

TU DORTMUND

INTRODUCTORY CASE STUDIES

Project 2: Comparison of Multiple Distributions

Lecturers:

Prof. Dr. Jörg Rahnenführer

Dr. Franziska Kappenberg

M. Sc. Marieke Stolte

Author: Prerana Rajeev Chandratre

Group number: 5

Group members: Janani Veeraghavan, Bushra Kiyani, Sathish
Ravindranath Kabatkar, Shivam Shukla

December 9, 2022

Contents

1	Introduction	1
2	Problem Statement	2
2.1	Data Source	2
2.2	Variables and Data Quality	2
2.3	Project Objectives	3
3	Statistical Methods	3
3.1	Hypothesis Testing	3
3.2	Assumptions for One-Way Variance Analysis	5
3.2.1	Normality using Q-Q Plots	5
3.3	One-Way ANOVA	6
3.4	Two Sample T-Test	7
3.5	Multiple Testing Problem	8
3.5.1	Bonferroni Correction	8
3.5.2	Bonferroni-Holm Correction	9
4	Statistical Analysis	9
4.1	Descriptive Analysis and Assumption's Test	9
4.2	Global Test : One way ANOVA	14
4.3	Pairwise T-Test and Bonferroni Correction	14
5	Summary	17
	Bibliography	19
	Appendix	20
A	Additional Figure	20

1 Introduction

The collection of sports data is one of the most important aspects of sports competition integrity. The analysis of sports data is taking sports beyond game performance. A vast amount of data is available in today's world because of advancements in technology. Recently, the European aquatics championship took place in Rome namely "Roma 2022". The European Aquatics Championship 2022 invited best athletes from all over the world and held 75 events in the divisions of swimming and diving.

A small dataset (MicroPlus, 2022) from the information collected at the European Aquatics Championship 2022 (MicroPlus, 2022) is considered for this project. It consists of data on women's semifinals in 200 meter in five different categories. The categories given in the data set are breaststroke, butterfly, backstroke, medley and freestyle. Name and finishing time of the swimming of every participant in five mentioned categories are given in the dataset. Analysis of the completion time of the competitions is the main focus of this project.

The finishing time of swimming for each category is examined under three conditions in order to perform the tasks of this project. The three conditions that the data has to fulfil are normality examined via Q-Q plots, independence and homogeneity examined via descriptive analysis. After checking for these assumptions a global test using one-way ANOVA is performed to check for the differences in the finishing time between the categories. Additionally, the swimming time of each pair of categories were compared using a pairwise t-test. To reduce the possibility of false positive results, test values were adjusted using the Bonferroni and Bonferroni-Holm methods. The tests that were performed in this report shows that there is a difference in the finishing time of the players in each category. It also demonstrates that some pairs of categories have a difference in their finishing time of the competition.

This report has five sections in total including the introduction section. Section 2 explains the dataset, the way it is collected, quality of data, and provides an overview of the objective of the project. Section 3 talks about the methods used for statistical analysis. Here the concepts of standard deviation, ANOVA, hypothesis testing, p-value, t-test, Bonferroni correction method and Bonferroni-Holm procedure are explained. Visualization graphs like box plot and Quantile-Quantile plots are also explained in section 3. In the subsequent section 4, statistical methods are applied to the given data set and

the results are interpreted. The last section contains a review of the findings and the valuable takeaways from the analysis.

2 Problem Statement

2.1 Data Source

The European Aquatics Championships 2022, arranged in Rome from August 11 to 21, featured men’s and women’s athletics. The results of the women’s 200 meter semifinals in five categories such as butterfly, freestyle, backstroke, breaststroke, and medley are included in the given dataset (MicroPlus, 2022). It involves the data on swimming completion time for each category, with an overall of 80 observations, 16 in each category, along with three variables. Seven female athletes competed in more than one category. To avoid inconsistencies while performing the analysis, these seven observations were removed from the second category of their participation.

2.2 Variables and Data Quality

The given dataset (MicroPlus, 2022) contains 3 variables. The variables name and respective data type are given in below (Table 1).

Table 1: Variables and their data type

Variable (name in dataset)	Data Type
Category	Nominal
Name	Nominal
Time	Numeric(Continuous)

The variable *Category* describes five different categories of women’s athletics, namely butterfly, freestyle, backstroke, breaststroke, and medley. All female participants’ names are included in the variable *Name*. The variable *Time* specifies the exact swimming completion time for each participant by category.

2.3 Project Objectives

There are three main objectives of this project. First, conducting a descriptive analysis of the variables of given dataset. Later on, verifying the data for homogeneity, normality, and independence in order to set up one-way ANOVA as a global test, to find out whether the swimming completion time differs from category to category for each player. Third task is to perform two-sample pairwise t-test to check pairwise differences between each category's swimming completion time. In order to solve the multiple testing problems, previous t-test results are corrected using Bonferroni and Bonferroni-Holm methods. A comparison of test results with and without adjusting for multiple testing is done at the end.

3 Statistical Methods

In this section, all the statistical methods which are used for analysis of this project are discussed in detail with formulas. For this analysis tasks, software Jupyter Notebook (Fernando Perez, 2015) Version 6.3.0, R (R Core Team, 2020) Version 4.1.0 is used. The graphs were created using R packages namely ggplot(Wickham, 2016), dplyr (Wickham, 2022), GGally (Schloerke, 2020) and gridExtra (Auguie, 2017).

3.1 Hypothesis Testing

In statistics, sample data is used to infer underlying population parameters and their distribution. This approach is commonly known as inferential statistics. Hypothesis testing is an important method in an inferential statistics, in which a population parameter assumption or assertion is tested using a statistical test to analyze the true parameters based on the assumptions. The null hypothesis H_0 is an assumption about the population parameter or distribution that is used to start the test. The statement is tested on the sample data, producing a test statistic that indicates whether or not the null hypothesis should be rejected. An alternative hypothesis H_1 is a statement or assumption that contradicts the null hypothesis. A test statistic is a criterion used to determine whether to accept or reject a hypothesis. It is calculated using measurements from the observation. It is a function $t(x)$ of the sample values x_1, x_2, \dots, x_n that are provided, preferably in a way that maximizes the difference between the distribution

$f(t | H_0)$ and the distributions that belong to H_1 . H_0 is rejected if $t(x) \in R$ for a specified rejection region R . on the other hand, we fail to reject H_0 in favor of assumption condition if $t(x) \notin R$ (Sheldon (2010, p. 388)).

Hypothesis tests are also categorised as, one-tailed and two-tailed tests, as shown in the figure 4 in the appendix. When testing for the possibility of positive or negative differences, the two-tailed test uses both the positive and negative tails of the distribution for the rejection region. A one-tailed test determines whether there is a difference in the direction of the difference between groups. A critical value C defines the rejection region. A null hypothesis is either rejected or not rejected when comparing the test statistic $t(x)$ with the crucial value C . For the two-tailed test, the rejection region is $R = (-\infty, C_1) \cup (C_2, +\infty)$ where C_1 and C_2 represents lower and upper limit of the critical region respectively. The rejection region is either, $R = (-\infty, C)$ or $R = (C, +\infty)$ for the one-tailed test (Sheldon (2010, p. 394-395)).

The conclusions drawn from the hypothesis testing are not entirely certain. In the actual world, an assumption made about a hypothesis based on the data's observations may be incorrect. There are two types of error which are considered in hypothesis testing: Type-I and Type-II error. A type-I (false positive) error occurs when a null hypothesis is rejected when the null hypothesis should not be rejected. A type-II (false negative) error is the failure to reject the null hypothesis when it should be rejected. The probability of making a type-I error is the significance level α . The significance level α of most statistical tests is established before the test. If the p-value derived from the test statistic is less than or equal to the significant level α , the null hypothesis is then rejected (Sheldon (2010, p. 388)). The below table (Table 2) represents the type-I and type-II errors based on the choice of null hypothesis.

Table 2: Type-I and Type-II error based on the hypothesis testing

	When H_0 is true	When H_0 is not true
Reject H_0	Type I error (α)	correct choice ($1-\beta$)
Fail to reject H_0	correct choice ($1-\alpha$)	Type II error (β)

Nevertheless, there is an another way to interprets results of a hypothesis test is by using the continuous parameter, p-value. It is the probability of observing sample data will be maximum as test statistics. A smaller value of p improves the evidence against H_0 . The p-value is the probability of the test statistic being the observed value or, more extreme,

if the null hypothesis is true. Based on the p-value, the null hypothesis can be rejected or not rejected (Sheldon (2010, p. 398)).

3.2 Assumptions for One-Way Variance Analysis

Three main assumptions are considered when performing the ANOVA test. The first is that the observations are random and the samples are taken from populations that are independent of each other. The second assumption is that the observations are taken from normally distributed populations. It can be verified with the help of Q-Q plots, which are discussed above. Lastly, it is assumed that the empirical variance of the data in different groups is equal. We can check this using a box plot (Sheldon (2010, p. 506-508)).

3.2.1 Normality using Q-Q Plots

To fulfil the purpose of calculation of cumulative distribution function, is to assume the normal distribution for the dataset. For assumption, the data can be plotted as k th smallest observation on an x-axis and the expected value of k th observation among all n standard normal distribution on the y-axis. For any mean and standard deviation, the graph can be straight line if the data is normally distributed. The graph that helps to make this task easy is called as "Quantile versus Quantile (Q-Q) plot" and has its own built in function in R i.e., *qqnorm*. In Q-Q plot if the outer part of the curve is deeper than the middle part then one can say that the distribution has long tails. For the Q-Q plot, a scatter plot is created by mapping two sets of quantiles against each other. If the two pairs of quantiles belong to the same distribution, the points form an approximate straight line. The straight line represents the point where the observed value is consistent with the theoretical value. As long as the points fall on this line, the data follow the assumed distribution. The given data points follow the assumed distribution if all pairs of quantiles lie (almost) on a straight line at an angle of 45 degree from the x-axis. In most cases, the Q-Q plot is used with a normal distribution or a Gaussian distribution (Dalgaard (2008, p. 74-75)).

3.3 One-Way ANOVA

The one-way analysis of the variance generally helps to determine the statistically significant differences among the mean values of two or more independent or unrelated groups. Assume k groups are being studied, and the one-way ANOVA test is used to test the hypothesis using the formula:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$$H_1 : \text{At least one of the means is different from the others}$$

The null hypothesis states that the means of all groups are the same. As per the alternative hypothesis, if at least one of the means deviates from the other means, the null hypothesis is rejected (Sheldon (2010, p. 505)).

Assume the i^{th} sample's mean and variance are denoted by \bar{X}_i and S_i^2 , respectively, for normally distributed sample data, each with size n and an unknown population variance σ^2 where $i = 1, 2, \dots, m$. By comparing the values of two estimators, σ^2 the test for the null hypothesis will be performed. The variance S_i^2 is unbiased estimator of σ^2 , without taking the null hypothesis into consideration. Then the average of the variances S_i^2 of all m samples indicates the first estimator of σ^2 with the formula:

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^m S_i^2,$$

The second estimator of σ^2 is evaluated with the assumption that the null hypothesis H_0 is true. As is assumed for the true null hypothesis (H_0), all m normally distributed samples have the equal mean μ and equal variance σ^2/n . Then the estimator $\hat{\sigma}^2/n$ given by the formula :

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^m S_i^2.$$

where, \bar{X} is average of all m mean values \bar{X}_i , which will be calculated as:

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m \bar{X}_i$$

Furthermore, it is assumed that, $n\bar{S}^2$ is second estimator of σ^2 . To calculate the value of f , which follows a F distribution with $m - 1$ numerator and $m(n - 1)$ denominator degrees of freedom, the suitable test statistic TS for one-way ANOVA. Then the value

of f using two estimators of σ^2 will be:

$$f = \frac{n\bar{S}^2}{\frac{1}{m} \sum_{i=1}^m S_i^2}.$$

For specific significance level of α , the p-value is calculated using,

$$p = 2P\left(F_{m-1, n(m-1), \alpha} \geq |f|\right),$$

where, $F_{m-1, n(m-1), \alpha}$ is F distributed random variable with $m-1$ numerator and $m(n-1)$ denominator degrees of freedom and chosen significance level of α then H_0 will be rejected if the p-value is less than or equal to the significance level of α (Sheldon (2010, p. 505-510))

3.4 Two Sample T-Test

As the name suggests, the two sample t-test is useful to compare the distribution which are come from the same mean of two independent populations. In general, it helps to test the hypotheses which were applied to the population. It is assume that, data comes from two groups, y_{11}, \dots, y_{1n_1} and y_{21}, \dots, y_{2n_2} with normal distributions, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively. The t-test analyzes t statistics and t distribution values to determine statistical significance. Here for the calculations, we have the requirement of three important data values to calculate the t-statistic, which are the number of data values in each group, the standard deviation in each group, and the difference between the means in each data set (referred to as the mean difference). It can be given by the formula:

$$t = \frac{\bar{y}_2 - \bar{y}_1}{\text{SEDM}}.$$

where SEDM stands for the *Standard Error of Difference of Means* which can be denoted as,

$$\text{SEDM} = \sqrt{\text{SEM}_1^2 + \text{SEM}_2^2}.$$

To calculate SEDM, there are two ways, first by assuming the same variance and consider the pooled s variance based on standard deviation. The t value can be calculated under the null hypothesis, which follows t-distribution and $n_1 + n_2 - 2$ degrees of freedom. The another approach is Welch procedure i.e., to calculate SEM for separate groups with standard deviation s_1 and s_2 . In this procedure the t is not actually defined by

t-distribution by it is defined approximately by t-distribution with n degrees of freedom (Dalgaard (2008, p. 100-102)).

Using the test statistic, the p-value can be computed and the hypothesis test is done which is explained above. Below are the null and alternative hypothesis for the t-test.

$$H_0 : \mu_{y_1} = \mu_{y_2}$$

$$H_1 : \mu_{y_1} \neq \mu_{y_2}$$

where μ_{y_1} and μ_{y_2} are the population means of the data set y_1 and y_2 respectively. The null hypothesis can be rejected or not rejected based on the p-value (Dalgaard (2008, p. 100-102)).

3.5 Multiple Testing Problem

If we run the t-test multiple times on the same data set, it may cause multiple testing problem. Consider the following scenario to understand the multiple testing problem. When we calculate the t test, we know that when the null hypothesis is indeed true, the type I error rate is most likely set to $\alpha = 0.05$. In other words, we will not throw a type I error (False Alarm) in $1 - 0.05$ of the cases when the null hypothesis is true. If we calculate two independent t tests, the probability of no false positives is $0.95^2 = 0.9$. Hence, the possibility of false positives increases with the number of comparisons. Generally speaking, the probability of making at least one type 1 error in an independent m test is:

$$1 - (1 - \alpha)^m.$$

Thus we need some correction method when we are performing t-tests multiple times same set of data (Bretz (2011, p. 20)). This will be discussed further in subsection 3.4.1 and 3.4.2 respectively.

3.5.1 Bonferroni Correction

Bonferroni test is a statistical test used to reduce the false positives. In particular, the Bonferroni method is designed to prevent the data from incorrectly appearing to be statistically significant. This method is considered to be the simplest but most conservative method for handling Type I errors. To implement Bonferroni correction, simply multiply the original p-values by the number of statistical tests to be performed

which can be considered to be n . The result of the equation is the Bonferroni-corrected p-value, and these new p-values must be less than α to be considered as a significant value (Bretz (2011, p. 31)).

$$\text{Corrected } p - \text{Value} = p * n$$

3.5.2 Bonferroni-Holm Correction

The Bonferroni-Holm method is also called the Holm method, which helps to handle multiple testing problems. It iteratively accepts and rejects the null hypothesis. It is similar to the Bonferroni method with a different threshold level. Assume α is the fixed significant threshold value for rejecting the null hypothesis and n is the number of hypotheses. We will select the lowest p-value and compare it to the α/n by iteratively arranging p-values alongside the n hypothesis. If the p-value is low, the null hypothesis is rejected, and the procedure is repeated on the unfinished $n - 1$ hypothesis with a threshold value of $\alpha/(n - 1)$. Will continue this iteration until the selected p-value is high enough that all remaining hypotheses should not be rejected any more. This method is preferred over the Bonferroni correction method because it uses $\alpha/n, \alpha/n - 1, \dots, \alpha$. Hence, the chances of rejecting the null hypothesis are less than or equal to those of the Bonferroni correction method (Bretz (2011, p. 32)).

4 Statistical Analysis

In this section, all the tests that were performed and the underlying assumptions required for this test are discussed. Interpretation of the results with regard to the problem statement are also discussed in this section.

4.1 Descriptive Analysis and Assumption's Test

In this section, the descriptive analysis is conducted to accommodate an overview of the dataset. Following that, an univariate analysis was run on the continuous variable time (in seconds) and the nominal variable category. All the values of central measure tendencies, i.e., mean, median, variance, standard deviation, and interquartile ranges are listed in the Table 3.

Table 3: Measures of Central Tendency over all Categories for Swimming Time(SD - Standard Deviation, IQR - Interquartile Range)

Category	Backstroke	Breaststroke	Butterfly	Freestyle	Medley
Median	131.115	146.680	130.820	119.500	133.455
Mean	131.380	146.400	131.675	119.358	134.040
SD	1.851145	1.490676	2.717344	1.558782	1.588752
Variance	3.426738	2.222114	7.383960	2.429803	2.524133
Min.	128.18	143.73	128.48	117.70	131.84
Max.	135.74	148.40	136.92	122.42	137.02
IQR	1.5175	2.1050	5.1000	2.2350	2.4750

The histogram in Figure 1, shows that the distribution of continuous variable time is symmetric with the mean and variance. To learn if the swimming finishing time differs or not between the categories, specific statistical tests were performed. At first, we chose one-way ANOVA method to conduct a global test. Here the one-way ANOVA test is the best fit since the dataset has one categorical and one continuous variable. Only when the data is independent, normally distributed, and has a constant variance, the one-way ANOVA test can be used. There is no evident relationship between each individual category in the data set. Hence within the scope of this study the data is assumed to be independent. That is, it is assumed that the data values are drawn independent of each other.

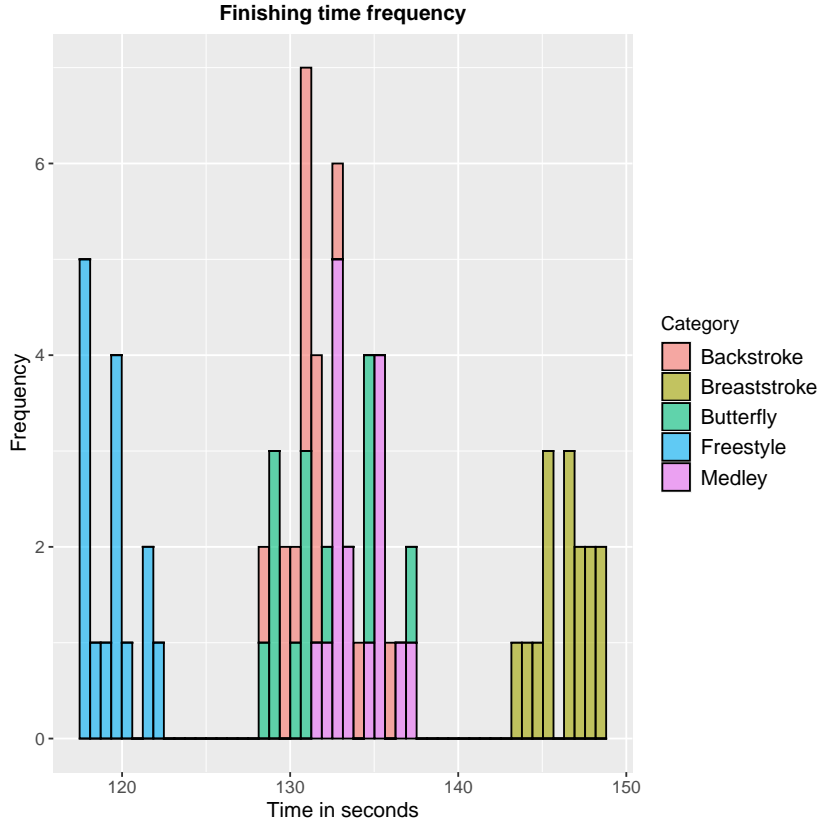
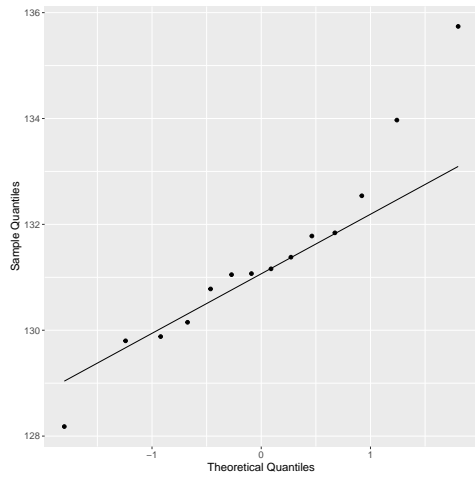
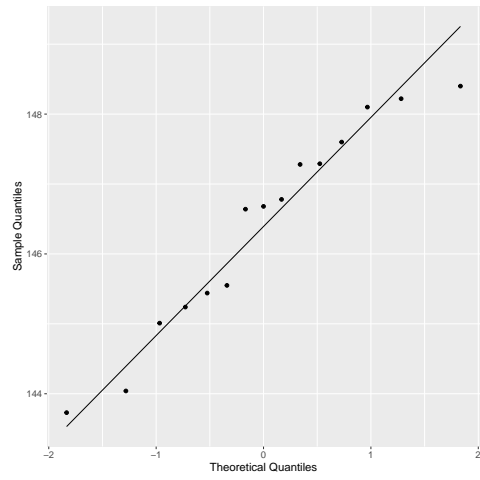


Figure 1: Histogram: Time in seconds for each category of swimming

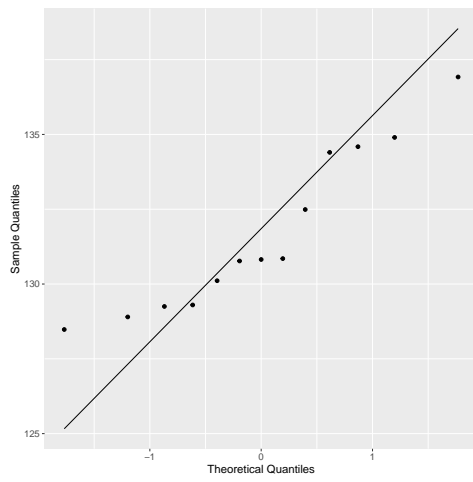
As mentioned previously, the given dataset contains five categories (Backstroke, Breaststroke, Butterfly, Freestyle, Medley) in athletic championship 2022. It has been observed that the values of time in seconds are not depending on any category and each data entry is independent for every competitor. Hence the values are independent. The normality is examined by the Q-Q plot to study the proportion of observed and predicted values, and is it matching with each other or not. The theoretical and sample quantiles are calculated and plotted against each other with a straight line. This straight line represents that the both sets of quantiles came from the same distribution. Figure 2 represents the Q-Q plots which are generated for each category. It interprets the linear relationship between the quantiles which shows the data is normally distributed. Some values are scattered, in every category. However, almost all values follow the straight line, which makes it easier to assume that the data is normally distributed among all the values.



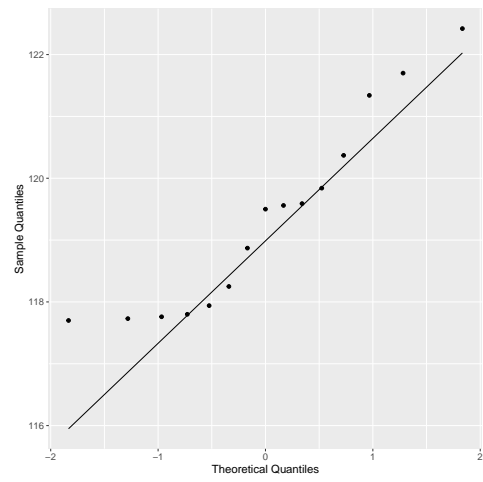
a) Backstroke



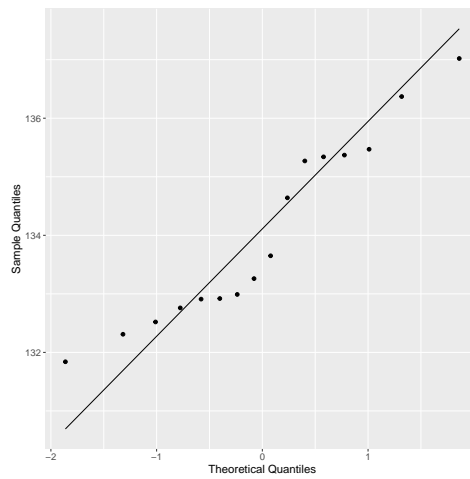
b) Breaststroke



c) Butterfly



d) Freestyle



e) Medley

Figure 2: Q-Q normality plots for the variable time in all five categories

To study the homogeneity of variance, empirical variances and the standard deviations of the data values were calculated for each group. Also, with the help of box plots, spread in the data is visualized. From the Table 3, all the values i.e. mean, median and variance is calculated for every category which shows in the box plot Figure 3. It is observed by the height of the boxes. The inter quartile ranges (IQR) for each categories are almost comparable, though not the same, with different median values. The median of the category Breaststroke contains highest values, whereas category Freestyle has the lowest median value. The variance of both categories are also close to each other, in the range of 2.222 to 2.429. This shows that there is not much difference in the variance between all categories. It is also worth noting that the category Butterfly has the highest variance with value 7.383. Also the median line for each category's box plot inside shows the homoscedastic (i.e.constant variance).

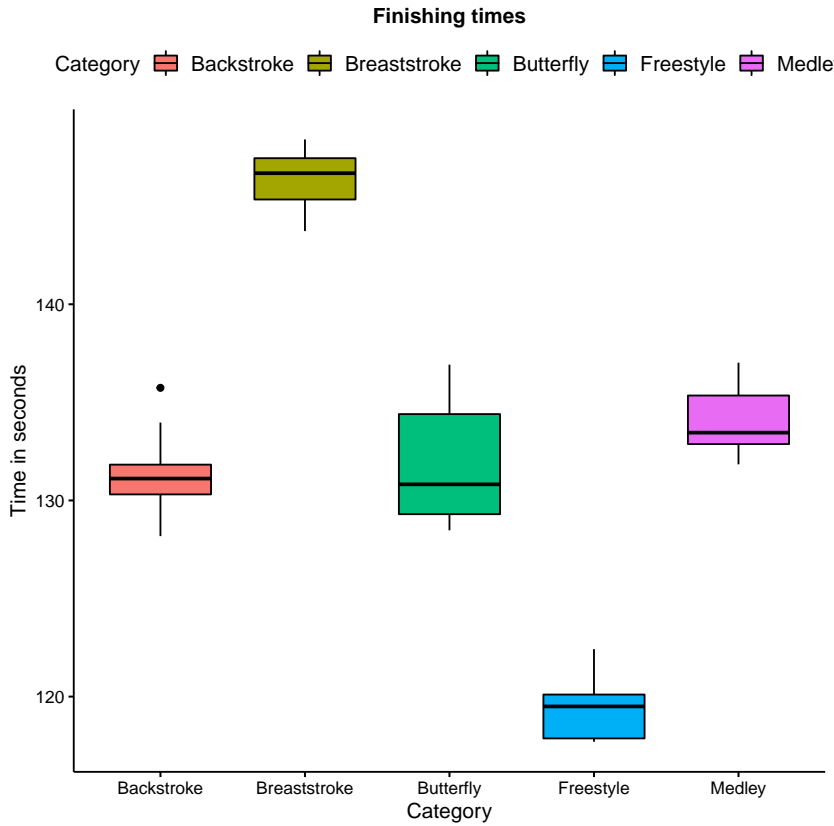


Figure 3: Boxplot: Represents finishing time in seconds against categories

4.2 Global Test : One way ANOVA

With the help of the above explanation, it is observed that the data is homoscedastic and independent. The null and alternative hypothesis were tested to study if there were any statistical differences in the mean values of the finishing time for each category. The results of the null hypothesis represent that the mean values of all five categories are equal, and the alternate hypothesis is taken as that all the means are not equal.

Holding all the assumptions made in the above section true, an original least square model is fit to the swimming finishing time for each category. The sum of the errors or the residuals and the mean sum of squares are calculated for the ANOVA method. The ANOVA output is presented in the Table 4, where sum-sq is the sum of squares, Mean-sq is the mean of squares, F is the F-statistic and the trailing values have been rounded off to three decimal places.

Table 4: One-way ANOVA output (Df - degrees of freedom, Sum-sq - sum of square error, Mean-sq - mean square error, Pr(>F) - Probability comparison with F-value)

	Df	Sum-sq	Mean-sq	F-Value	Pr(>F)
Category	4	5552	1388.0	399.7	<2e-16
Residuals	68	236	3.5		

For this test, as $p < \alpha$, the null hypothesis H_0 is rejected. At 5 percent significance level, the data provides strong evidence to indicate that roughly at least one pair of group (Backstroke-Butterfly) means is different from each other. The variation among the groups dominates over the variance within the groups. Therefore, the null hypothesis can be rejected and the mean values of finishing time of each categories are different to each other.

4.3 Pairwise T-Test and Bonferroni Correction

In this section, pairwise tests for comparing the means of each pair of categories are carried out. It is checked for a pairwise difference between the means of time per second, considering all pairs of categories. To analyze this, we performed a two-sample independent t-test. The results of this test are given in the Table 5. The pair of categories

along with their p-values are presented. A two-tailed, two sample t-test is conducted as the difference in the mean values. When means are unequal, they can tend to be positive or negative. The null hypothesis is proposed as, the means of the two categories groups in question being equal. The alternate hypothesis is stated as the means of the two categories groups not being equal. The assumptions to perform t-test are same as the ANOVA i.e. independence, normality, and homoscedasticity. In this two-tailed, two-sample t-test the value of significance level of $\alpha = 0.5$ is used. The results of these pairwise tests are included in Table 5. It represents that we rejected the null hypothesis (Yes) if the p-value is less than or equal to the significance level $\alpha = 0.05$ or do not rejected the null hypothesis (No) if the p-value is greater than the significance level 0.05. While performing pairwise test, we considered ten pairs of categories which are unique and independent.

Table 5: Tabuling p-values for pair of categories

Categories-Pair	p-values	Reject Yes/No
Backstroke-Breaststroke	8.180062e-20	Yes
Backstroke-Butterfly	7.424315e-01	No
Backstroke-Freestyle	3.900106e-17	Yes
Backstroke-Medley	2.220449e-04	Yes
Breaststroke-Butterfly	2.913045e-16	Yes
Breaststroke-Freestyle	1.411879e-28	Yes
Breaststroke-Medley	8.422435e-20	Yes
Butterfly-Freestyle	2.710592e-14	Yes
Butterfly-Medley	6.882146e-03	Yes
Freestyle-Medley	1.275451e-21	Yes

To address the multiple testing problem, Bonferroni correction is made by multiplying all the p-values. Table 6 shows the summery of the adjusted p-values after Bonferroni correction and the corresponding results after the adjustment. From the Table 5 and Table 6, we can state that, before applying the Bonferroni correction method, all the nine values for pair of all categories except "Backstroke-Butterfly" holds 'Yes', it means, $p\text{-value} < \alpha$ rejecting the null hypothesis. Decisions are made for rejecting H_0 based on these corrected p -values. When $p \leq \alpha$, the data has sufficient evidence against H_0 , so we reject H_0 . When $p > \alpha$, H_0 is not rejected. The results with and without Bonferroni

correction are compared. It shows that, only one pair of categories has the difference between the mean values of finishing time. Because of multiple testing problems, there might be several false positives among the other pairs of categories above, rejecting the null hypothesis. A Bonferroni correction method can be applied to the p -values, which values are acquire earlier with the false positives. While correcting here we multiplied all the p -values by the number of statistical tests performed. If these p -values $<$ significance level of 0.05, the test be considered statistically significant. The table 6 shows adjusted p -values after applying the Bonferroni correction method. It is observed that, for the pair of categories "Butterfly-Medley", the null hypothesis is not rejected, which was rejected before applying the Bonferroni correction method. It shows a wrong assumption due to the multiple testing problem in the t-test which is performed initially. Hence, after the Bonferroni correction the pairs of categories "Backstroke-Butterfly" and "Butterfly-Medley" have significant difference in the mean value of finishing time for each category.

Table 6: Adjusted p -values by Bonferroni method

Categories-Pair	Adjusted p -values	Reject Yes/No
Backstroke-Breaststroke	8.180062e-19	Yes
Backstroke-Butterfly	1.000000e+00	No
Backstroke-Freestyle	3.900106e-16	Yes
Backstroke-Medley	2.220449e-03	Yes
Breaststroke-Butterfly	2.913045e-15	Yes
Breaststroke-Freestyle	1.411879e-27	Yes
Breaststroke-Medley	8.422435e-19	Yes
Butterfly-Freestyle	2.710592e-13	Yes
Butterfly-Medley	6.882146e-02	No
Freestyle-Medley	1.275451e-20	Yes

Along with the Bonferroni correction, we used another method i.e. Holm-Bonferroni to overcome the multiple testing problem. The table 7 contains the adjusted p -values after applying the Holm correction method with corresponding results to reject the null hypothesis or not. It is observed that only the pair of categories "Backstroke-Butterfly" have difference in the mean value of finishing time for each category. This results are similar to the results of p -values table. So it is concluded with the help of both

Bonferroni and Holm correction method, the two categories i.e. "Backstroke-Butterfly" and "Butterfly-Medley" of swimming has an influence on the finishing time.

Table 7: Adjusted p-values by Holm-Bonferroni method

Categories-Pair	Adjusted p-values	Reject Yes/No
Backstroke-Breaststroke	6.544050e-19	Yes
Backstroke-Butterfly	7.424315e-01	No
Backstroke-Freestyle	2.340064e-16	Yes
Backstroke-Medley	6.661346e-04	Yes
Breaststroke-Butterfly	1.456523e-15	Yes
Breaststroke-Freestyle	1.411879e-27	Yes
Breaststroke-Medley	6.544050e-19	Yes
Butterfly-Freestyle	1.084237e-13	Yes
Butterfly-Medley	1.376429e-02	Yes
Freestyle-Medley	1.147906e-20	Yes

5 Summary

Sports is one of the important industry. Especially to collect the sports data for analysis purpose helps the players and coaches of professional teams to create a personalized training and plan other strategies for different competitions for better achievement of their players and teams. One of the major sports in the Olympics and world championships is athletics. It helps to improve the blood circulation and stimulates the brain cells in the human body. The European Aquatics Roma 2022 were recently held in Italy. The top athletes from around the world were invited to the European Aquatics Championship 2022, which featured 75 events in the swimming and diving divisions. In this project, we used a small portion of the data for analysis purposes, which included information on the 200 meter semifinals for women in five separate categories. The categories listed in the data set are freestyle, medley, backstroke, breaststroke, and butterfly. Every participant's name and swimming finish time in the five categories specified are provided in the dataset.

The main objective of this project was to find the differences between the finishing time in different categories and also to check the pairwise differences between them. Both the

tasks were implemented based on three assumptions. They are , variance homogeneity, independent and normality. All three assumptions were discussed in statistical method section. Initially a descriptive analysis was done on the data set to confirm those three assumptions. By excluding a few competitors' observations from their second category of competition, the disparities for the independence assumption were resolved. Afterwards, Q-Q plots were constructed to prove variance homogeneity. Once the assumptions are confirmed we performed global testing using one-way ANOVA method. The level of significance, α was set to 0.05 ANOVA method was tested to find the difference in mean values of finishing time of each category and the results of the test showed that there was significant differences in their mean.

In the end, pairs of categories were used to perform two-sample t-tests. In the null hypothesis, it was assumed that the two pair categories' mean values of time were equal. On the other hand, the alternative hypothesis made the assumption that these mean values are not equivalent for the same pair. All potential pairs of categories were subjected to this paired t-test, and the results were recorded for study. The results showed that, with the exception of one pair, "backstroke and butterfly," there is a significant difference in the average finishing times of nearly all pairings of categories. Since multiple tests were conducted, the multiple testing issue was addressed by adjusting the p-values of the results and findings using the Bonferroni and Holm methods. The outcomes of the Holm technique and the paired t-test were the same, according to these test results. The decision to accept the null hypothesis of the pair "butterfly and medley" was altered by the test results of the Bonferroni method. All other pairs of categories showed a significant variation in the average finishing times, with the exception of the "backstroke and butterfly" and "butterfly and medley" pairs. The results of the pairwise tests also suggest that the category of swimming has a significant influence on the finishing times. In future investigations, the effects of additional factors on the participants' swimming time could make significant difference. By carefully examining all other variables, like sports physiotherapy, player health, food and diet, the probability of winning championships may be enhanced.

Bibliography

- Auguie, B. (2017), *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.
- Bretz, H. (2011), *Multiple Comparision using R*, CRC Press. ISBN 978-1-58488-574-0.
- Dalgaard, P. (2008), *Introductory Statistics with R*, Springer. ISBN 978-0-387-79053-4.
- Fernando Perez, B. G. (2015), *Jupyter: Interactive data science and scientific computing in all programming languages* URL: <https://jupyter.org>. version 6.3.0.
- Ibe, O. C. (2014), *Fundamentals of Applied Probability and Random Processes* URL: <https://www.sciencedirect.com/topics/mathematics/tail-test> (visited on 8th December 2022).
- MicroPlus, V. (2022), *European Aquatics Championships 2022 Database* URL: <https://www.roma2022.eu/en/> (visited on 8th December 2022).
- R Core Team (2020), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Schloerke, B. (2020), *GGally: Extension to ggplot2*. R package version 2.1.2.
- Sheldon, M. (2010), *Introductory Statistics*, Elsevier. ISBN 978-0-12-374388-6.
- Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag. R package version 3.4.0.
- Wickham, H. (2022), *dplyr: A Grammar of Data Manipulation*. R package version 1.0.10.

Appendix

A Additional Figure




One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$
		

Figure 4: Hypothesis Testing: Types and Rejection/Acceptance Criteria (Ibe (2014))