Final Project

Brooks Piper (brookspiper)

2024-12-11

Introduction

Analysis of Variance (ANOVA) is a cornerstone statistical method used to test differences among group means, playing a crucial role in fields ranging from psychology to engineering. However, the validity of its results depends on meeting key assumptions, including normality, homogeneity of variance, and sufficient sample sizes. Despite its commonhood, questions remain about how deviations from these assumptions—such as unequal variances, skewed means, or small sample sizes—affect the robustness and reliability of ANOVA. This study systematically investigates how manipulating aspects of sample composition, such as size, variance, and mean, influences the results of ANOVA. By doing so, we aim to provide deeper insights into the method's strengths and limitations, offering practical guidance for researchers designing experiments and interpreting their findings. The formal question driving this research study was: Are the presence of ANOVA assumptions necessary for its efficacy?

It was hypothesized that:

1. Disparities in variances result in the largest errors.
2. Violation to the normality assumption poses minimal concern.
3. In general, ANOVA remains fairly robust regardless of assumption validity.

Methods

ANOVA

ANOVA is a parametric statistical test used to analyze the equivalence of group means by partitioning variability into two components: systemic factors and random factors (Quirk, 2012). In sample data, variability can arise either from genuine differences between groups or from random chance, quantified as MST (mean square for treatment) and MSE (mean square for error), respectively. ANOVA evaluates the likelihood of these sources of variability against one another.

Within the calculations of metrics, each data point is compared to its group mean (attributing variability to group differences) or the overall mean of the dataset (attributing variability to error). The deviations are squared and then averaged to compute MST and MSE. These values are then used to calculate the test statistic, F, defined as

$$F = \frac{MST}{MSE}$$

An F-statistic follows an F-distribution, which arises from the unique distributions of the MST and MSE, both of which are chi-square distributions (Awan, 2001). When the ratio of these two chi-squares is calculated, the resulting test statistic follows an F-distribution. By assessing the probability of observing a specific F-ratio under the null hypothesis, ANOVA determines whether the variability between group means (MST) is sufficiently large relative to the overall error (MSE), indicating a significant difference between the groups. The null and alternative hypotheses for this test are as follows:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_A : \text{at least one of the means is different}$$

While this sounds complicated, in the era of statistical software, performing an ANOVA only requires one simple line of code without any of the arduous mathematical calculation. However, you can't just *do* an ANOVA—or at least you shouldn't (Emerson, 2022). Due to the nature of the test and the calculations involved, several assumption are made about the quality of the data, found below.

Table 1: ANOVA Assumptions

Normality	The data within each group should be normally distributed.
Homogeneity of Variance	The variance of the data within each group should be equal.
Independence	The observations within each group should be independent.
Random Sampling	Observations within each group have been sampled randomly and are independent of each other.

As convenient as it would be for these assumption to always be satisfied, statisticians are confronted with the reality that most of the time this is not the case. However, there is a test that performs the same tasks as ANOVA, but without all of the tedious assumptions: the Permutation test.

Permutation Test

Unlike ANOVA, the permutation test is a non-parametric method, meaning it does not rely on specific distributional assumptions. It leverages the null hypothesis of ANOVA: there is no difference in the true means of the groups. The implication of this hypothesis is that even if the data were completely randomized—rendered unrecognizable to human interpretation—the group means would remain equivalent. This randomization approach enables the permutation test to eliminate the influence of randomness inherent to ANOVA.

As the name suggests, the permutation test operates by systematically reordering, or permuting, the data. Theoretically, the test could analyze every possible permutation, but the number of permutations increases factorially with sample size, quickly becoming a computational nightmare. Instead, the test is typically limited to a manageable subset, such as 10,000 permutations. Despite this limitation, the permutation test achieves greater accuracy than ANOVA at the cost of runtime.

The calculation of the test statistic is straightforward. Data is randomly permuted to generate a theoretical sample, and ANOVA is performed on each permutation to compute the F-statistic. This process is repeated for all permutations, and the F-statistics are stored. The proportion of F-values from the permutations (F_{perm}) that are as extreme or more extreme than the original observed F-statistic (F_{obs}) provides the p-value, calculated as:

$$P = \frac{\text{number of permutations such that } F_{perm} \geq F_{obs}}{\text{total number of permutations}}$$

A key strength of the permutation test is its independence from assumptions about the data's underlying distribution, such as normality. The reliance on numerous permutations naturally eliminates this concern, making the test robust to assumption violations. For this reason, the permutation test serves as an ideal baseline for evaluating the performance of ANOVA, particularly in scenarios where its assumptions may be violated.

Data Generation

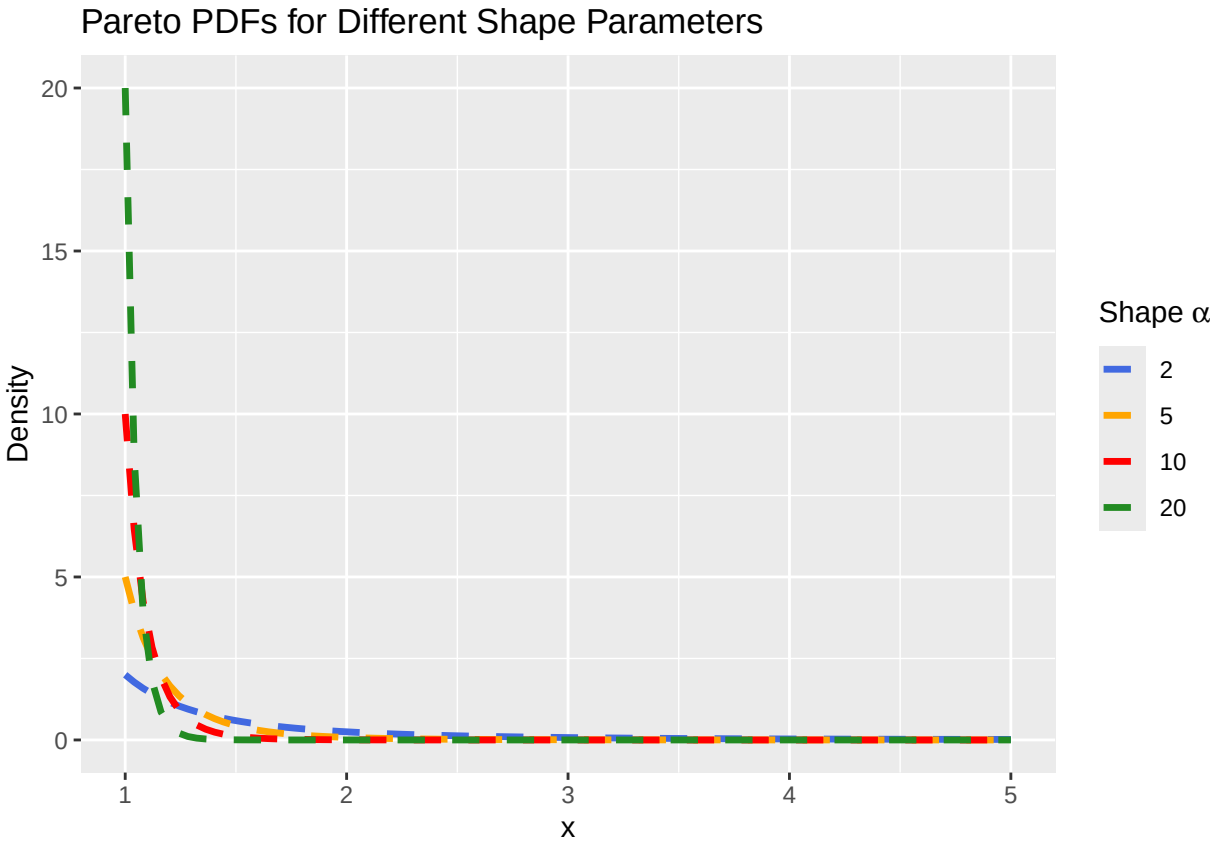
The Pareto Distribution

In the real world, statistical assumptions, particularly normality, are often violated (Sheng, 2008). To address this, the simulation study will generate data from the Pareto distribution. This distribution is widely known for its connection to the Pareto principle, which suggests that 80% of outcomes arise from 20% of inputs. While initially applied in wealth management to describe income distributions, the Pareto distribution has since proven versatile across various fields, including social science, quality control, geophysics, and risk analysis. Although not as commonly recognized as other distributions, the Pareto distribution is highly applicable and serves as a representative model for numerous real-world phenomena.

The Pareto distribution uses two parameters: the scale parameter α and the shape parameter θ. The PDF of the distribution is as follows:

$$f(x; \alpha, \theta) = \frac{\theta \alpha^{\theta}}{x^{\theta+1}}$$

The support of the Pareto distribution requires $\alpha, \theta > 0$. The distribution has mean $\mathbb{E}[X] = \frac{\theta\alpha}{\theta-1}$ and variance $Var(X) = \frac{\theta\alpha^2}{(\theta-1)^2(\theta-2)}$, however, the mean and variance are only finite when $\theta > 1$ and $\theta > 2$, respectively. For the sake of consistency in this simulation, all values of θ will be restricted to be greater than 2.



Note: In mathematical terms, α dictates the steepness of the distribution around θ (which is fixed at 1 here). Intuitively, the larger α is, the more extreme the Pareto principle is represented (as seen in comparing the distributions with $\alpha = 2$ and $\alpha = 20$).

The graphical representation reflects the non-normal shape of the Pareto distribution, which is far more closely characterized by the exponential distribution.

Procudure

For the actual data collection, samples will be drawn from the Pareto distribution using the **VGAM** package which contains the rpareto function.

VGAM::rpareto(n, scale, shape)

where n designates the sample size, and the scale and shape parameters determine the shape of the Pareto distribution.

To investigate the hypotheses, sample distributions will be partitioned into cases based on three criteria: means, variances, and sample sizes. Furthermore, these will be varied based on equivalence. The levels are also broken down in the following table:

Table 2: Mean-Variance-Sample Size Equivalence Chart

	Mean Equivalence	Variance Equivalence	Sample Size Equivalence
1	Equal	Equal	Equal
2	Equal	Equal	Mixed
3	Equal	Equal	Unequal

	Mean Equivalence	Variance Equivalence	Sample Size Equivalence
4	Equal	Mixed	Equal
5	Equal	Mixed	Mixed
6	Equal	Mixed	Unequal
7	Equal	Unequal	Equal
8	Equal	Unequal	Mixed
9	Equal	Unequal	Unequal
10	Mixed	Equal	Equal
11	Mixed	Equal	Mixed
12	Mixed	Equal	Unequal
13	Mixed	Mixed	Equal
14	Mixed	Mixed	Mixed
15	Mixed	Mixed	Unequal
16	Mixed	Unequal	Equal
17	Mixed	Unequal	Mixed
18	Mixed	Unequal	Unequal
19	Unequal	Equal	Equal
20	Unequal	Equal	Mixed
21	Unequal	Equal	Unequal
22	Unequal	Mixed	Equal
23	Unequal	Mixed	Mixed
24	Unequal	Mixed	Unequal
25	Unequal	Unequal	Equal
26	Unequal	Unequal	Mixed
27	Unequal	Unequal	Unequal

As previously mentioned, for each of the 27 cases, 10000 permutations will be performed. This function is coded below.

```
aov_perm_test <- function(df, reps){
  colnames(df) <- c("values","factors")

  perm_F <- NA

  for(i in 1:reps){
    df_perm <- df
    df_perm$values <- sample(df_perm$values)
    perm_F[i] <- summary(aov(values ~ factors,
                         data=df_perm))[[1]][1,4]
  }
  return(perm_F)
}
```

Mean Equivalence

To compare the results of the tests, Type I error, Type II error, and statistical power will be calculated. The strength of this simulation study lies in its ability to manually control how each sample is generated, allowing us to definitively know whether true differences exist among the group means. Consequently, calculating these error rates is straightforward, as they can be derived from the proportion of tests that correctly or incorrectly reach a conclusion. Moreover, the study incorporates variations in group means to examine how effectively each test detects differences under different conditions, and as such acts as the first factor of investigation.

Recall that the mean of a Pareto distribution is $\mathbb{E}[X] = \frac{\theta\alpha}{\theta-1}$. To obtain samples of equal, mixed, and unequal means, we can select various scale-shape pairs.

More specifically, we can solve for one of the parameters, say α, and then plug in different values of $\mathbb{E}[X]$ and θ to get Pareto distributions with the desired mean. This gives,

$$\alpha = \frac{\mathbb{E}[X](\theta - 1)}{\theta}$$

```
fix_alpha_mean <- function(E, theta){
  alpha <- E*(theta-1)/theta
  return(alpha)
}
```

Variance Equivalence

For the second aspect of this analysis, we will be examining the effectiveness of the tests in the context of the equal variance assumption. Recall that the variance of a Pareto distribution is $Var(X) = \frac{\theta\alpha^2}{(\theta-1)^2(\theta-2)}$. Once more, we can select various scale-shape pairs to obtain samples of equal, mixed, and unequal variances.

More specifically, we can solve for one of the parameters, say α, and then plug in different values of $Var(X)$ and θ to get Pareto distributions with the desired variances. This gives,

$$\alpha = \sqrt{\frac{Var(X)(\theta - 1)^2(\theta - 2)}{\theta}}$$

```
fix_alpha_variance <- function(Var, theta){
  alpha <- sqrt(Var*(theta-1)^2*(theta-2)/theta)
  return(alpha)
}
```

However, recall that unlike the normal distribution, the mean and variance for the Pareto distribution are parameterized by the same variables: α and θ. Therefore, when selecting a variance, we will also need to verify the equivalence of means across samples. We can do that with the formula for the mean of a Pareto distribution.

$$\mathbb{E}[X] = \frac{\theta\alpha}{\theta - 1}$$

```
pareto_mean <- function(theta, alpha){
  mean <- (theta*alpha)/(theta-1)
  return(mean)
}
```

If we instead decide to fix the mean, we can also derive the variance to verify equivalence.

$$Var(X) = \frac{\theta\alpha^2}{(\theta - 1)^2(\theta - 2)}$$

```
pareto_variance <- function(theta, alpha){
  variance <- (theta*(alpha)^2)/((theta-1)^2*(theta-2))
  return(variance)
}
```

Sample size equivalence

For the third and final part of this analysis, the effect of sample size will be investigated. Unlike the previous two assumptions, this one requires no mathematical computation, but more so mathematical reasoning. A large theme within statistics is the central limit theorem (CLT), which suggests that as the sample size grows infinitely large, the sample mean approaches the population mean.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mu$$

This leads into the next aspect of the CLT, which says that for large sample sizes, sample means become normally distributed.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

As a general rule of thumb, statisticians typically consider $n = 30$ as the threshold where the sample size is sufficiently large to satisfy the normality assumption. Larger sample sizes naturally make this assumption more plausible. In the context of ANOVA, this serves as a safeguard for the normality assumption, reducing the impact of deviations from normality (Wilcox et al., 2013). Within this study, we anticipate that larger sample sizes will significantly mitigate errors arising from non-normally distributed data and sample means. To explore this effect, the sample sizes were divided into three groups: $n = 4, n = 30$, and $n = 100$. Adjusting the sample size was straightforward, requiring only a change to the n parameter in the rpareto function.

Simulation Function

The entire simulation procedure is contained within the following function pareto_simulation.

pareto_simulation(shape1, scale1, n1, shape2, scale2, n2, shape3, scale3, n3, H_0, alpha)

where shape_i, scale_i, and n_i denote the parameters of the ith Pareto sample, H_0 denotes the truthy value of the null hypothesis, and alpha determines what the p-value significance cutoff is for the ANOVA tests in determining mean difference.

```
pareto_simulation <- function(shape1, scale1, n1,
                              shape2, scale2, n2,
                              shape3, scale3, n3,
                              H_0, alpha = 0.05) {
  simulation_perm_data <- NA
  simulation_aov_data <- NA
  perm_reps <- 10000
  loop_reps <- 1000

  for (i in 1:loop_reps) {
    pareto_data1 <- rpareto(n1, scale=scale1, shape=shape1)
    pareto_data2 <- rpareto(n2, scale=scale2, shape=shape2)
    pareto_data3 <- rpareto(n3, scale=scale3, shape=shape3)

    pareto_data <- data.frame(
      y = c(pareto_data1, pareto_data2, pareto_data3),
      x = factor(c(rep("Group 1", n1), rep("Group 2", n2), rep("Group 3", n3)))
    )

    perm_test <- aov_perm_test(pareto_data, perm_reps)
    F0 <- summary(aov(y ~ x, data=pareto_data))[[1]][1, 4]

    simulation_perm_data[i] <- mean(abs(perm_test) >= abs(F0))
    simulation_aov_data[i] <- summary(aov(y ~ x, data=pareto_data))[[1]][1, 5]
  }

  if (H_0) {
    typeI_error_perm <- mean(simulation_perm_data < alpha)
    typeI_error_aov <- mean(simulation_aov_data < alpha)

    table_data <- data.frame(
      `Type I Error` = c(typeI_error_perm, typeI_error_aov),
      Power = NA,
      `Type II Error` = NA,
      row.names = c("Permutation Test", "ANOVA")
    )
    caption <- "Error Rates and Power for Permutation Test and ANOVA"
  } else {
    power_perm <- mean(simulation_perm_data < alpha)
    power_aov <- mean(simulation_aov_data < alpha)

    typeII_error_perm <- mean(simulation_perm_data >= alpha)
    typeII_error_aov <- mean(simulation_aov_data >= alpha)
```

```
    table_data <- data.frame(
      Power = c(power_perm, power_aov),
      `Type II Error` = c(typeII_error_perm, typeII_error_aov),
      `Type I Error` = NA,
      row.names = c("Permutation Test", "ANOVA")
    )
  }

  colnames(table_data) <- gsub("\\.", " ", colnames(table_data))

  return(table_data)
}
```

The function returns a data frame which contains the errors of each test for later comparison.

Technical Issues

While there were no technical issues, the simulation does have its limitations. The need for accuracy, which necessitated numerous iterations, coupled with the decision to use the permutation test as a comparative baseline, demanded substantial computational power. As a result, running the entire simulation took multiple days. This created challenges in maintaining accuracy within time constraints, as even with caching, any small adjustment required a complete re-run of the simulation.

Results

For the sake of conciseness and clarity, we will depict three of the 27 cases of the simulation and leave the rest behind the scenes. Computationally and procedurally, they will operate in complete equivalence.

Case Breakdown

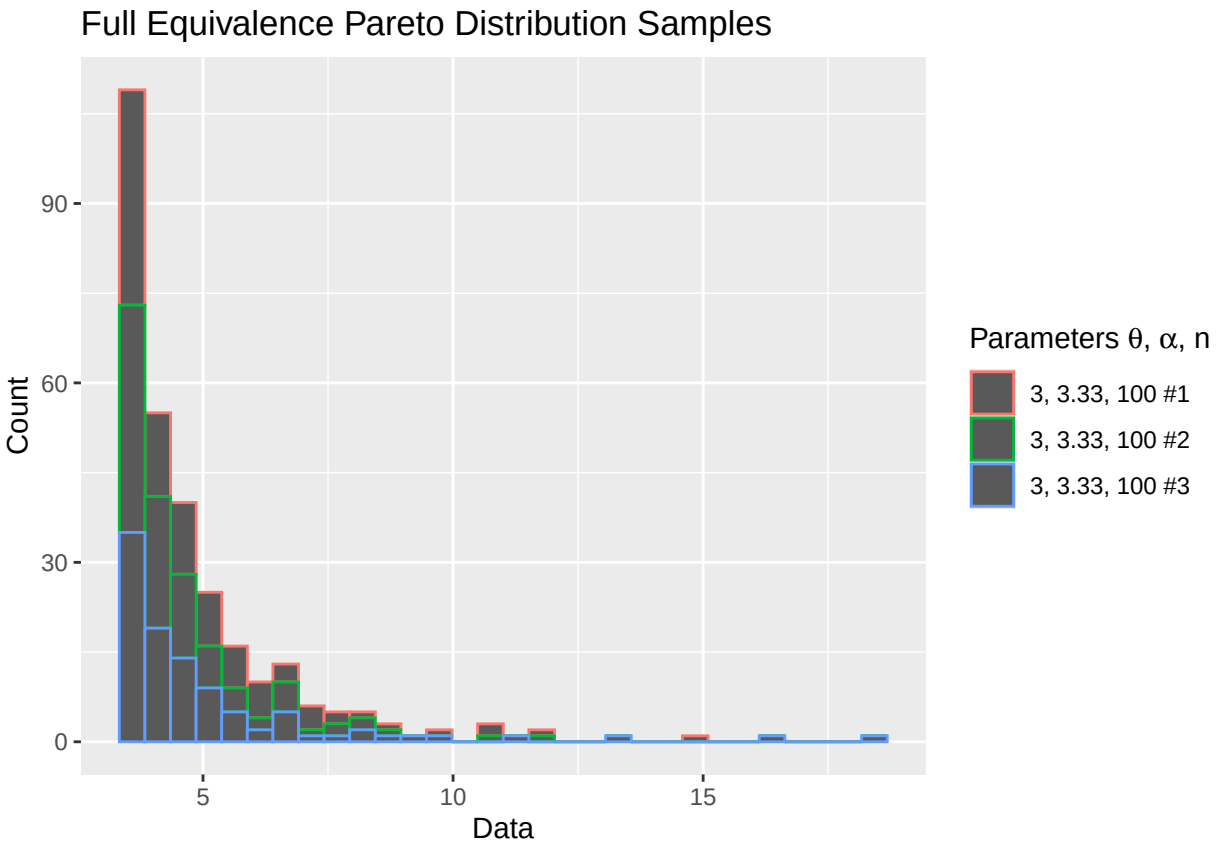
Case 1: Equivalent means, variances, and sample size

For the case of total equivalence, the resulting distributions are trivial (ie. require no functional computation to derive) due to the relationships of the mean and variance. That is, all of the distributions will be identical, so we can arbitrarily fix the mean to be 5.

Table 3: Full Equivalence Pareto Distribution Parameters

	n	θ	α	$\mathbb{E}[X]$	$\text{Var}(X)$
Distribution 1	100	3	3.333333	5	8.333333
Distribution 2	100	3	3.333333	5	8.333333
Distribution 3	100	3	3.333333	5	8.333333

A sample from these three equal Pareto distributions may look like the following:



With sample sizes of 100 each, the general shapes of the distributions appear very similar. Additionally, we can observe the qualities of equal means and variances within the samples, and as such, expect the tests to not reject.

Table 4: Error Rates and Power for Permutation Test and ANOVA

	Type I Error
Permutation Test	0.063
ANOVA	0.054

As expected, both the permutation and standard ANOVA tests reflect a low proportion of rejection-level p-values. That is, both tests come to the conclusion that there is no evidence to suggest any difference in the means. This was to be expected as all assumptions (aside from normality) were satisfied.

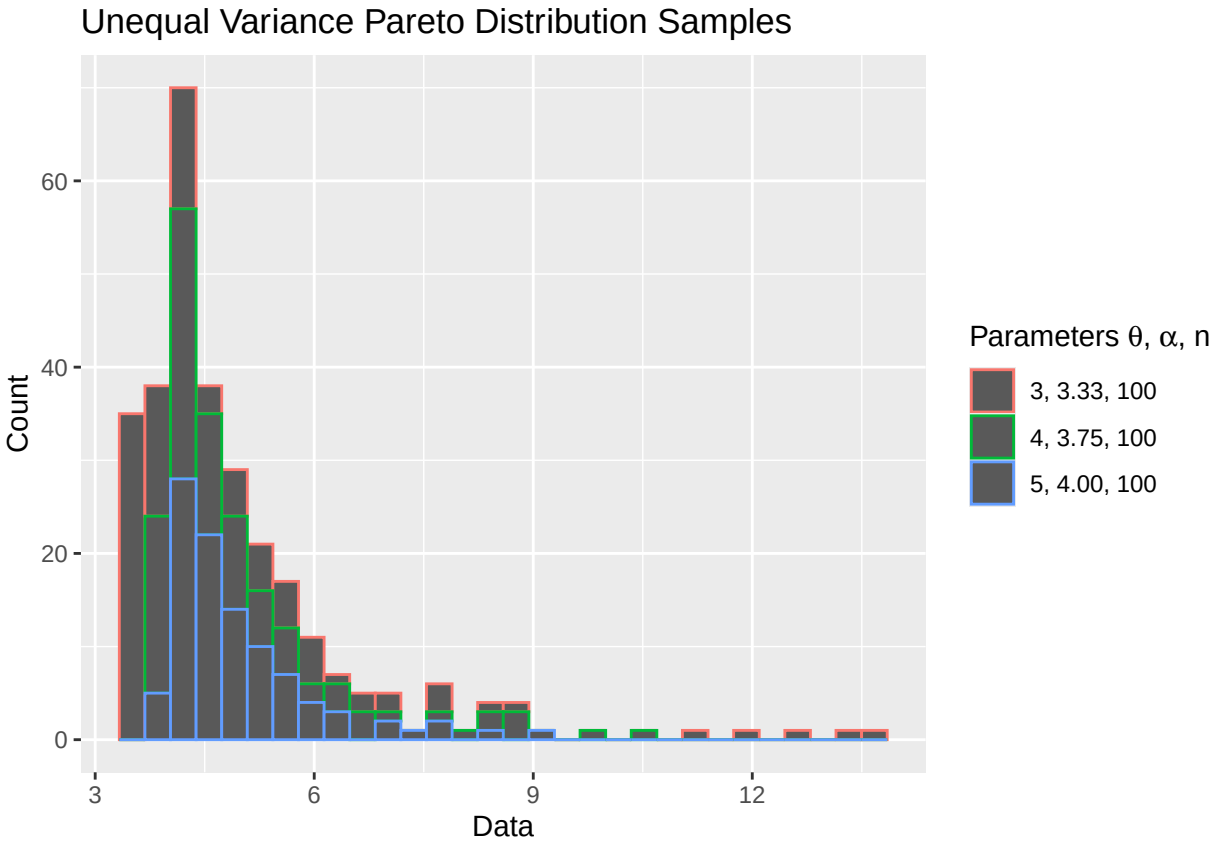
Case 7: Equivalent means and sample size but unequal variances

For the case of equal means and sample sizes but unequal variances, we can use the pareto_mean function while varying the θ argument across the three calls. This will result in three distributions with equal means and unequal variances. For this case, we will fix the mean to be 5 and let θ equal 3, 4, and 5.

Table 5: Unequal Variance Pareto Distribution Parameters

	n	θ	α	$\mathbb{E}[X]$	$\text{Var}(X)$
Distribution 1	100	3	3.333333	5	8.333333
Distribution 2	100	4	3.750000	5	3.125000
Distribution 3	100	5	4.000000	5	1.666667

A sample from these three Pareto distributions may look like the following:



With sample sizes of 100 each, the general shapes of the distributions begin to deviate away from one another. Additionally, we can observe the qualities of equal means and unequal variances within the samples, and once again, expect the tests to not reject.

Table 6: Error Rates and Power for Permutation Test and ANOVA

	Type I Error
Permutation Test	0.079
ANOVA	0.071

As expected, both the permutation and standard ANOVA tests reflect a low proportion of rejection-level p-values. That is, both tests come to the conclusion that there is no evidence to suggest any difference in the means. However, while the tests come to the correct conclusion, the Type I Error rate increased from the previous case. Despite the change being marginal, it reflects a possible vulnerability to unequal variances for ANOVA. Further exploration of additional test cases will provide a fuller understanding of this property.

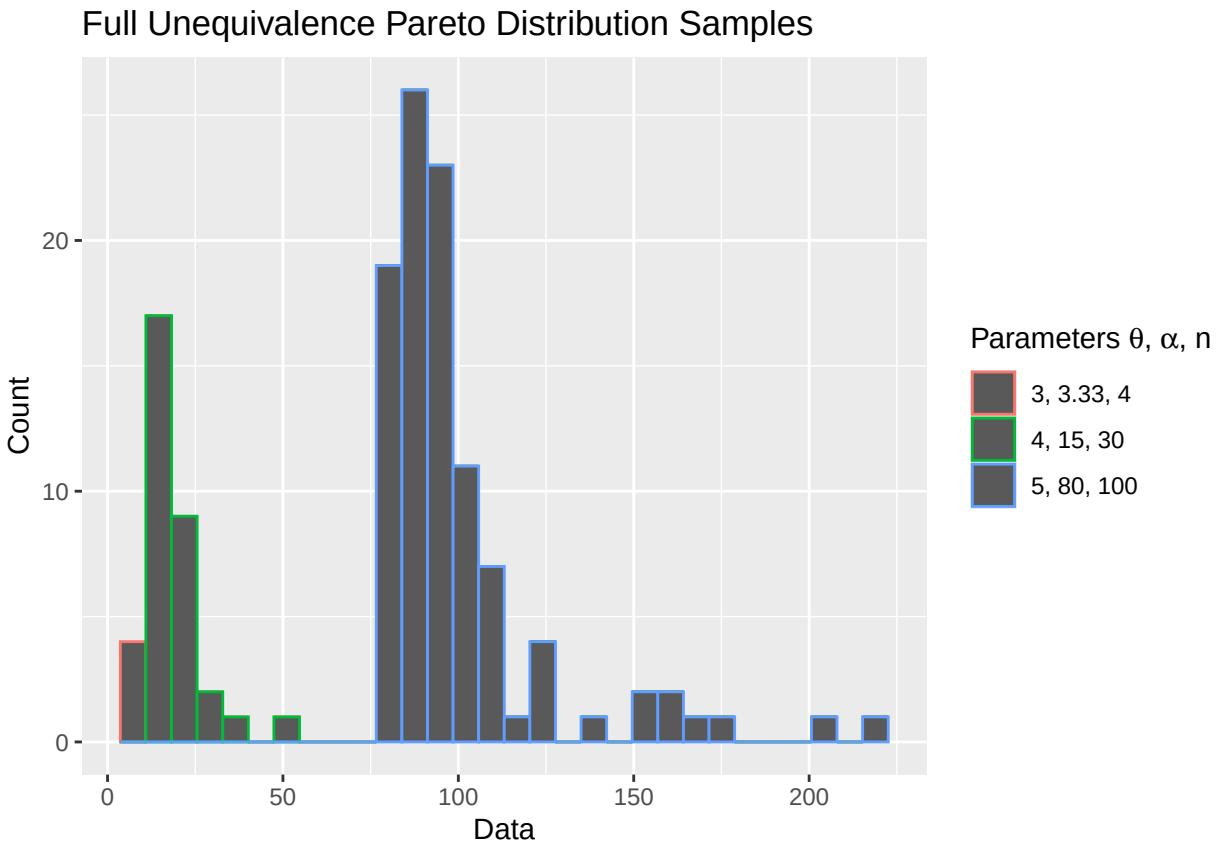
Case 27: Unequivalent means, variances, and sample size

For the case of total unequivalence, we can once again use the pareto_mean function while varying the θ argument across the three calls. This will result in three distributions with unequal means and unequal variances. For this case, we will fix the mean to be 5, 20, 100 and let θ equal 3, 4, and 5.

Table 7: Full Unequivalence Pareto Distribution Parameters

	n	θ	α	$\mathbb{E}[X]$	$\text{Var}(X)$
Distribution 1	4	3	3.333333	5	8.333333
Distribution 2	30	4	15.000000	20	50.000000
Distribution 3	100	5	80.000000	100	666.666667

A sample from these three unequal Pareto distributions may look like the following:



With the sparse difference in sample sizes, the general shapes of the distributions have noticeable differences. Additionally, we can observe the qualities of unequal means and variances within the samples, and as such, expect the tests to reject.

Table 8: Error Rates and Power for Permutation Test and ANOVA

	Power	Type II Error
Permutation Test	1	0
ANOVA	1	0

As expected, both the permutation and standard ANOVA tests yield rejection-level p-values, indicating evidence to suggest a difference in the means. These results highlight ANOVA's ability to handle unequal variances and introduce its tolerance for unequal sample sizes when the test is expected to reject H_0. Analysis of all 27 cases will provide a more comprehensive understanding of these properties, especially in how they differ between cases of true equivalence and true difference in group means.

Simulation

Cases where the means were equal (ie., H_0 was true) will be grouped into one table, while cases with mixed or unequal means (ie., H_0 was false) will be separated into two additional tables.

Table 9: H_0 True—Equal Means

Case	Test	Type I Error
1	Permutation	0.061
1	ANOVA	0.051
2	Permutation	0.051
2	ANOVA	0.054
3	Permutation	0.051
3	ANOVA	0.056
4	Permutation	0.083
4	ANOVA	0.078
5	Permutation	0.125
5	ANOVA	0.132
6	Permutation	0.353
6	ANOVA	0.387
7	Permutation	0.100
7	ANOVA	0.093
8	Permutation	0.243
8	ANOVA	0.261
9	Permutation	0.437
9	ANOVA	0.473

Table 10: H_0 False—Mixed Means

Case	Test	Power	Type II Error
10	Permutation	1.000	0.000
10	ANOVA	1.000	0.000
11	Permutation	1.000	0.000
11	ANOVA	1.000	0.000
12	Permutation	0.997	0.003
12	ANOVA	0.997	0.003
13	Permutation	1.000	0.000
13	ANOVA	1.000	0.000
14	Permutation	1.000	0.000
14	ANOVA	1.000	0.000
15	Permutation	1.000	0.000
15	ANOVA	1.000	0.000
16	Permutation	1.000	0.000
16	ANOVA	1.000	0.000
17	Permutation	0.998	0.002
17	ANOVA	0.997	0.003
18	Permutation	0.998	0.002
18	ANOVA	0.998	0.002

Table 11: H_0 False—Unequal Means

Case	Test	Power	Type II Error
19	Permutation	1	0
19	ANOVA	1	0
20	Permutation	1	0
20	ANOVA	1	0
21	Permutation	1	0
21	ANOVA	1	0
22	Permutation	1	0
22	ANOVA	1	0
23	Permutation	1	0
23	ANOVA	1	0
24	Permutation	1	0
24	ANOVA	1	0
25	Permutation	1	0
25	ANOVA	1	0
26	Permutation	1	0
26	ANOVA	1	0
27	Permutation	1	0
27	ANOVA	1	0

- Code used to create these tables written by Deanmoran, 2023; See full citation in **References**

Hypothesis 1: The Effect of Variance

All tables illustrate how violations of the variance assumption impact ANOVA's accuracy. In Table 9, error rates increase progressively as variance assumptions are ignored, with the largest violations toward the tail end driving errors up by 10% to 40%.

Table 12: Summary of Analysis for Mixed and Unequivalent Variance

Case	Test	Type I Error
4	Permutation	0.083
4	ANOVA	0.078
5	Permutation	0.125
5	ANOVA	0.132
6	Permutation	0.353
6	ANOVA	0.387
7	Permutation	0.100
7	ANOVA	0.093
8	Permutation	0.243
8	ANOVA	0.261
9	Permutation	0.437
9	ANOVA	0.473

When ANOVA underperformed compared to the permutation test, two consistent factors emerged: unequal variances or unbalanced sample sizes. Although sample size imbalance contributed to errors, prior observations (seen in cases 4 through 9) highlight unequal variances as having the most significant impact on error rates.

Conversely, cases with equal variances—such as 1, 2, 3, 10, 11, 12, 19, 20, and 21—consistently exhibited significantly lower error rates. Even when sample sizes varied, the resulting error increase was marginal compared to the sharp rise observed under unequal variances.

Table 13: Summary of Analysis for Equal Variance

Case	Test	Type I Error	Power	Type II Error
1	Permutation	0.061	NA	NA
1	ANOVA	0.051	NA	NA
2	Permutation	0.051	NA	NA
2	ANOVA	0.054	NA	NA
3	Permutation	0.051	NA	NA
3	ANOVA	0.056	NA	NA
10	Permutation	NA	1.000	0.000
10	ANOVA	NA	1.000	0.000
11	Permutation	NA	1.000	0.000
11	ANOVA	NA	1.000	0.000
12	Permutation	NA	0.997	0.003
12	ANOVA	NA	0.997	0.003
19	Permutation	NA	1.000	0.000
19	ANOVA	NA	1.000	0.000
20	Permutation	NA	1.000	0.000
20	ANOVA	NA	1.000	0.000
21	Permutation	NA	1.000	0.000

Case	Test	Type I Error	Power	Type II Error
21	ANOVA	NA	1.000	0.000

Hypothesis 2: The Unimportance of Normality

All tables highlight how violations of normality influence ANOVA results. When the normality assumption was formally ignored—aside from considerations of sample size—the results collectively suggest that its impact on accuracy is minimal. In all but three cases, the error rate remained at or below roughly 13%, indicating that deviations from normality do not strongly affect ANOVA's performance.

Table 14: Summary of Analysis for non-Normality

Case	Test	Type I Error
1	Permutation	0.061
1	ANOVA	0.051
2	Permutation	0.051
2	ANOVA	0.054
3	Permutation	0.051
3	ANOVA	0.056
4	Permutation	0.083
4	ANOVA	0.078
5	Permutation	0.125
5	ANOVA	0.132
7	Permutation	0.100
7	ANOVA	0.093

Examining sample size as a mitigating factor reinforces this conclusion. In cases where the sample size was equal at 100 observations—specifically cases 1, 4, 7, 11, 14, 17, 21, 24, and 27—ANOVA outperformed the permutation test, though only by a margin of up to 10%. Compared to the variance assumption violations observed earlier, these effects are minimal.

Table 15: Summary of Analysis for High and Equal Sample Size

Case	Test	Type I Error	Power	Type II Error
1	Permutation	0.061	NA	NA
1	ANOVA	0.051	NA	NA
4	Permutation	0.083	NA	NA
4	ANOVA	0.078	NA	NA
7	Permutation	0.100	NA	NA
7	ANOVA	0.093	NA	NA
11	Permutation	NA	1.000	0.000
11	ANOVA	NA	1.000	0.000
14	Permutation	NA	1.000	0.000
14	ANOVA	NA	1.000	0.000
17	Permutation	NA	0.998	0.002
17	ANOVA	NA	0.997	0.003
21	Permutation	NA	1.000	0.000
21	ANOVA	NA	1.000	0.000
24	Permutation	NA	1.000	0.000
24	ANOVA	NA	1.000	0.000
27	Permutation	NA	1.000	0.000

Case	Test	Type I Error	Power	Type II Error
27	ANOVA	NA	1.000	0.000

Hypothesis 3: The Benefit of Assumptions

All tables demonstrate ANOVA's robustness across various scenarios where assumptions are violated. Consistent with previous observations, in most cases, the error rate remained below approximately 13%, suggesting that, in general, assumption violations have minimal impact on accuracy.

However, cases with notably higher error rates reveal a different narrative. Cases 6, 8, and 9—characterized by mixed variances and sample sizes, unequal variances and sample sizes, or a combination of these factors—highlight ANOVA's limitations. For example, in case 9 from Table 9, where both variances and sample sizes were unequal, ANOVA produced a Type I error rate nearing 50%. This extreme error rate, along with those found in the other two cases, reflects the cumulative effect of violating multiple assumptions simultaneously, demonstrating a threshold where ANOVA results become unreliable.

Table 16: Summary for Analysis with High Error

Case	Test	Type I Error
6	Permutation	0.353
6	ANOVA	0.387
8	Permutation	0.243
8	ANOVA	0.261
9	Permutation	0.437
9	ANOVA	0.473

In contrast, Tables 10 and 11, which examined cases where ANOVA was expected to reject the null hypothesis, showed near-perfect accuracy. This finding marks a critical aspect of ANOVA's robustness: when population means are truly different, assumption violations have minimal effect. However, Table 9 emphasizes the opposite scenario—when population means are identical—where certain assumption violations can significantly undermine reliability.

Discussion

This simulation study examined how violations of key assumptions affect ANOVA's accuracy, hypothesizing that homogeneity of variance would be the most critical factor influencing error rates. By sampling from various Pareto distributions with varying shape and scale parameters, we analyzed how differences in means, variances, and sample sizes contributed to accuracy reduction.

First, the results confirmed that homogeneity of variance was the most influential assumption, with accuracy declining as this assumption was progressively violated, supporting our first hypothesis. Our second hypothesis, which suggested that ANOVA would maintain accuracy even when the normality assumption was violated, was also supported. This finding was particularly evident in cases with large sample sizes, where the central limit theorem mitigated the impact of non-normality. Finally, our third hypothesis, proposing that ANOVA's accuracy would remain robust despite major assumption violations, was mostly upheld. However, in cases with significant violations—such as lack of normality, heterogeneity of variance, and small, unequal sample sizes—ANOVA's vulnerability became apparent.

These findings demonstrate that while ANOVA often performs reliably even under imperfect conditions, its accuracy is contingent on the degree of assumption violations. In particular, homogeneity of variance proved to be the most critical factor, highlighting the importance of verifying this assumption in practice. This is especially relevant in applied fields where small or unequal sample sizes and heterogeneous data are common. By carefully validating assumptions, researchers can minimize errors and improve the reliability of their analyses.

This study also emphasizes the need for practical tools and guidelines to diagnose and address assumption violations effectively. Procedural and computational constraints limited our ability to examine all assumptions comprehensively, such as data independence and random sampling. Future research should explore these assumptions, either independently or alongside normality and homogeneity of variance, to develop a more holistic understanding of ANOVA's robustness. Moreover, broader investigations could assess whether the robustness observed here is unique to ANOVA or a general characteristic of statistical tests.

In conclusion, this study validates the robustness of ANOVA in many scenarios but cautions against disregarding its assumptions entirely. Researchers should prioritize assumption testing and mitigation strategies to optimize the accuracy and reliability of their results, particularly when working with complex or imbalanced datasets. By doing so, they can ensure that statistical conclusions remain both valid and impactful.

References

Awan, Hayat M. "Effect of Departures From Standard Assumption Used In Analysis of Variance." *Journal of Research (Science) 12.2*, 2001, pp. 180-188.

Deanmoran, Julian. "Align multiple tables side by side." Stack Overflow, 2023, https://stackoverflow.com/questions/38036680/align-multiple-tables-side-by-side.

Emerson, Robert Wall. "ANOVA Assumptions." *Journal of Visual Impairment & Blindness*, vol. 116, no. 4, 2022, pp. 585–586.

Sheng, Yanyan. "Testing the Assumptions of Analysis of Variance." *Best Practices in Quantitative Methods*, 2008, pp. 324–340.

Wilcox, Rand, Mike Carlson, Stan Azen, and Florence Clark. "Avoid Lost Discoveries, Because of Violations of Standard Assumptions, by Using Modern Robust Statistical Methods." *Journal of Clinical Epidemiology*, vol. 66, no. 3, 2013, pp. 319–329. ScienceDirect, https://doi.org/10.1016/j.jclinepi.2012.09.003.

Quirk, Thomas J. "One-Way Analysis of Variance (ANOVA)." *Excel 2007 for Educational and Psychological Statistics: A Guide to Solving Practical Problems*, 2012, pp. 163–179.