof of a structure. Current design snow load requirements for the United States are introduced in Bean et. al [9]. These design requirements rely on an estimate of the ratio between the maximum observed ground snow load and the maximum observed roof load, a ratio referred to hereafter as GR. This GR information is obtained from a series of Canadian snow surveys. As might be expected, the distribution of snow on a roof is rarely uniform, with blowing snow and/or complex roof geometries creating pockets of heavier than average, or lighter than average, snow on the roof. This motivates consideration of GR as an interval-valued ratio, denoted [GR_min, GR_max], or equivalently [GR_C, GR_R]. This interval is constructed by determining the maximum max snow load on a roof divided by the maximum ground load across the season. The minimum is constructed similarly, using the smallest snow load measurement on the day that the maximum average roof load measurement was observed. This interval characterization provides valuable information about the unbalanced nature of the snow load, which may result in different failure modes than would happen in a uniform loading scenario. In this paper, we are interested in determining if the GR interval ratio remains constant across the geographical regions represented in the previously cited Canadian snow surveys [24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. This question is explored using two different partitions of the data and results are compared to results obtained using standard ANOVA using the standard GR measurements.

Figure 3.12 shows a map of the observations spread throughout Canada. In total, there are 81 observations in our data set. The center is represented by the color of the dot, while the range is the size of the dot. Due to the imprecise geographic nature of the data, many points in major cities overlap one another, but this graphic still provides information about the general location of observations. Of note, but difficult to see in this graphic, our center and range are positively correlated, with $\rho = 0.821$.

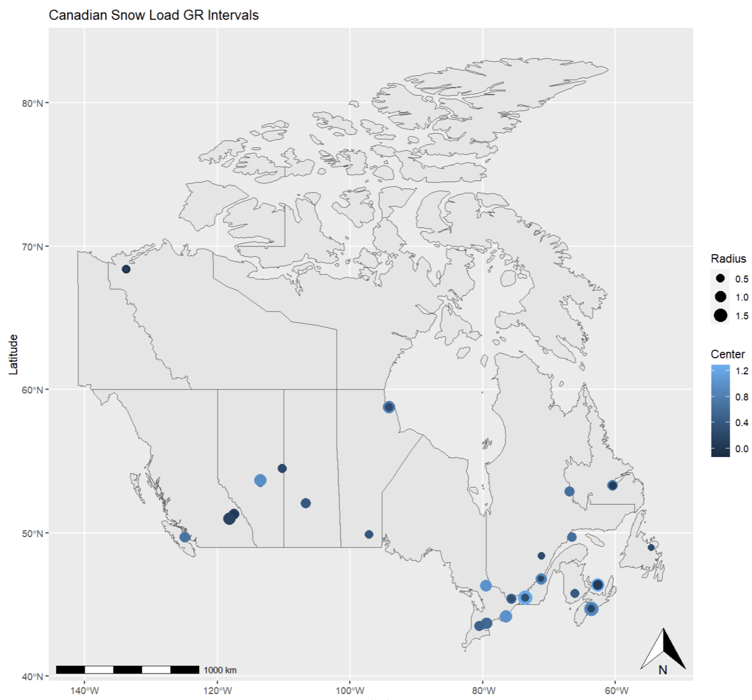

Figure 3.12: Locations of Observations

The distribution of points show a clear lack of balance between the northern and southern provinces. This is reflective of the population distribution in Canada, and where we're often interested in studying, but this isn't to say observations in the northern portions of Canada wouldn't be worth further study.

3.2.2

Political Partition

We first partition observations by province. Specifically, we separate our observations into five groups: British Columbia, The Great Plains Provinces (AB, SK, MB), Ontario, Quebec, and

The Maritime Provinces (NB, PEI, LB, NS, NF). This provides a decent balance for our groups, but it does omit five observations in the Northwest Territories. If we only had the provinces of these observations, this would probably be the most reasonable way to separate them. This political split is mostly motivated by accessibility; it's easy to see why these observations should go together. Legislative decisions are often made at this scale, which would be useful for updating building codes to reflect differences in GR intervals between provinces. The shortcoming, of course, is that a partition along these lines may fail to capture differences in GR interval that occur due to ecological mechanisms. This shortcoming will be discussed in the next section, but this remains a justifiable starting point. The box plots for the province groups are shown in Figure 3.13.

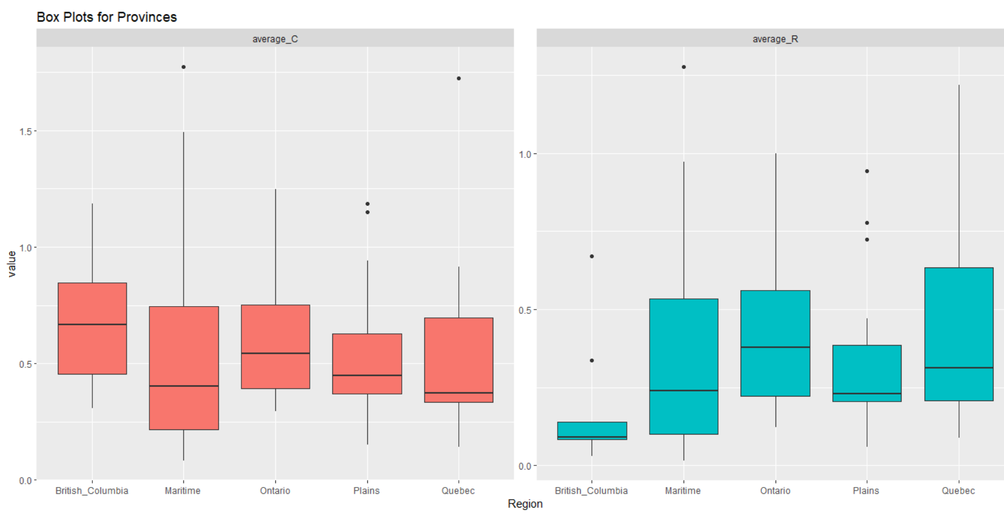

Figure 3.13: Box-plot of Province Partitioning

It's difficult to discern any patterns in the box-plots, and our analysis also shows a lack of any statistically significant differences in the groups. Our test-statistic for this partition is 0.730, whereas the 95th quantile for our simulated limiting distribution is 2.263. This shows insufficient evidence to suggest that the province has an effect on the GR interval. There are potentially meaningful splits along province lines, but this partition does not appear to have any impact.

3.2.3

Ecological Partition

Our second partition separates observations based upon the Level I North American Ecoregions. Figure 3.14 shows the observations in each region [34]. Of note, due to a lack of observations in some regions, 6 and 7 are combined, as are 9 and 5. 2 an 4 are also combined in the graphic.. Our groups, broadly speaking, correspond to west coast forests and mountains, tundra/taiga, eastern temperate forests, and the plains/northern forests regions.

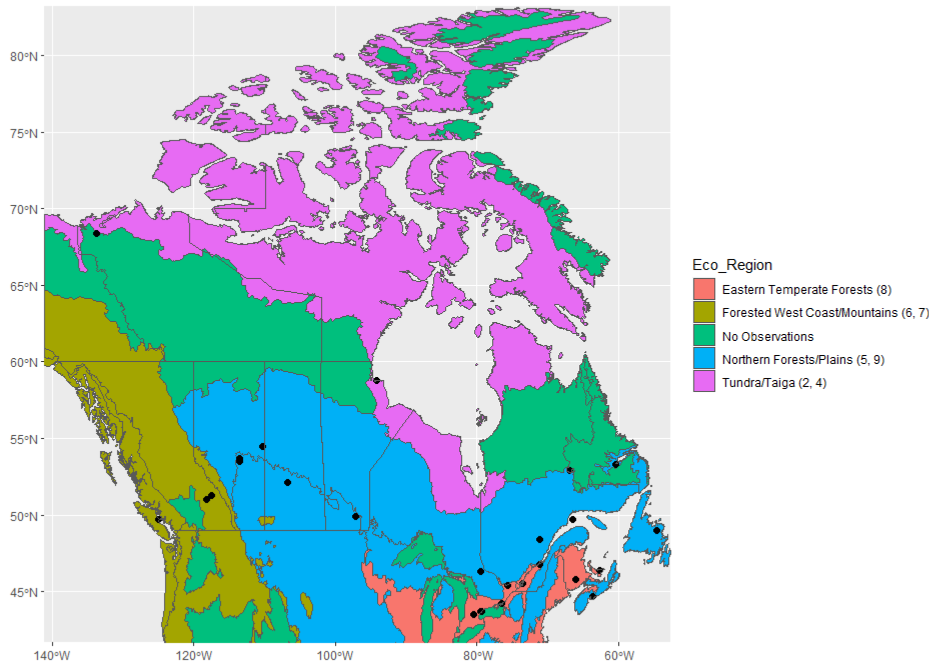

Figure 3.14: Observations within their Eco-Regions

Splitting along these lines is more consistent with the problem at hand. The ecological conditions are going to generally be more reflective of how snowfall works in a given region. Any of the mechanisms that promote a roofs propensity for drifting should be better captured by the eco-region the roofs fall in. Continuing with our analysis, Figure 3.15 shows the box-plots for our groups partitioned by eco-region.

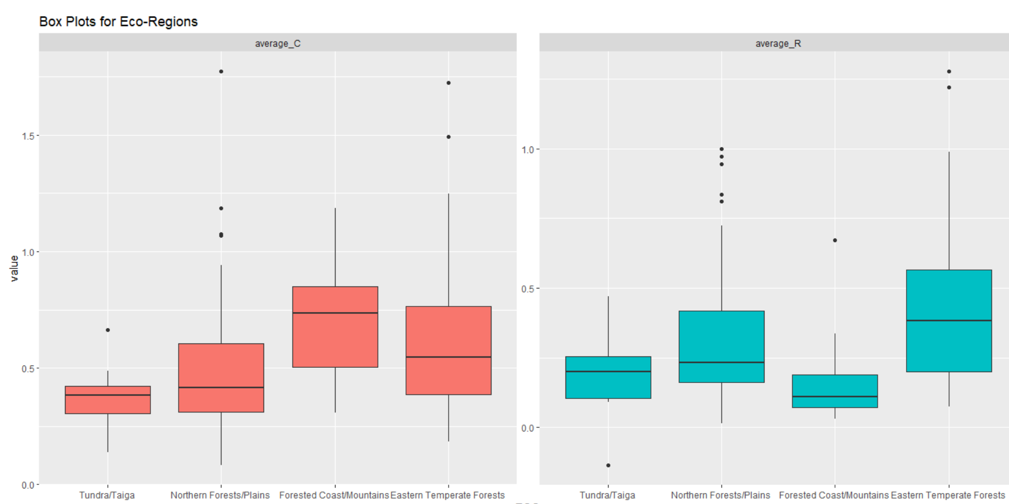

Figure 3.15: Box-plot of Eco-region Partitioning

While there are again no clear pattern to the data, it looks possible that the Eco-Regions 6 and 7 have a higher center and lower range than average. Additionally, while we are not testing for it, the variance of Region 2 appears to be lower than the other groups for both center and range. This more meaningful partition returns more meaningful results. Our null hypothesis for this test is that GR interval mean is constant across regions, while our alternative is GR interval mean is non-constant across regions. We use an $\alpha = 0.05$ cutoff for this test. Our test-statistic for this partition is 2.48, whereas the 95th quantile for our simulated limiting distribution is 2.46. For our chosen α cutoff, these results are statistically significant. There is evidence to suggest GR_interval is non-constant across eco-regions. This is quite close to our cut-off, but this is likely due to relatively few observations in our different looking groups (6 and 7), vs our similar groups (8, 5 and 9). It's possible that clearer patterns would emerge with a larger sample size, but these results show that the GR interval with respect to Eco-Region would be worth investigating further.

3.2.4

Classical ANOVA

We now move to a classical ANOVA on the averages for GR. The average is point valued, and thus we are able to use classical statistical methods. Box plots for both the political and eco-region partition are shown below in Figures 3.16 and 3.17. a

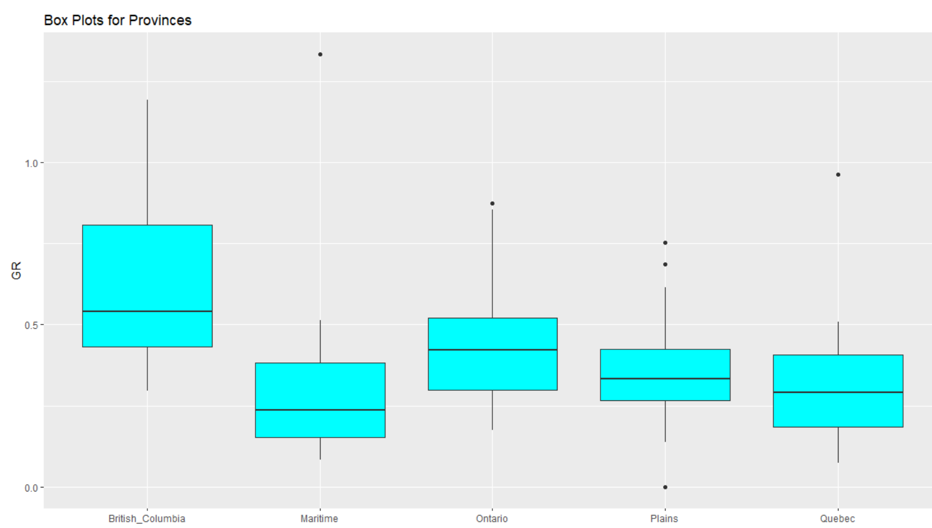

Figure 3.16: Box-plot of Political Partitioning (Point-Valued)

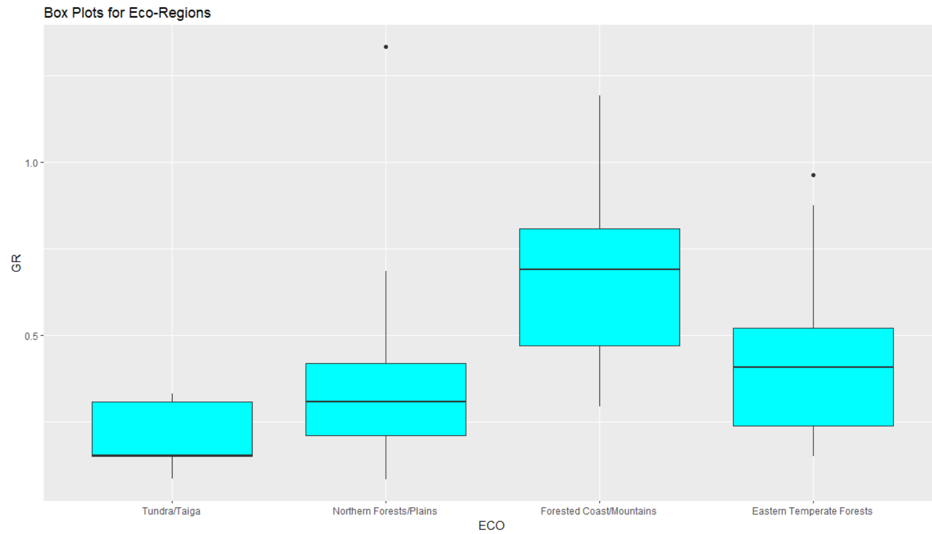

Figure 3.17: Box-plot of Eco-region Partitioning (Point-Valued)

In both cases, there does appear to potentially be a difference in the average based on the groups. The differences in the political groups seem to be driven by British Columbia, while the differences in the ecological groups are driven by the group containing eco-region 6 and 7. Of note, eco-region 6 and 7 most strongly correspond to British Columbia, so this effect looks to be occurring around the same place. Running the ANOVA, we have significant results for *both* partitions. Our political split has an F-statistic of 3.982 and a p-value of 0.005, which are fairly strong results indicating that the province has an effect on the average GR. Additionally, our ecological split has an F-value of 6.180 and a p-value of 0.0008, which are highly significant results. It's possible that the province partition is capturing some of the eco-region effect, as there is a fair amount of overlap between the two. Both these results suggest that assuming GR is constant across may regions may not be well-founded in practice. The primary driver in the higher power of classical ANOVA relative to IANOVA is likely the amount of data we had for each test. We have a larger sample size of GR than we do GR intervals, stemming from average data being more available than min-max data. In particular, the greater number of observations in the regions that seem notably different (Forested Coast/Mountains), will naturally make those results more significant.

3.2.5 Discussion

When we pay attention to the ecological factors surrounding an observation, we become more likely to reject the null hypothesis. This holds for both IANOVA and classical ANOVA, where the consistency is promising for our new method. Classical ANOVA is quite good at detecting significant results for small samples, whereas IANOVA's use of an asymptotic limiting distribution means that it can run into problems with small sample size. We do not have significant results for the province partition in the interval-valued case, while we do for the point-valued case. While we do have significant results for the Eco-Region based partition, they are quite close to our $\alpha = 0.05$ cutoff. This could be due to a lack of observations in some of our groups, which weakens the validity of using an asymptotic limiting distribution. This example, however, really highlights the strength of IANOVA. If we were just interested in how GR varies across regions, the ANOVA on averages would be sufficient. If we are interested in how both GR *and* propensity for drifting vary across regions, then the interval-valued approach is quite useful. Again, this provides extra information on the stresses a roof might actually incur, given that some areas of the roof will experience greater pressure. The use of IANOVA allows us to use this extra information, which can help illuminate distinct and important results. The results of both analyses indicate a potential need to further investigate the variance of GR across regions, and the drifting effect across regions.

CHAPTER 4
CONCLUDING REMARKS

IANOVA provides another tool for interval valued data analysis, allowing for comparisons across greater than two groups of intervals. This thesis provides theoretical justification for IANOVA, proofs of important asymptotic results, simulations verifying those results, and a real data problem where IANOVA provides some illuminating results.

The structure of interval data and the approach of our test lead to a fairly natural comparison to Multiple ANOVA (MANOVA). Our interval data is represented by centers and radii, so we could consider using MANOVA with center/radii as our dependent variables. There are a number of similarities between the two procedures, and a few key differences. The assumptions between the two are nearly identical, with the exception of MANOVA requiring normality for all dependent variables, while IANOVA does not. This is a consequence of IANOVA being an asymptotic procedure, but it's worth noting that MANOVA is robust to non-normality for large sample sizes. MANOVA defines a linear combination of the dependent variables and a generalization of the sum of squares, which serve essentially the same purpose as the w-distance used to derive the test statistic of IANOVA. One of the notable advantages of MANOVA is its flexibility in choice of a test statistic, but this comes with the notable disadvantage that many of these statistics lack a straightforward distribution under the null hypothesis. For IANOVA, our statistic develops from theoretical results and has a manageable null distribution, while remaining reasonably flexible via the ω parameter. While either procedure will likely provide functionally very similar results, the decision of which to use will depend on the nature of your data, with IANOVA having a potentially stronger theoretical justification.

The work done here is by no means exhaustive, and there are a number of ways the IANOVA procedure could be expanded and improved upon. Notably, IANOVA is not currently generalized to account for multiple explanatory variables in the way a multi-way ANOVA procedure would. Additionally, while we have always assumed equal variances across groups, it would be worthwhile to explore building explicit tests for this assumption. Other statistical testing analogues like non-parametric tests and contrasts could add important tools for interval data analysis. Finally, a more comprehensive comparison to MANOVA (e.g. power analysis) would provide valuable insight into important differences between the two procedures. Interval data analysis is still a relatively new field, and developing additional tools/methodologies strengthens the ability of practitioners to deal with modern data problems.

REFERENCES

[1] G. Matheron, *Random sets and Integral Geometry.* John Wiley & Sons, 1975.

[2] R. Körner, *About a Variance of Compact Convex Random Sets.* Preprint, TU Bergakademie, 1995.

[3] F. d. A. T. de Carvalho, E. d. A. Lima Neto, and C. P. Tenorio, "A new method to fit a linear regression model for interval-valued data," in *KI 2004: Advances in Artificial Intelligence* (S. Biundo, T. Frühwirth, and G. Palm, eds.), (Berlin, Heidelberg), pp. 295–306, Springer Berlin Heidelberg, 2004.

[4] E. de A. Lima Neto and F. de A.T. de Carvalho, "Centre and range method for fitting a linear regression model to symbolic interval data," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1500–1515, 2008.

[5] E. de A. Lima Neto and F. de A.T. de Carvalho, "Constrained linear regression models for symbolic interval-valued variables," *Computational Statistics & Data Analysis*, vol. 54, no. 2, pp. 333–347, 2010.

[6] L. Billard, *Dependencies and Variation Components of Symbolic Interval-Valued Data*, pp. 3–12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[7] M. ángeles Gil, G. González-Rodríguez, A. Colubi, and M. Montenegro, "Testing linear independence in linear models with interval-valued data," *Computational Statistics & Data Analysis*, vol. 51, no. 6, pp. 3002–3015, 2007.

[8] B. Bean, Y. Sun, and M. Maguire, "Interval-valued kriging for geostatistical mapping with imprecise inputs," *International Journal of Approximate Reasoning*, vol. 140, pp. 31–51, 2022.

[9] B. et al., "The 2020 national snow load study," *USU Digital Commons*, 2020.

[10] Y. Sun, "Asymptotic tests for interval-valued means," *Statistics & Probability Letters*, vol. 121, p. 70–77, 2017.

[11] M. Montenegro, A. Colubi, M. Casals, and M. A. Gil, "Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable," *Metrika*, vol. 59, pp. 31–49, 02 2004.

[12] M. á. Gil, M. Montenegro, G. González-Rodríguez, A. Colubi, and M. Rosa Casals, "Bootstrap approach to the multi-sample test of means with imprecise data," *Computational Statistics & Data Analysis*, vol. 51, no. 1, p. 148–162, 2006.

[13] A. Colubi, "Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data," *Fuzzy Sets and Systems*, vol. 160, no. 3, p. 344–356, 2009.

[14] G. Debreu, "Integration of correspondences," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability II*, pp. 351–372, 1967.

[15] I. Molchanov, *Theory of Random Sets*. Springer-Verlag, London, 2005.

[16] Z. Artstein and R. Vitale, "A strong law of large numbers for random compact sets," *Annals of Probability*, vol. 5, pp. 879–882, 1975.

[17] R. Aumann, "Integrals of set-valued functions," *J. Math. Anal. Appl.*, vol. 12, pp. 1–12, 1965.

[18] N. Lyashenko, "Limit theorem for sums of independent compact random subsets of euclidean space," *Journal of Soviet Mathematics*, vol. 20, pp. 2187–2196, 1982.

[19] N. Lyashenko, "Statistics of random compacts in euclidean space," *Journal of Soviet Mathematics*, vol. 21, pp. 76–92, 1983.

[20] H. Rå dström, "An embedding theorem for spaces of convex sets," *Proc. Amer. Math. Soc.*, vol. 3, pp. 165–169, 1952.

[21] H. Hörmander, "Sur la fonction d'appui des ensembles convexes dans un espace localement convexe," *Arkiv för Mat*, vol. 3, pp. 181–186, 1954.

[22] M. Gil, M. Lopez, M. Lubiano, and M. Montenegro, "Regression and correlation analyses of a linear relation between random intervals," *Test*, vol. 10, no. 1, pp. 183–201, 2001.

[23] A. Joarder, M. Omar, and A. Gupta, "The distribution of a linear combination of two correlated chi-square variables," *Columbian Journal of Statistics*, vol. 36, pp. 211–221, 12 2013.

[24] D. E. Allen, "Snow loads on roofs. the present requirements and a proposal for a survey of snow loads on roofs," tech. rep., National Research Council, Division of Building Research, 1956.

[25] D. E. Allen, “Snow loads on roofs 1956-57: a progress report,” tech. rep., National Research Council, Division of Building Research, 1958.

[26] C. Allen and B. Peter, “Snow loads on roofs 1962-63: Seventh progress report,” tech. rep., National Research Council, Division of Building Research, 1963.

[27] Y. Faucher, *Snow Loads on Roofs 1964-65: Ninth Progress Report.* National Research Council of Canada, Division of Building Research, 1967.

[28] P. Hebert and B. Peter, “Snow loads on roofs 1961-62: Sixth progress report with an appendix on roof to ground load ratios,” tech. rep., National Research Council, Division of Building Research, 1963.

[29] M. Ho and D. A. Lutes, *Snow Loads on Roofs 1965-66: Tenth Progress Report.* National Research Council of Canada, Division of Building Research, 1968.

[30] I. Kennedy and D. Lutes, *Snow Loads on Roofs 1966-67: Eleventh Progress Report*, 12 1968.

[31] G. Pernica and B. Peter, *Snow Loads on Roofs 1963-64: Eighth Progress Report.* National Research Council Canada, Division of Building Research, 1966.

[32] J. Scott and B. Peter, “Snow loads on roofs 1960-61: Fifth progress report,” tech. rep., National Research Council, Division of Building Research, 1961.

[33] W. Watt and H. J. Thorburn, “Snow loads on roofs 1959-60: Fourth progress report,” tech. rep., National Research Council, Division of Building Research, 1960.

[34] CEC, “Ecological regions of North America: toward a common perspective,” tech. rep., Commission for Environmental Cooperation, 1997. Accessed: 12-1-2020.

APPENDICES

APPENDIX

Proof of Correlation

Theorem 4.0.1. *Under the null hypothesis, the chi-square random variables in our limiting distribution have correlation* ρ^2*, where* ρ *is the correlation between* $\bar{X}^C_{i\cdot} - \bar{X}^C_{\cdot\cdot}$ *and* $\bar{X}^R_{i\cdot} - \bar{X}^R_{\cdot\cdot}$*.*

Proof. Let

$$U = \bar{X}^C_{i\cdot} - \bar{X}^C_{\cdot\cdot}, \qquad V = \bar{X}^R_{i\cdot} - \bar{X}^R_{\cdot\cdot}.$$

Under the null hypothesis, U and V are asymptotically bivariate normal, with $\mu_U = \mu_V = 0$, variances σ^2_U, σ^2_V, and correlation ρ.
Let Z_1, Z_2 be independent standard normal random variables. We can define U and V as transformations of Z_1, Z_2 via the following

$$U = \sigma_U Z_1$$
$$V = \sigma_V(\rho Z_1 + \sqrt{1-\rho^2} Z_2)$$

The covariance of U^2 and V^2 can thus be represented by

$$\mathrm{Cov}(U^2, V^2) = \mathrm{Cov}(\sigma^2_U Z^2_1, \sigma^2_V(\rho Z_1 + \sqrt{1-\rho^2} Z_2)^2)$$

Properties of the covariance allow us to split this further into

$$\begin{aligned}
\mathrm{Cov}(U^2, V^2) &= \mathrm{Cov}(\sigma^2_U Z^2_1, \sigma^2_V(\rho Z_1 + \sqrt{1-\rho^2} Z_2)^2) \\
&= \sigma^2_U \sigma^2_V [\mathrm{Cov}(Z^2_1, (\rho Z_1 + \sqrt{1-\rho^2} Z_2)^2)] \\
&= \sigma^2_U \sigma^2_V [\mathrm{Cov}(Z^2_1, \rho^2 Z^2_1) + \mathrm{Cov}(Z^2_1, (1-\rho^2) Z^2_2) + \mathrm{Cov}(Z^2_1, \rho\sqrt{1-\rho} Z_1 Z_2)] \\
&= \sigma^2_U \sigma^2_V [2\rho^2 + 0 + 0]
\end{aligned}$$

Where we've used the independence of Z_1 and Z_2 and the moments of the standard normal distribution to simplify. To get the correlation, we use

$$\begin{aligned}
\mathrm{Corr}(U^2, V^2) &= \frac{\mathrm{Cov}(U^2, V^2)}{\sqrt{\mathrm{Var}(U^2)\mathrm{Var}(V^2)}} \\
&= \frac{2\sigma^2_U \sigma^2_V \rho^2}{\sqrt{2\sigma^4_U 2\sigma^4_V}} \\
&= \frac{2\sigma^2_U \sigma^2_V \rho^2}{2\sigma^2_U \sigma^2_V} \\
&= \rho^2
\end{aligned}$$

□