ALGORITHM ON HYPOTHESIS TESTING ON THE MEANS OF TWO NORMAL POPULATION AND ITS' IMPLEMENTATION ON COMPUTER USING R

BY

BETHEL EJEDAWE

PSC1909241

DEPARTMENT OF STATISTICS

FACULTY OF PHYSICAL SCIENCES

UNIVERSITY OF BENIN

BENIN CITY NIGERIA

APRIL, 2024

ALGORITHM ON HYPOTHESIS TESTING ON THE MEANS OF TWO NORMAL POPULATIONS AND ITS' IMPLEMENTATION ON COMPUTER USING R

BY

BETHEL EJEDAWE

PSC1909241

A PROJECT WRITTEN AND SUBMITTED TO THE DEPARTMENT OF STATISTICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE (B.SC.) IN STATISTICS OF THE UNIVERSITY OF BENIN, BENIN CITY, NIGERIA

APRIL, 2024

CERTIFICATION

We certify that this work was carried out by Bethel Ejedawe of the Department of Statistics, University of Benin, Nigeria

Prof. J.I. Mbegbu **Date**
Supervisor

Prof. N. Ekhosuehi **Date**
Head of Department

External Examiner **Date**

UNDERTAKING

This project work was carried out by Bethel Ejedawe with Matriculation Number PSC1909241. I have neither copied nor duplicated the work of any other author(s). All works used have been duly cited and acknowledged.

Bethel Ejedawe
Name of Student **Signature/Date**

DEDICATION

This project is dedicated to God and my parents Mr. and Mrs. Ejedawe for the moral and financial support given to me.

ACKNOWLEGEMENT

With deep appreciation, I would like to extend my gratitude to my project supervisor, Prof. J.I. Mbegbu for his invaluable guidance. Special thanks to my course adviser Mr. C.O. Odijie who has taken his time to ensure that each student is on the right track in regards of his/her academics. I'm also thankful for the teachings rendered to us by our lecturers. A heartfelt thanks to each contributor for their role in bringing my project to life.

To my lovely parents, I want to thank you for the unconditional love you showed towards me. I want to also thank my friends Osaretin Favour, Bamigboye David, Uruvwewhu Courage, Racheal, Precious, Delight, Faith, Chiedoziem, Loretta, Beatrice, Cecilia, Blessing, Sophia, Esther, Timothy, Victory, Miracle, Joshua Oluwaseyi, Joshua(Luther), Chebe, Japhet, my departmental Precident Tracy and entire coursemates that have made my stay in this institution a good one.

Special thanks my fellow project collegues/Friends, Anosike Beatrice, Odafehere Loveth, Osaze-ilori Osayande David, Akhigbe Eucharia, and above all, I thank God Almighty.

ABSTRACT

This study evaluated and compared the performance of three statistical methods for hypothesis testing when comparing means between two populations: the t-test, Welch's t-test, and the z-test. The t-test assumes normally distributed data and equal variances, while Welch's t-test accounts for unequal variances, and the nonparametric Mann-Whitney U test is an alternative for non-normal data. The research aimed to determine the optimal test by formulating hypotheses, selecting appropriate test statistics, determining sample sizes, and implementing the tests using R programming. The data analyzed were the mean heights of NBA guards and forwards during the 2022-2023 season. A power analysis assessed the reliability, validity, and assumptions of the tests. The results indicated a significant difference in mean heights between guards and forwards, with guards being slightly taller on average. Importantly, the Welch's t-test consistently outperformed the standard t-test and z-test across varying sample sizes, demonstrating higher power and a greater ability to detect true effects while minimizing Type I and Type II errors. This superior performance is attributed to the robustness of Welch's t-test in handling unequal variances between groups, a common scenario in real -world data analysis.

TABLE OF CONETENTS

CHAPTER THREE: METHODOLOGY

CHAPTER FOUR: DATA ANALYSIS AND RESULT

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATION

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

Did you know that the weight of human beings varies depending on their type of diet, and that this variation can be modeled by the normal distribution? According to a study conducted by Dawczynski et al. (2022), flexitarians, vegetarians, and vegans had a lower body weight, BMI, and fat percentage in comparison to omnivores ($p \le 0.05$).

In the realm of statistical analysis, hypothesis testing serves as a foundation for drawing meaningful insights from data. This study specifically delves into the comparison of means for two normally distributed populations, a fundamental aspect within hypothesis testing, drawing inspiration from seminal research works by Moore et al (2009)

Hypothesis testing serves as the analytical compass guiding researchers through the maze of uncertainties, enabling them to validate assumptions and draw evidence-based conclusions. Within this framework, the act of comparing means takes centre stage, offering insights into the differences or similarities between two groups.

The normal distribution, a statistical archetype, plays a key role in this study. The symmetrical bell –shaped curve serves as the canvas upon which the comparison unfolds, highlighting the elegance and precision inherent in statistical analysis.

However, this study is not merely a statistical journey, the study ventures into the practical domain of algorithm development. A bespoke algorithm, fine-tuned for comparing means of two normal distributions, takes shape, showcasing the fusion of theoretical principles with computational pragmatism.

As our narrative progresses, the study seamlessly weaves into the fabric of computer science. The relevance of mean comparison in computational scenarios is explored, drawing connections between methodologies and computer science applications.

However, as with any study, certain limitations frame its boundaries. The assumptions of normal distribution, algorithmic specificity, and the simplification of real-world scenarios are acknowledged .These limitations serve as navigational markers, guiding readers through the study's scope while highlighting the need for a nuanced interpretation of its findings.

1.2 AIM AND OBJECTIVES

The primary aim of this study is to explore and implement algorithms for hypothesis testing of the means of two normal populations and assess their performance through practical simulations on a computer, and the objectives are to;

1. Implement and evaluate algorithms for hypothesis testing of the means in two normal population using R.

2. Conduct comparative analysis in assessing which statistical test provides the most reliable and valid results for our analysis, given the data characteristics and assumptions inherent in each test.

3. Identify and document of challenges encountered during the implementation and assessment of algorithms.

1.3 SCOPE

This study aims to implement and evaluate three distinct hypothesis testing algorithms for comparing means of two normal populations. The selected algorithms include the z-test, independent sample t-test, and the Welch's t-test. The implementation would be carried out using R (a programming language), providing practical insights to the application of these algorithms. Simulation studies will be conducted to assess the effectiveness and efficiency of the implemented tests under various conditions, offering valuable comparison of mean differences.

1.4 STATEMENT OF THE PROBLEMS

- Ensuring the normality assumption and homogeneity of variances was challenging for real-world scenarios.

- Determining the appropriate sample size to achieve sufficient statistical power was crucial, as both small and large sample sizes can introduce issues.
- Obtaining accurate and representative samples, as well as dealing with unknown population parameters, was problematic.
- Implementation of the various algorithm in a programming language was challenging as some of the test e.g Z-test did not have an in-built function.
- Interpreting the statistical significance and limitations, as well as generalizing the findings, required careful consideration.

1.5 DEFINITION OF TERMS

1.5.1. HYPOTHESIS (POPPER, 1959)

A hypothesis is a proposed explanation for a phenomenon or a statement about the relationship between two or more variables that can be tested through observation and experiment.

1.5.2. HYPOTHESIS TESTING (FISHER, 1925)

Hypothesis testing is a statistical method used to determine whether a hypothesis about a parameter of a population is likely to be true or false.

1.5.3. NORMAL DISTRIBUTION (GAUSS, 1809)

The normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

Mathematically:

A continuous random variable X is said to have a normal distribution, with mean μ and variance σ^2, that is, $X \sim N(\mu, \sigma^2)$, if its pdf $f_x(x)$ and the cdf $F_x(x) = P(X \leq x)$ are, respectively

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2}, \quad -\infty<x<\infty,$$

and

$$F_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2} dy$$

$$= \frac{1}{2}[1+\mathrm{erf}\,(\frac{x-\mu}{\sigma\sqrt{2}})], \quad -\infty<x<\infty, \quad -\infty< \mu <\infty, \quad \sigma>0,$$

Where erf(.) denotes error function, and μ and σ are location and scale parameters respectively

1.5.4. ALGORITHM (KNUTH, 1973)

An algorithm is a step-by-step procedure for solving a problem or accomplishing a task.

1.5.5. MEAN (PEARSON, 1896)

The mean is the sum of all the values in a dataset divided by the total number of values.

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

Where:

$\bar{y}$= mean

n = number of observations

$\sum_{i=1}^{n} y_i$ = sum of all the individuals in the data set

1.5.6. P-VALUE (FISHER, 1925)

The p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is true.

1.5.7. CONFIDENCE INTERVAL (NEYMAN, 1937)

A confidence interval is a range of values that is likely to contain an unknown population parameter, with a specified probability.

1.5.8. STANDARD DEVIATION (PEARSON, 1894)

The standard deviation is a measure of the dispersion or spread of a dataset, calculated as the square root of the variance.

1.5.9. VARIANCE PEARSON (1894).

Variance is a measure of the average squared deviation from the mean in a dataset.

1.5.10. Z-TEST (GOSSET, 1908)

The Z-test is a statistical test used to determine whether the mean of a population is significantly different from a hypothesized value, assuming the population follows a normal distribution.

One sample case

$$Z = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Where;

Z= z-stat

$\bar{y}$= sample mean of the dataset

σ = standard deviation of the dataset

μ_0 = hypothesized mean

n = sample size

Two sample case

$$Z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where;

Z= z-stat

$\bar{y}_1$= sample mean of 1st dataset

$\bar{y}_2$ = sample mean of 2nd-dataset

μ_1 =population mean of 1st dataset

μ_2 = population mean of 2nd dataset

σ^2_1= variance 1st dataset

σ^2_1= variance 2nd dataset

n_1= sample size 1st dataset

n_2 = sample size 2nd dataset

1.5.11. INDEPENDENT T-TEST (STUDENT, 1908)

The independent t-test is a statistical test used to determine whether the means of two independent samples are significantly different from each other.

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Where $\bar{y}_1$ and $\bar{y}_2$ are the means of the two groups, n_1 and n_2 are the sizes of the groups, and $s^2{}_1$ and $s^2{}_2$ are the variances of the two groups. The degrees of freedom for this test are n_1+n_2-2.

1.5.12. WELCH'S T-TEST (WELCH, 1947)

Welch's t-test is a variant of the independent t-test that is used when the two samples have unequal variances.

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)}}$$

Where $\bar{y}_1$ and $\bar{y}_2$ are the means of the two groups, n_1 and n_2 are the sizes of the groups, and $s^2{}_1$ and $s^2{}_2$ are the variances of the two groups.

The degrees of freedom for this test are estimated as:

$$df=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

1.5.13. POWER OF A TEST (WEISS, 2012)

The power of a statistical test is the probability of rejecting the null hypothesis when it is false.

1.5.14. COHEN'S D (COHEN, 1988)

Cohen's d is a measure of the standardized mean difference between two groups

a. Cohen's d for two independent samples with equal variance:

$$d = \frac{\bar{y}_1-\bar{y}_2}{\sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}}$$

Where:

$\bar{y}_1$= sample mean of group 1

$\bar{y}_2$= sample mean of group 2

n_1= sample size of group 1

n_2= sample size of group 2

s_1^2 and s_2^2= sample standard deviations of the two groups

b. Cohen's d for Z-test with two independent samples:

$$d = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where:

σ_1^2 and σ_2^2 = population standard deviations of the two groups

c. Cohen's d for Welch t-test with two independent samples:

$$d = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2 + s_1^2}{2}}}$$

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Hypothesis testing is a methodology to systematically quantify how certain one can be of the result of a statistical experiment (gabrieletolomei.wordpress.com). Hypothesis tests are central to quantitative research in the social sciences. The standard format for research papers is to draw on theory to make a case for a positive or negative association between the values of two variables, report a measure of association between those variables after controlling for other potentially relevant factors, and present a test of the “null hypothesis” that the association is zero (David et al, 2016).

In research, there are two types of hypotheses: null and alternative. They work as a complementary pair, each sating that the other is wrong. **The null hypothesis (H_0)** can be thought of as the implied hypothesis. “Null” meaning “nothing”. This hypothesis states that there is no difference between groups or no relationship between variables. The null is a presumption of the status quo or no change, while the **alternative hypothesis (H_a)** should state what you expect the data to show, based on your research on the topic (resources.nu.edu).

2.2 HISTORICAL CONTEXT AND EVOLUTION

The concept of hypothesis testing has undergone significant evolution since its inception. The Trial of the Pyx serves as an early example of hypothesis testing principles, its formalization and evolution into the sophisticated methods we use today unfolded over centuries.

Following the Trial of the Pyx, innovative thinkers started applying similar logic to diverse areas. In 1710, John Arbuthnot's statistical analysis of sex ratios paved the way for using data to explore phenomena beyond physical measurements. Later, Michell (1767) studied on star distribution marked an early foray into applying hypothesis testing to astronomical observations. These examples highlight the gradual shift from purely physical applications to broader spheres of inquiry.

However, the 20th century witnessed the most significant advancements in hypothesis testing methodology. Pearson's chi-square test (1900) offered a powerful tool for assessing categorical data, paving the way for rigorous analysis in medical research. Notably, William Gosset ("Student") made groundbreaking contributions to small sample testing in 1908, significantly impacting medical studies with limited participant groups.

Meanwhile, in ecology, hypothesis testing found fertile ground. Ronald Fisher's work, particularly his emphasis on p-values (1920), became instrumental in analyzing ecological data and drawing inferences about populations and environmental factors. Later, Neyman and Pearson (1930) refined the framework with their null hypothesis approach, further solidifying the methodology's foundation in ecological research. "*Historical Hypothesis Testing*"(www.usu.edu). However, the formulation and philosophy of hypothesis testing was largely created in the period 1915-1933 by three men Fisher (1890-1962), Neyman (1894-1981), and Pearson (1895-1980). Since then it has expanded into one of the most widely used quantitative methodologies, and has found its way into nearly all areas of human endeavor Lehmann (1993)

2.3 STATISTICAL METHODS FOR ANALYZING MEAN DIFFERENCES

The problem of comparison of two samples obtained in different measurements appears in a wide range of tasks starting from physical research and ending with social and political studies. The comparison includes the tests of the samples' distributions and their parameters, and the result of the comparison specifies whether the samples were drawn from the same population or not (Alexander and Eugene (2023)). At the heart of comparing means between two groups lies the t-test also known as "Student's t

-test”, a statistical method developed by William Sealy Gosset in 1908. The Student t-test is a parametric method that needs the observations/populations of study to be normally distributed as well as having equal variances. The Student t-test is a powerful test if the homogeneity of the variances, which is considered to be the most important assumption of parametric tests, is violated (Ergin and Koskan (2023)). (Zimmerman and Zumbo (1993)), reported that the heterogeneity of variances in experiments where the number of observations is not equal (unbalanced design) causes the probability of Type I error determined at the beginning of the experiment to not be maintained at 5%.
The parametric alternative to the Student t-test is the Welch t-test, which was developed by correcting the degrees of freedom of the independent two groups t-test in experiments where group variances were not homogeneous Derrick et al.(2016). Additionally, Winter (2013) supported through a simulation study that applying the Welch t-test on experiments with very small sample sizes is problematic.
The Wilcoxon Mann-Whitney test is one of the commonly used two sample means tests and sometimes considered as the nonparametric counterpart of the Welch t-test. In a study conducted by (Tsagaris et al (2020)), the Wilcoxon Mann-Whitney test was manifested to be highly inaccurate in terms of type I error, even if the exact p-value was calculated. In addition the exact p-value cannot be computed when ties are present in the data and it was later concluded that Wilcoxon Mann-Whitney test should not be considered as a competing non-parametric alternative to Welch test.

Sedgwick (2015) conducted a study to test the effectiveness of corticosteroids in reducing respiratory disorders in infants born at 34-36 weeks’ gestation and two broad categories of statistical methods were used; parametric and non-parametric tests.

Assumptions such as Normality about the distribution of the data and equal variance in the birth weight and treatment groups were made when using the parametric test, but none needed to be made when using the non-parametric test. He compared the treatment groups in mean birth weight using the Student's t-test (independent samples t -test) which is a parametric method. The Apgar score at five minutes was measured on the ordinal scale; therefore, the distributional assumption of normality could not be made and the Student's t-test could not be used. The Mann-Whitney U test the non-parametric equivalent of the Student's t-test was used instead.

The t-test has a little more power when it is valid, and the result is easier to interpret, and the alternate hypothesis is more simply stated: specifically that the means of the group differ. The t-test is robust against departures from normality as long as the distribution is reasonably symmetric (West (2021)). West (2021) carried out a simple test, he drew samples for group 1 (n=11), having a normal distribution with mean 4.0 and standard deviation 1.0, and for the group2 (n=22) with mean 3.0 and standard deviation 1.5. The sample gave means of 4.18 and 2.84. Student t-test falsely assumed that the standard deviations in each sample were equal, and

the test statistic t= 2.79, with 31 degrees of freedom so that P = 0.009. Welch's t-test was found to be valid, and it yielded a test statistic t = 3.174 with 27.9 degrees of freedom, so that P = 0.004. West RM also used the Mann-Whitney U-test and found P = 0.009. He came to a conclusion that the Welch t-test is preferred to the Student's t-test, whenever the distribution of measurements is close to normal or symmetric with at least 50 measurements. There is little difference in statistical power and the Mann-Whitney U-test is almost as powerful and has no distributional assumptions. He finally concluded by saying Welch t-test tests for a difference in means while the Mann-Whitney U-test tests for a difference in medians.

Takiar (2021) carried out a study to evaluate the performance of the t-test as compared to the Z-test in testing the significant or non-significant differences between two sample means. He generated four Normal populations (Population A, B, C and D) and then drew 30 samples each form the populations (thereby having a sample size of 120). Overall, the study covered 14400 comparisons to test for significant differences and 18240 comparisons for non-significant differences between means. At α=5%, the validity of the t-test remained below 50% for picking the significant difference between two sample means. The t-test performed far better when it came to testing the expected non-significant differences between two sample means and the validity was observed to be more than 94%. Low validity of t-test, especially in picking up the expected significant differences, suggested that probably, there is a need to raise the level from 5% to 20% to improve overall validity of the t-test. This was also true in the case of Z-test. In view of Z-test performing better as compared to t-test in picking up the significant differences, correctly, and not lagging behind much in picking up the non-significant differences between two sample means, suggests that Z-test can be used even for small sample sizes in place of hitherto used t-test.

Similarly, Takiar (2023) conducted another study to evaluate the performance of t-test, Mann-Whitney test as compared to Z-test in testing the possible significant or non-significant differences between two sample means or between a sample mean and the population mean. 500 samples of size 10, 6 and 3 were drawn from predefined four Normal populations. Overall, the study covered 18000 comparisons between sample means and the respective population mean. It also covered an equal number of comparisons for testing the possible significant differences between two sample means by three selected significance tests. It was discovered that for samples of size 10, at α=5%, t-test can pick up only 31.1% of the expected significant differences between two sample means which decreased to 11.4% for the sample size 3. This suggested that t-test is not valid when the sample size is 10 or below. In comparison, at α=5%, for the sample size of 10, Mann-Whitney test showed the validity of 30.4% while Z test with estimated variance (Z-EV) showed the validity of 39.9%. At Sample size 3, the validity of Mann-Whitney test and Z-EV test is observed to be less and is 20.1% and 31.5%, respectively. In view of very low validity observed, it was concluded that neither t-test nor Mann-Whitney test is suitable to be used when the sample size is 10 or below. In view of higher negative validity seen for Z-EV test as compared to t-test in the present study, as well as in his previous study for the sample size 9, 13 and 20, it was recommended that for sample size above 10 and below 30, Z-EV test can be used in the place of t-test, preferably with = 10%

2.4 CONCLUSION

This chapter delved into the realm of statistical methods used to compare means between two groups. We explored the commonly employed t-test, also known as "Student's t-test," which thrives under the assumptions of normally distributed data and equal variances in both groups. However, we learned that this method can be susceptible to Type I errors if the assumptions are violated.

When these assumptions falter, alternative options emerge. The Welch t-test addresses unequal variances, offering a robust solution for maintaining accurate results. However, for data with non-normal distributions, the Mann-Whitney U-test emerges as a non-parametric alternative, free from the constraints of normality assumptions.

Through careful consideration of the data's characteristics and adherence to appropriate assumptions, researchers can leverage these valuable tools to effectively compare means and draw sound conclusions from their investigations.

The following facts are vital:

- The t-test is a powerful tool for comparing means, but requires adherence to assumptions of normality and equal variances.
- The Welch t-test offers an alternative when variances differ.
- The Mann-Whitney U-test provides a non-parametric solution for non-normal data.
- Careful selection of the appropriate method is crucial for obtaining reliable results when comparing means.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

In this chapter, we delve into three principal tests for comparing the means of two normally distributed populations: the **Welch's t-test,** the **Z-test,** and the **Independent t-test**. Also outlining the scenarios where each test is most applicable.

3.2 STUDY DESIGN

In this study a quantitative research design was employed to investigate the difference in the means of two normally distributed populations. The research design involves the following steps

1 Formulation of Hypothesis:

$H_0 : \mu_1 = \mu_2 = 0$

$H_1 : \mu_1 \neq \mu_2$

Where

μ_1 is the mean of the first population

μ_2 is the mean of the second population

H_0 is the null hypothesis which states that there is no significant difference between the population means

H_1 is the alternative hypothesis which states that there is a significant difference between the population means.

2 Test statistics:

Given the nature of our data and research questions, we chose the t-test, Welch's t-test and z-test approximation for our analysis. These tests were selected based on considerations of data normality, variance homogeneity, and the size of our dataset.

- **Independent t-test:** Compares means of two groups

<u>Algorithm</u>

Step1. Calculate the test statistics by using the formula;

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$, assuming equal variances

Where $\bar{y}_1$ and $\bar{y}_2$ are the means of the two groups, and are calculated as;

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

n_1 and n_2 are the sizes of the groups

s^2_1 and s^2_2 are the variances of the two groups, and are calculated as;

$$s^2 = \frac{\sum(y-\bar{y})^2}{n-1}$$

Step2. Determine the degrees of freedom using the formula; $df = n_1+n_2-2$

Step3. Determine the critical value from the t-distribution table for the chosen level of significance(α) and degrees of freedom(df)

Step4. Compare the value of the test statistics with the critical value:

- If ($|t_{cal}| > |t_{critical}|$), reject the null hypothesis.
- Otherwise, fail to reject the null hypothesis.

- **Z-test:** Tests difference between two population means with known variances, but in our study the population variance is unknown and was estimated from the sample data using;

<u>Algorithm</u>

Step1. Calculate the test statistics by using the formula;

$$Z = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where

Z= z-stat

$\bar{y}_1$= sample mean of 1st dataset,

$\bar{y}_2$ = sample mean of 2nd-dataset. They are calculated using the formula below;

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

s^2_1= variance 1st dataset,

s^2_2= variance 2nd dataset. They are calculated using the formula below;

$$s^2 = \frac{\sum(y-\bar{y})^2}{n-1}$$

n_1= sample size 1st population

n_2 = sample size 2nd population

Step2. Determine the critical value from the standard normal distribution table for the chosen level of significance(α)

Step3. Compare the value of the test statistics with the critical value:

- If ($|Z_{cal}| > |Z_{critical}|$), reject the null hypothesis.
- Otherwise, fail to reject the null hypothesis.

- **Welch t-test:** Similar to independent t-test but for unequal variances and/or sample size

  Algorithm

  Step1. Calculate the test statistics using the formula;

$$t=\frac{(\bar{y}_1-\bar{y}_2)}{\sqrt{(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2})}}$$

Where $\bar{y}_1$ and $\bar{y}_2$ are the means of the two groups, n_1 and n_2 are the sizes of the groups, and

s^2_1 and s^2_2 are the variances of the two groups.

Step2. Determine the degrees of freedom;

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Step3. Determine the critical value from the t-distribution table for a chosen significance level(α) and degrees of freedom(df).

Step4. Compare the value of the test statistics with the critical value:

- If ($|t_{cal}| > |t_{critical}|$), reject the null hypothesis.
- Otherwise, fail to reject the null hypothesis.

3 Sampling Design:

This study employs a stratified random sampling approach to ensure that the sample accurately represents the two distinct subgroups within the NBA player population: guards and forwards in the 2022-2023 season. The primary objective of the sampling design is to compare the average heights of these subgroups, thereby necessitating a method that ensures both groups are adequately and fairly represented.

- **Population:** The broader population encompasses all active NBA players; however, for the purpose of this study, attention is exclusively directed towards active players classified as guards and forwards during the 2022-2023 season. Although the NBA features a variety of positions that contribute to the dynamics of the game, this research narrows its scope to these two categories to investigate differences in height, that are hypothesized to be positionally influenced.

- **Stratification:** To ensure a rigorous comparative analysis, the sampling process stratifies the extensive pool of NBA players into two specific groups: guards and forwards. This methodological foundation ensures that the ensuring analysis delivers position-specific insights and captures significant differences in physical attributes between these roles.

- **Sample Size Determination:** In this study sample sizes of 9, 20 and 31 were selected to test the performance (power) of the three tests on small, medium and large sample sizes.
- **Sampling Procedure:**

1. **List Compilation**: An exhaustive list of active NBA players identified as guards or forwards for the 2022-2023 season was compiled from the official NBA statistics and player registry (nba.com). This list serves as the foundation for the sampling process.
2. **Randomized Selection**: Within each targeted group (guards and forwards), 9, 20 and 31 players were selected through a computer-generated random number sequence, ensuring that each player had an equal chance of being included in the sample.
3. **Data Compilation**: The official NBA profiles provided the height data for the selected players. These records are regarded as accurate, reflecting the players' height without shoes, as officially measured by the NBA.

Rationale: Employing stratified random sampling is crucial for conducting nuanced, position-specific height comparisons within the NBA's 2022-2023 player roster. Focusing on guards and forwards allows the study to investigate the influence of positional roles on players' physical stature, enhancing the analysis's relevance and representativeness. Random selection within these strata further strengthens the study's integrity by reducing selection bias.

Considerations:

1. The methodology presumes that the player roster and corresponding height data for the 2022-2023 season are current and accurate. Discrepancies or roster changes could influence the sample's representativeness

2. By selectively examining guards and forwards, the study acknowledges the hypothesis that these positions may have distinct height profiles, warranting an in-depth exploration.

3.3 STUDY QUESTIONS

Our investigation is driven by a set of study questions that not only seek to uncover differences in mean heights between guards and forwards in the NBA but also aim to critically assess the effectiveness of various statistical tests in analyzing these differences. These questions serve as the foundation for our methodological approach:

1. **Primary Study Question:** "What are differences in mean heights between guards and forwards in the NBA during 2022-2023 season?"

- This question necessitates a comparison of two independent samples within our dataset, guiding our choice of statistical tests suitable for such analyses.

2. **Methodological Study Questions:**

- How do results of the t-test, Welch's t-test, z-test compare when applied to our dataset?

- Which statistical test provides the most reliable and valid results for our analysis, given the data characteristics and assumptions inherent in each test?

3.4 IMPLEMENTATION OF HYPOTHESIS TESTING ON COMPUTER USING R

The hypothesis testing procedure is implemented on a computer using the following steps:

1. **Data Preparation**

In preparation for our analysis, we undertook a comprehensive data preparation phase. This phase was critical or ensuring the data's integrity and relevance to our study's objectives, which involve comparing the mean heights of NBA guards and forwards.

- **Data source identification**: The official NBA statistics database served as our primary data source, chosen for its comprehensive and up-to-date records of player heights and positions for the 2022-2023 season.
- **Data Scope**: We focused exclusively on active players identified as guards or forwards, aligning our data collection with the study's specific objectives.
- **Data Storage and Organization**: The acquired data were structured into a CSV file, facilitating straightforward data management and analysis within the R environment. This file included essential details such as player names, positions and heights.
- **Preliminary Data Review**: An initial examination of the dataset was conducted to identify any anomalies or missing values, setting the stage for comprehensive data cleaning and validation.

2 **Data Import and Pre-processing**

Following data preparation, we imported the dataset into R using the **read.csv( )** function. The data cleaning phase was crucial for ensuring the quality of our dataset, Involving steps such as:

- **Handling Missing Values**: Utilizing R's **na.omit( )** function, we removed any records with missing height values to maintain the integrity of our analysis.
- **Verification of Player Positions**: We confirmed the accuracy of the player position labels, ensuring that our dataset accurately reflected the guards and forwards categories for analysis.

3 **Conducting the Tests**

We utilized R's statistical capabilities to conduct our hypothesis test:

- **T-Test and Welch's T-Test**: Through R's **t.test( )** function, we performed both the standard t-test and Welch's t-test, adjusting parameters to cater to our data's characteristics.
- **Z-Test**: We approximated a z-test for large samples, manually calculating the z-statistic and corresponding p-value based on our dataset's mean, standard deviation, and sample size.

3 **Analysis and Interpretation**

The outcomes of our statistical tests were critically evaluated against a 0.05 significance level. We interpreted p-values in this context to determine the statistical significance of height differences between guards and forwards, discussing both statistical and practical implications of our findings.

4 **Reporting Results**

Our analysis results were comprehensively summarized and presented, highlighting the key statistical findings and their implications. We leveraged R’s graphical tools, such as **hist( )**, to visually illustrate the height distributions and the differences between guards and forwards, enhancing the clarity and interpretability of our findings.

CHAPTER 4

ANALYSIS

4.1 INTRODUCTION

In this chapter, we delve into the analysis phase, where we employ statistical methods to draw meaningful insights from the data collected. This chapter focuses on three fundamental hypothesis testing techniques: the z-test, t-test, and the Welch's t-test. These tests serve as a crucial tool for comparing sample means, making inference about populations, and determining the significance of observed differences. The R (version 4.2.2) was used in performing the series of statistical tests presented in this chapter.

4.2 DESCRIPTION AND VISUALIZATION

The table below shows the descriptive statistics of NBA players (guards, forwards) height in inches for the 2022-2023 season.

Table 1.0

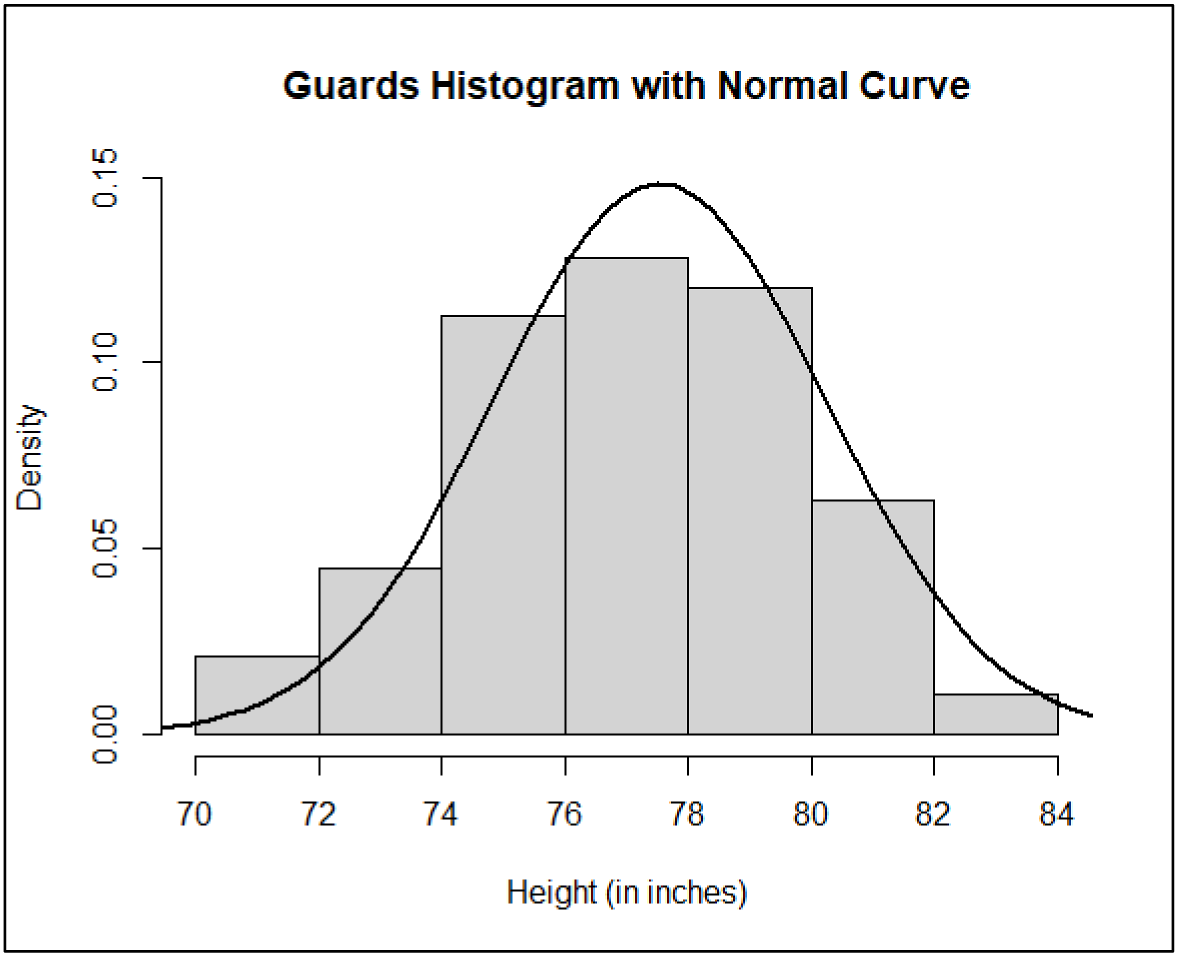
Height (in inches)

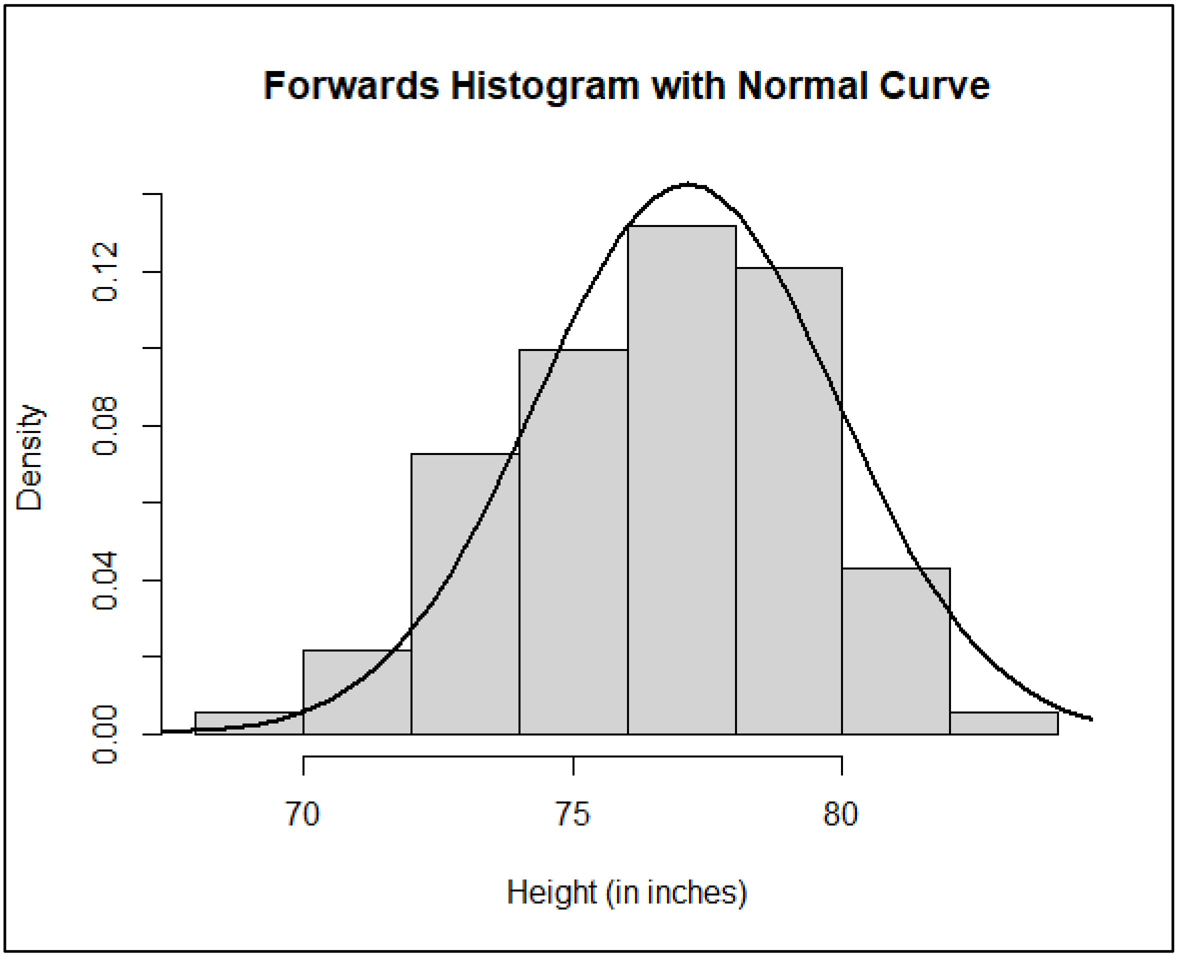
Position	Population size	Mean	Median	Standard Deviation
Guards	213	77.529	77	2.623
Forwards	164	77.113	77	2.796
Total	377			

Interpretation: From the table above we can see that the total population size is 377, where guards take up to 213 of the population size and forwards take up to 164.

The mean height of guards is 77.529, median is 77 while the standard deviation is 2.623, and for the forwards the mean height is 77.113, median is 77 and standard deviation is 2.796.

4.2.1 NORMAL DISTRIBUTION CURVE





4.3 ANALYSIS AND INTERPRETATION

In this analysis sample sizes of 9, 20 and 31 were taken from both population resulting to a total number of 6 samples.

4.3.1 Study Objective 1

Is there any difference in the mean height of guards and forwards?

H_o: There is no significant differences in the mean height of guards and forwards.

H_1: There is a significant differences in the mean height of guards and forwards.

NOTE: A α (significance level) of 5% was chosen for this research.

1. **Independent Two Sample T-test**

Code: The code below is used for determining some useful statistics using Independent two sample t-test in which our the p-value is a part of and it is our variable of interest and the results are displayed at the table after the code.

```
# T-TEST
ttest_result1 = t.test(guards_h1,forwards_h1, var.equal = TRUE) # n1,n2 = 9
print(ttest_result1)

ttest_result2 = t.test(guards_h2,forwards_h2, var.equal = TRUE) # n1,n2 = 20
print(ttest_result2)

ttest_result3 = t.test(guards_h3,forwards_h3, var.equal = TRUE) # n1,n2 = 31
print(ttest_result3)
```

Table1.1

n_1	n_2	$\bar{y}_1$	$\bar{y}_2$	t	Df	P-value
9	9	75.667	79.556	-4.055	16	0.0009194
20	20	75.55	80.1	-7.6981	38	2.852×10^{-9}
31	31	75.097	79.581	-9.2557	60	3.674×10^{-13}

2. **Welch's T-test**

Code: The code below is used for determining some useful statistics using Welch's t-test in which our the p-value is a part of and it is our variable of interest and the results is displayed at the table after the code.

```
Welch_result1 = t.test(guards_h1,forwards_h1, var.equal = FALSE) # n1,n2 = 9
print(Welch_result1)

Welch_result2 = t.test(guards_h2,forwards_h2, var.equal = FALSE) # n1,n2 = 20
print(Welch_result2)

Welch_result3 = t.test(guards_h3,forwards_h3, var.equal = FALSE) # n1,n2 = 31
print(Welch_result3)
```

Table 1.2

n_1	n_2	$\bar{y}_1$	$\bar{y}_2$	Welch's t-test	Df	P-value
9	9	75.667	79.556	-4.055	15.879	0.0009324
20	20	75.55	80.1	-7.6981	35.24	4.694×10^{-9}
31	31	75.097	79.581	-9.2557	49.043	2.439×10^{-12}

3. Z-test

Code: The code below is used for determining some useful statistics using z-test in which our the p-value is a part of. The p-value is our variable of interest and the results are displayed on the table after the code.

```r
#1st n1,n2 =9
mean1 = mean(guards_h1)
mean2 = mean(forwards_h1)
sd1 = sd_height_guards
sd2 = sd_height_forwards
n1 = length(guards_h1)
n2 = length(forwards_h1)

SE = sqrt(sd1^2/n1+ sd2^2/n2)
z_stat = (mean1-mean2)/SE
p_value = 2*pnorm(-abs(z_stat)) # two-tailed test
 print(p_value)

#2nd n1,n2 = 20
mean11 = mean(guards_h2)
mean21 = mean(forwards_h2)
n11 = length(guards_h2)
n21 = length(forwards_h2)
SE = sqrt(sd1^2/n11+ sd2^2/n21)
z_stat1 = (mean11-mean21)/SE
p_value1 = 2*pnorm(-abs(z_stat1)) # two-tailed test
print(p_value1)

#3rd n1,n2 = 31
mean12 = mean(guards_h3)
mean22 = mean(forwards_h3)
n12 = length(guards_h3)
n22 = length(forwards_h3)
SE = sqrt(sd1^2/n12+ sd2^2/n22)
z_stat2 = (mean12-mean22)/SE
p_value2 = 2*pnorm(-abs(z_stat2)) # two-tailed test
print(p_value2)
```

Table 1.3

n_1	n_2	$\bar{y}_1$	$\bar{y}_2$	σ_1	σ_2	P-value
9	9	75.667	79.556	2.623	2.796	0.002653
20	20	75.55	80.1	2.623	2.796	1.5914×10^{-7}
31	31	75.097	79.581	2.623	2.796	1.2679×10^{-10}

INTERPRETATION

Looking at the p-values on **Table 1.1, Table 1.2** and **Table 1.3** one can see that they are less than our chosen level of significance **(0.05)**. Therefore the null hypothesis was rejected by the three tests used for the three different sample sizes. We then conclude that there is a significant difference in the mean height of guards and forwards.

4.3.2 Study Objective 2

Which statistical test provides the most reliable and valid results for our analysis, given the data characteristics and assumptions inherent in each test? The power analysis would be used to address this objective.

The effect size was first determined before carrying out the power analysis, the code for the effect size is given below.

Code:

```
#  calculate effect size (d)
#for t-test
# 9
pooled_sd = sqrt(((9-1)*s_d1^2+s_d2^2*(9-1))/9+9-2)
d = (mean1-mean2)/pooled_sd
print(d)
# 20
pooled_sd1 = sqrt(((20-1)*s_d11^2+s_d21^2*(20-1))/20+20-2)
d1 = (mean11-mean21)/pooled_sd1
print(d1)
# 31
pooled_sd2 = sqrt(((31-1)*s_d12^2+s_d22^2*(31-1))/31+31-2)
  d2 = (mean12-mean22)/pooled_sd2
  print(d2)

#for  Welch t-test
  # 9
w_pooled_sd = sqrt((s_d1^2+s_d2^2)/2)
  w_d = (mean1-mean2)/w_pooled_sd
print(w_d)
# 20
w_pooled_sd1 = sqrt((s_d11^2+s_d21^2)/2)
  w_d1 = (mean11-mean21)/w_pooled_sd1
  print(w_d1)

# 31
w_pooled_sd2 = sqrt((s_d12^2+s_d22^2)/2)
  w_d2 = (mean12-mean22)/w_pooled_sd2
  print(w_d2)

```

```
#for z-test
# 9
z_pooled_sd = sqrt((sd1^2+sd2^2)/2)
 z_d = (mean1-mean2)/z_pooled_sd
 print(z_d)
# 20
z_pooled_sd1 = sqrt((sd1^2+sd2^2)/2)
z_d1 = (mean11-mean21)/z_pooled_sd1
print(z_d1)
# 31
z_pooled_sd2 = sqrt((sd1^2+sd2^2)/2)
z_d2 = (mean12-mean22)/z_pooled_sd2
print(z_d2)

```

Power Analysis: The code below was used to obtain the power of each test for the different sample sizes.

```r
#Power analysis
library(pwr)
# t-test
# 9
power_t1 = pwr.t.test(n=9, d = d, sig.level= 0.05,
                      type = "two.sample", alternative = "two.sided")
print(power_t1)
# 20
power_t2 = pwr.t.test(n=20, d = d1, sig.level= 0.05,
                      type = "two.sample", alternative = "two.sided")
print(power_t2)
# 31
power_t3 = pwr.t.test(n=31, d = d2, sig.level= 0.05,
                      type = "two.sample", alternative = "two.sided")
print(power_t3)

```

```r
# welch-test
# 9
powerwelch = function(alpha = 0.05, sigma1, sigma2, n1, n2, deltaw){
  se1 = sigma1/sqrt(n1)
  se2 = sigma2/sqrt(n2)
  s =sqrt(se1^2+se2^2)
  df = (se1^2+se2^2)^2/ ((se1^4 / (n1-1)) + (se2^4/(n2-1)))
  tcr = qt(1-alpha,df)
  delta_standardized = deltaw/s
  powert = 1-pt(tcr-delta_standardized,df)
  return(powert)
}
power_t = powerwelch(alpha = 0.05, sigma1=s_d1,
                     n1=9,sigma2=s_d2, n2=9, deltaw=abs(w_d)) #9
print(power_t)
power_t1 = powerwelch(alpha = 0.05, sigma1=s_d11,
                      n1=20,sigma2=s_d21, n2=20, deltaw=abs(w_d1)) #20
print(power_t1)
power_t2 = powerwelch(alpha = 0.05, sigma1=s_d12,
                      n1=31,sigma2=s_d22, n2=31, deltaw=abs(w_d2)) #31
print(power_t2)
```

```r
# z-test
# 9twosample
powerztest = function(alpha= 0.05, sd1, n1, sd2, n2, delta){
  zcr = qnorm(p=1-alpha,mean=0, sd=1)
  s = sqrt((sd1^2/n1)+ (sd2^2/n2))
  power = 1-pnorm(q= zcr, mean = delta/ s, sd=1)
  return(power)
}
power_z = powerztest(alpha = 0.05, sd1=sd1,
                     n1=9,sd2=sd2, n2=9, delta=abs(z_d)) #9
print(power_z)
power_z1 = powerztest(alpha = 0.05, sd1=sd1,
                      n1=20,sd2=sd2, n2=20, delta=abs(z_d1)) #20
print(power_z1)
power_z2 = powerztest(alpha = 0.05, sd1=sd1,
                      n1=31,sd2=sd2, n2=31, delta=abs(z_d2)) #31
print(power_z2)
```

Table 1.4

Sample Size		Independent Sample T-test	Welch's T-test	Z-test
n1	n2	Power	Power	Power
9	9	0.534	0.596	0.291
20	20	0.806	0.989	0.604
31	31	0.825	0.998	0.757

Interpretation: Based on the power analysis conducted, it is evident that the Welch t-test consistently outperforms both the standard t-test and the z-test across varying sample sizes. The power values obtained for the Welch t-test indicate a higher likelihood of correctly detecting true effects, thereby enhancing the reliability and validity of the statistical analysis. This superiority of the Welch t-test can be attributed to its ability to accommodate unequal variances between groups, which is a common scenario encountered in real-world data analysis. By robustly addressing this issue, the Welch t-test minimizes the risk of Type I and Type II errors, ensuring more accurate inference and interpretation of the results. Therefore, based on the observed power values and considering the data characteristics and assumptions inherent in each test, it is recommended to employ the Welch t-test for hypothesis testing in this particular analysis. This choice not only enhances the credibility of the findings but also highlights the importance of selecting a statistical test that aligns with the underlying data structure and assumptions, ultimately leading to more robust and defensible conclusions within the project.

CHAPTER 5

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 SUMMARY

The chapters provide an overview of statistical methods for hypothesis testing, used for comparing means between two populations. Specifically, the t-test, Welch's t-test, and z-test are discussed in detail.

The t-test is a parametric method that assumes normally distributed data and equal variances between groups. However, if these assumptions are violated, it can lead to Type I errors. The Welch's t-test is a robust alternative that accounts for unequal variances, while the Mann-Whitney U-test is a non-parametric option for non-normal data.

The research design involved formulating hypotheses, selecting appropriate test statistics (t-test, Welch's t-test, and z-test), determining sample sizes, and implementing the tests using R programming. The study focused on comparing the mean heights of NBA guards and forwards during the 2022-2023 season.

The analysis revealed a significant difference in mean heights between guards and forwards, with guards being slightly taller on average. A power analysis was conducted to assess the reliability and validity of the tests, considering their assumptions and the data characteristics.

5.2 CONCLUSION

Based on the power analysis, the Welch's t-test consistently outperformed the standard t-test and z-test across varying sample sizes. It demonstrated higher power values, indicating a greater ability to detect true effects while minimizing Type I and Type II errors. The superiority of the Welch's t-test is attributed to its robustness in accommodating unequal variances between groups, a common scenario in real-world data analysis.

5.3 RECOMMENDATION

Given the observed power values and the data characteristics, it is recommended to employ the Welch's t-test for hypothesis testing in this particular analysis. The Welch's t -test enhances the credibility of the findings by minimizing the risk of errors and ensuring accurate inference and interpretation of results. Selecting an appropriate statistical test that aligns with the underlying data structure and assumptions is crucial for robust and defensible conclusions within the project.

Additionally, it is essential to carefully consider the assumptions and limitations of each statistical method and choose the most suitable one based on the research objectives and data characteristics. This approach ensures the validity and reliability of the analysis, ultimately leading to more robust and meaningful conclusions.

REFERENCES

Alexander N and Eugene K (2023). A Distance Based Two-Sample Test of Means Difference for Multivariate Datasets. doi: 0.220944/preprints202311.1732.v2.

Arbuthnott, J. (1710). An Argument for Divine Providence, Taken from the Constant Regularity Observ'd in the Births of Both Sexes. By Dr. John Arbuthnott, Physitian in Ordinary to Her Majesty, and Fellow of the College of Physitians and the Royal Society. Philosophical Transactions, 27(325-336), 186-190.

Cohen, J. (1988). Statistical poweranalysis for the behavioral sciences (2^{nd} ed.). Lawrence Erlbaum Associates Publishers, USA.

Dawczynski.C, Weidauer.T, Richert.C, Schlattmann.P, Dawczynski.K and Kiehntopf.M. (2022). Nutrient intake and Nutrition Status in Vegetarians and Vegans in Comparison to Omnivores- the Nutritional Evaluation (NuEva) Study. https://www.frontiersin.org/articles/10.3389/fnut.2022.819106/full

Derrick B, Toher D and White P (2016). Why Welch's test is type 1 error robust. The Quantitative Methods in Psychology,12(1). Doi:10.20982/tqmp.12.1.p030

Lehmann E.L (1993), The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?, Journal of the American Statistical Association, 88(424) (Dec., 1993), 1242-1249

Ergin M and Koskan O (2023). Comparison of Student-t, Welch's t, and Mann_Whitney U Tests in Terms of Type I Error Rate and Test Power. Selcuk Journal of Agriculture and Food Sciences. 37 (2), 223-231.DOI:10.15316/SJAFS.2023.022. https://dergipark.org.tr/tr/pub/selcukjafsci

Fisher, R. A. (1925). Statistical methods for research workers. Oliver and Boyd.

Edinburgh.

Fisher, R.A. (1973). Statistical Methods and Scientific Inference, third edition. Collier Macmillan. London.

Gauss, C. F. (1809). Theoria motus corporum coelestium in sectionibus conicis solem ambientium.

Gosset, W. S. (1908). The probable error of a mean. Biometrika, 6(1), 1-25.

Historical Hypothesis Testing. Utah State University Mathematics Department: StatsStuff. Accessed February 6, 2024. https://www.usu.edu/math/schneit/StatsStuff/Inference/HistoricalHT/.

Hypothesis Testing. (https://gabrieletolomei.wordpress.com/miscellanea/probability-statistics/hypothesis-testing/).

Knuth, D. E. (1973). The art of computer programming. Volume 1: Fundamental algorithms. Addison-Wesley.Reading, Massachusetts (MA).

Michell, J. (1767). An Inquiry into the Probable Parallax, and Magnitude of the Fixed Stars, from the Quantity of Light Which They Afford us, and the Diminution of their Light, in its Passage from so Immense Distances. Philosophical Transactions of the Royal Society of London, 57, 234-264.

Moore D, McCabe G and Craig B (2009). Introduction to the practice of statistics 10th edition. W.H. Freeman and Company. Houndmills, Basingstoke RG21 6XS, England.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 236(767), 333-380.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London. A, 185, 71-110.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 187, 253-318.

Popper, K. R. (1959). The logic of scientific discovery. Hutchinson, London.

Sedgwick, P. (2015). A comparison of parametric and non-parametric statistical tests. BMJ: British Medical Journal, 350. https://www.jstor.org/stable/26520708
Statistical Resources, Hypotheses. National University Academic Success Center (ASC) (https://resources.nu.edu/statsresources/hypothesis).

Student (1908). The probable error of a mean. Biometrika, 6(1), 1-25.

Takiar R (2021). The Validity of t-test and Z-test for Small One sample and Small Two Sample test. Bulletin of Mathematics and Statistics Research. 9(4).

Takiar R (2023). The Validity of t-test, Mann_Whitney test and Z-test for Testing Significant difference between two Small Means when sample size is 10 or below. BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH. Vol.11.Issue.2.2023 (April-June).

The Royal Mint. (2016). The History of the Trial of the Pyx. Retrieved from www.royalmint.com/discover/uk-coins/history-of-the-trial-of-the-pyx

Tsagris M, Alenazi A, Verrou K.M, Pandis N. (2020). Hypothesis testing for two population means: parametric or non-parametric test? Journal of statistical computation and simulation 90 (2), 252-270.

Weakliem, David L. (2016). Hypothesis testing and model selection in the social sciences
Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. Biometrika, 34(1/2), 28-35.

West RM. Best practice in statistics: Use the Welch t-test when testing the difference between two groups. Annals of Clinical Biochemistry. 2021;58(4):267-269. doi:10.1177/0004563221992088

Winter JCF (2013). Using the Student's t-test with extremely small sample sizes. Practical Assessment, Research, and Evaluation Practical Assessment, 18(1): 10. Doi:10.7275/e4r6-dj05

Zimmerman DW, Zumbo BD (1993). Rank transformations and the power of the Student t test and Welch t test for non-normal populations with unequal variances. Canadian Journal of Experimental Psychology, 47(3): 523. Doi:10.1037/h0078850

APPENDICES

R Codes: Data Extraction and Cleaning

```r
df =read.csv("C:/Users/user/Documents/Projectdata/dataproj/players.csv",
             header= TRUE)
na.omit(df)
View(df) # displays the data in a spreadsheet
new_df = subset(df, position %in% c("Guard",
"Forward")) # filters the data for just Guard and Forward
new_d1 = subset(df, position %in% c("Guard"))
new_d2 = subset(df, position %in% c("Forward"))
View(new_d1)
View(new_d2)
feet_inches = strsplit(as.character(new_df$height), "-")
height_inches = sapply(feet_inches, function(x)
  as.numeric(x[1]) * 12 + as.numeric(x[2]))
print(height_inches)
var(height_inches)
sample_data = new_df
# Descriptive statistics
mean_height_guards= mean(height_inches[new_df$position == "Guard"])
mean_height_forwards= mean(height_inches[new_df$position == "Forward"])
median_height_guards= median(height_inches[sample_data$position == "Guard"])
median_height_forwards= median(height_inches[new_df$position == "Forward"])
sd_height_guards= sd(height_inches[new_df$position == "Guard"])
sd_height_forwards= sd(height_inches[new_df$position == "Forward"])
```

```
# Print descriptive statistics
print(paste("mean guards:", mean_height_guards))
print(paste("median guards:", median_height_guards))
print(paste("sd guard:", sd_height_guards))

print(paste("mean forwards:", mean_height_forwards))
print(paste("median forwards:", median_height_forwards))
print(paste("sd forwards:", sd_height_forwards))
```

```
# Visualizations
guards_height = height_inches[sample_data$position == "Guard"]


# Creates the histogram and adjust y-axis limits for density curve visibility
hist(guards_height, freq = FALSE, breaks =6,
     main = "Guards Histogram with Normal Curve", xlab = "Height (in inches)",
     ylim = c(0, max(dnorm(seq(min(guards_height), max(guards_height),
                               length = 100),
                           mean(guards_height), sd(guards_height)))))

# Adds the density curve with appropriate x-axis range
curve(dnorm(x, mean = mean(guards_height), sd = sd(guards_height)),
from = min(guards_height) - 3*sd(guards_height),
to = max(guards_height) +
  3*sd(guards_height),  # Extend x-axis to cover entire dataset
add = TRUE, lwd = 2)

#forwards
forwards_height = height_inches[sample_data$position == "Forward"]
hist(forwards_height, freq = FALSE, breaks =6,
     main = "Forwards Histogram with Normal Curve", xlab = "Height (in inches)",
     ylim = c(0, max(dnorm(seq(min(forwards_height),
                               max(forwards_height), length = 100),
                           mean(forwards_height), sd(forwards_height)))))

# Adds the density curve with appropriate x-axis range
curve(dnorm(x, mean = mean(forwards_height), sd = sd(forwards_height)),
      from = min(forwards_height) - 3*sd(forwards_height),
      to = max(forwards_height) +
        3*sd(forwards_height),  # Extend x-axis to cover entire dataset
      add = TRUE,lwd = 2)
```

```r
# for the first scenario i.e small sample size
# Randomly sample 9 players from each position
set.seed(123) # to set the seed for generating random numbers
guards1 = new_df[new_df$position == "Guard",]
forwards1 = new_df[new_df$position == "Forward",]
sample_guards1 = guards1[sample(nrow(guards1), 9), ]
sample_forwards1 = forwards1[sample(nrow(forwards1), 9), ]

df1 = rbind(sample_guards1, sample_forwards1)

# for the second scenario i.e moderate sample size
# Randomly sample 20 players from each position
set.seed(123) # to set the seed for generating random numbers
guards2 = new_df[new_df$position == "Guard",]
forwards2 = new_df[new_df$position == "Forward",]
sample_guards2 = guards2[sample(nrow(guards2), 20), ]
sample_forwards2 = forwards2[sample(nrow(forwards2), 20), ]

df2 = rbind(sample_guards2, sample_forwards2)
```


```r
# for the second scenario i.e moderate sample size
# Randomly sample 20 players from each position
set.seed(123) # to set the seed for generating random numbers
guards2 = new_df[new_df$position == "Guard",]
forwards2 = new_df[new_df$position == "Forward",]
sample_guards2 = guards2[sample(nrow(guards2), 20), ]
sample_forwards2 = forwards2[sample(nrow(forwards2), 20), ]

df2 = rbind(sample_guards2, sample_forwards2)
```


```r

# for the third scenario i.e moderate sample size
# Randomly sample 31 players from guard and 20 from forward
set.seed(123) # to set the seed for generating random numbers
guards3 = new_df[new_df$position == "Guard",]
forwards3 = new_df[new_df$position == "Forward",]
sample_guards3 = guards3[sample(nrow(guards3), 31), ]
sample_forwards3 = forwards3[sample(nrow(forwards3), 31), ]

df3 = rbind(sample_guards3, sample_forwards3)
View(df1)
View(df2)
View(df3)
```

Height data on forwards

<n= 9>

S/N	First name	Last name	Position	Height	Height in inches
1	Keita	Bates-Diop	Forward	6-8	80
2	Aleksej	Pokusevski	Forward	7-0	84
3	P.J.	Washington	Forward	6-7	79
4	Anthony	Lamb	Forward	6-6	78
5	Kawhi	Leonard	Forward	6-7	79
6	Patrick	Williams	Forward	6-7	79
7	E.J.	Liddell	Forward	6-6	78
8	Cole	Swider	Forward	6-9	81
9	Cody	Martin	Forward	6-6	78

<n=20>

S/N	First name	Last name	Position	Height	Height in inches
1	Isaiah	Jackson	Forward	6-9	81

2	John	Butler Jr.	Forward	7-0	84
3	Marvin	Bagley III	Forward	6-10	82
4	Cole	Swider	Forward	6-9	81
5	James	Johnson	Forward	6-7	79
6	Nikola	Jovic	Forward	6-10	82
7	Dorian	Finney-Smith	Forward	6-7	79
8	Khris	Middleton	Forward	6-7	79
9	Cedi	Osman	Forward	6-7	79
10	Cameron	Johnson	Forward	6-8	80
11	Isaiah	Todd	Forward	6-9	81
12	Robert	Covington	Forward	6-7	79
13	Marcus	Morris Sr.	Forward	6-8	80
14	Zion	Williamson	Forward	6-6	78
15	Ziaire	Williams	Forward	6-9	81
16	LeBron	James	Forward	6-9	81
17	Jarrell	Brantley	Forward	6-5	77
18	RaiQuan	Gray	Forward	6-7	79
19	Jeremy	Sochan	Forward	6-8	80
20	Jarred	Vanderbilt	Forward	6-8	80

<n=31>

S/N	First name	Last name	Position	Height	Height in inches
1	Cameron	Johnson	Forward	6-8	80
2	Isaiah	Todd	Forward	6-9	81
3	Robert	Covington	Forward	6-7	79
4	Marcus	Morris Sr.	Forward	6-8	80
5	Marvin	Bagley III	Forward	6-10	82
6	Cole	Swider	Forward	6-9	81
7	LeBron	James	Forward	6-9	81
8	Jarrell	Brantley	Forward	6-5	77
9	Trendon	Watford	Forward	6-8	80
10	RaiQuan	Gray	Forward	6-7	79
11	Jeremy	Sochan	Forward	6-8	80
12	Jack	White	Forward	6-7	79
13	Jae	Crowder	Forward	6-6	78
14	Serge	Ibaka	Forward	6-11	83
15	Isaiah	Jackson	Forward	6-9	81
16	Thaddeus	Young	Forward	6-8	80
17	Gordon	Hayward	Forward	6-7	79

18	JT	Thor	Forward	6-9	81
19	Naji	Marshall	Forward	6-7	79
20	Kawhi	Leonard	Forward	6-7	79
21	Ousmane	Dieng	Forward	6-9	81
22	Leandro	Bolmaro	Forward	6-6	78
23	Kessler	Edwards	Forward	6-7	79
24	Anthony	Lamb	Forward	6-6	78
25	Tobias	Harris	Forward	6-7	79
26	MarJon	Beauchamp	Forward	6-7	79
27	Eugene	Omoruyi	Forward	6-6	78
28	Kenneth	Lofton Jr.	Forward	6-6	78
29	Deni	Avdija	Forward	6-9	81
30	Kevin	Knox II	Forward	6-7	79
31	Lindy	Waters III	Forward	6-6	78

Height data on guards

<n= 9>

S/N	First name	Last name	Position	Height	Height in inches
1	Joshua	Primo	Guard	6-6	78
2	Donovan	Williams	Guard	6-6	76
3	Landry	Shamet	Guard	6-4	77
4	Bogdan	Bogdanovic	Guard	6-5	72
5	Kemba	Walker	Guard	6-0	75
6	Derrick	Rose	Guard	6-3	73
7	Devon	Dotson	Guard	6-1	77
8	Theo	Maledon	Guard	6-5	75
9	Matthew	Dellavedova	Guard	6-3	73

<n=20>

S/N	First name	Last name	Position	Height	Height in inches
1	Joshua	Primo	Guard	6-6	78
2	Donovan	Williams	Guard	6-6	78
3	Landry	Shamet	Guard	6-4	76
4	Bogdan	Bogdanovic	Guard	6-5	77
5	Kemba	Walker	Guard	6-0	72
6	Derrick	Rose	Guard	6-3	75
7	Devon	Dotson	Guard	6-1	73
8	Theo	Maledon	Guard	6-5	77
9	Matthew	Dellavedova	Guard	6-3	75
10	Delon	Wright	Guard	6-5	77
11	Alondes	Williams	Guard	6-4	76
12	Cameron	Payne	Guard	6-1	73
13	Jrue	Holiday	Guard	6-5	77
14	Talen	Horton-Tucker	Guard	6-4	76
15	Duane	Washington Jr.	Guard	6-2	74
16	Blake	Wesley	Guard	6-4	76
17	Ish	Smith	Guard	6-0	72
18	Caleb	Houstan	Guard	6-8	80
19	Malik	Monk	Guard	6-3	75
20	Reggie	Jackson	Guard	6-2	74

<n=31>

S/N	First name	Last name	Position	Height	Height in inches
1	Joshua	Primo	Guard	6-6	78
2	Donovan	Williams	Guard	6-6	78
3	Landry	Shamet	Guard	6-4	76
4	Bogdan	Bogdanovic	Guard	6-5	77
5	Kemba	Walker	Guard	6-0	72
6	Derrick	Rose	Guard	6-3	75
7	Devon	Dotson	Guard	6-1	73
8	Theo	Maledon	Guard	6-5	77
9	Matthew	Dellavedova	Guard	6-3	75
10	Delon	Wright	Guard	6-5	77
11	Alondes	Williams	Guard	6-4	76
12	Cameron	Payne	Guard	6-1	73
13	Jrue	Holiday	Guard	6-5	77
14	Talen	Horton-Tucker	Guard	6-4	76
15	Duane	Washington Jr.	Guard	6-2	74
16	Blake	Wesley	Guard	6-4	76
17	Ish	Smith	Guard	6-0	72

18	Caleb	Houstan	Guard	6-8	80
19	Malik	Monk	Guard	6-3	75
20	Reggie	Jackson	Guard	6-2	74
21	Jalen	Green	Guard	6-4	76
22	Facundo	Campazzo	Guard	5-10	70
23	D.J.	Augustin	Guard	5-11	71
24	Jeenathan	Williams	Guard	6-5	77
25	Vince	Williams Jr.	Guard	6-4	76
26	Austin	Reaves	Guard	6-5	77
27	R.J.	Hampton	Guard	6-4	76
28	Gary	Harris	Guard	6-4	76
29	Lindell	Wigginton	Guard	6-1	73
30	Tre	Jones	Guard	6-1	73
31	Kyle	Lowry	Guard	6-0	72

Data source : https://www.kaggle.com/datasets/szymonjwiak/nba-active-players-data-images