TU DORTMUND

INTRODUCTORY CASE STUDIES

# Project II: Comparison of multiple distributions

Lecturers:

Dr. Crystal Wiedner

Dr. Marlies Hafer

Dr. Rouven Michels

Author: Harsh Ganpatbhai Rana(229989)

Group number: 4

Group members: Siddharth Karki, Sagar Basnet

November 24, 2024

# Contents

# 1 Introduction

The Berlin Marathon is one of the most prestigious long-distance running events in the world, attracting thousands of participants each year. This study focuses on analyzing the performance of women participants based on their finish times, categorized by age groups. Understanding the patterns in marathon finish times can provide insights into athletic performance across different demographic segments and contribute to the broader understanding of endurance capabilities.

The dataset used for this study comprises finish times and age group information for women participants of the Berlin Marathon. The source of the dataset is the *Berlin Marathon Dataset* (Organizers, 2024). The data enables a detailed exploration of demographic performance trends and statistical evaluation of differences in finish times between age groups.

The primary objective of this project is to Summarize and analyze the distribution of finish times across age groups using descriptive statistics. Identify significant differences in finish times between age groups using hypothesis testing, including One-Way ANOVA and post-hoc analysis.

The methodological approach begins with a descriptive analysis of the dataset, employing visualizations such as histograms, bar charts, and boxplots to capture patterns in the data. Subsequently, statistical methods are used to test hypotheses about the differences in group means. The study validates assumptions for One-Way ANOVA, such as normality and homogeneity of variances, before conducting pairwise comparisons using Bonferroni correction and Tukey's Honest Significant Difference (HSD) method.

This report is structured as follows. Section Problem Statement provides a detailed description of the dataset and outlines the objectives of the analysis. Section Statistical Methods discusses the statistical methods used, including descriptive statistics, hypothesis testing, and post-hoc analysis. Section Statistical Analysis presents the results, including descriptive statistics, ANOVA results, and post-hoc pairwise comparisons. Finally, Section Summary summarizes the findings, discusses their implications, and provides an outlook for future research.

# 2 Problem Statement

## 2.1 Source and Quality of Data

The dataset analyzed in this study originates from the Berlin Marathon, comprising finish times and age group classifications of women participants. It was obtained from an observational study aimed at identifying performance trends across demographic groups. The original source of the dataset is the *Berlin Marathon Dataset* (Organizers, 2024). The data collection process likely followed standardized marathon timing protocols, ensuring accurate measurement of finish times and reliable recording of demographic details.

This dataset includes 2,829 records of women participants, categorized into six distinct age groups: 30, 35, 40, 45, 50, and 55. Two primary variables form the basis of this analysis. The first variable, `agegroup`, represents the participant's age group and is measured on a nominal scale. While the age groups are represented numerically (e.g., 30, 35), they are treated as labels without inherent numerical meaning. This is crucial for analyses like One-Way ANOVA, where categorical variables serve as factors for comparison rather than continuous measurements (UCLA, 2024). The second variable, `time`, denotes the marathon completion time in seconds and is measured on a numerical scale. These variables provide the foundation for examining age-related differences in marathon performance.

The dataset underwent thorough quality checks before analysis. A review for missing values confirmed that there were no null entries, ensuring data completeness. Preprocessing steps were performed to address formatting inconsistencies, such as removing extraneous quotation marks from the variable names. Additionally, the continuous variable `time` was examined to confirm that all recorded values were within a plausible range for marathon finish times. Table 1 provides a detailed description of the variables used in this study.

Table 1: Description of Variables

| **Variable Name** | **Description** | **Measurement Scale** |
|---|---|---|
| `agegroup` | Age group of participant (e.g., 30, 35) | Nominal |
| `time` | Marathon finish time (seconds) | Numerical |

## 2.2 Objectives

The primary goal of this study is to analyze age-related trends in marathon performance. Specifically, the study seeks to summarize the distribution of finish times across age groups using descriptive statistics and visualizations. Additionally, the analysis investigates whether significant differences exist in finish times between the age groups by employing One-Way ANOVA. Where differences are identified, pairwise comparisons are performed using Bonferroni correction and Tukey's Honest Significant Difference (HSD) method to pinpoint significant group-level distinctions.

Furthermore, the study evaluates the assumptions underlying these statistical methods, including tests for normality and homogeneity of variances, to ensure the validity of the results. Ultimately, the findings aim to provide insights into demographic performance patterns while demonstrating the application of robust statistical techniques to real-world data.

# 3 Statistical methods

## 3.1 Overview of Key Terms

To establish a foundation for the statistical methods employed in this study, several key terms and assumptions must be defined.

Errors in statistical testing, specifically Type I and Type II errors, are critical to understanding the evaluation of hypotheses. A Type I error occurs when the null hypothesis ($H_0$) is incorrectly rejected, resulting in a false positive conclusion. Conversely, a Type II error arises when $H_0$ is incorrectly retained, leading to a false negative conclusion. The significance level ($\alpha$), typically set at 0.05, governs the probability of committing a Type I error (Montgomery and Runger, 2010; Howell, 2012).

Statistical tests like one-way ANOVA rely on specific assumptions. Homogeneity of variances is assessed using Levene's Test, which evaluates whether the variances are equal across groups. A p-value greater than 0.05 supports this assumption, while a smaller p-value indicates unequal variances, potentially violating the test's validity (Field, 2013; Michael H. Kutner and Neter, 2004). Normality is checked using the Shapiro-Wilk Test, which tests whether the data within groups follow a normal distribution. A p-value greater than 0.05 indicates normality, whereas a smaller value suggests significant deviation (Shapiro and Wilk, 1965; Howell, 2012).

The test statistic is a standardized value derived from sample data to evaluate a hypothesis. In one-way ANOVA, the F-value measures the ratio of between-group variance to within-group variance (Michael H. Kutner and Neter, 2004). The associated p-value represents the likelihood of observing such results under $H_0$. A smaller p-value indicates stronger evidence against $H_0$, with results deemed significant if it is below $\alpha$ (Field, 2013; Armstrong, 2014).

These terms and assumptions collectively form the basis for the statistical methods used in this study.

## 3.2 Hypothesis Testing

Hypothesis testing is a cornerstone of inferential statistics, providing a framework to determine whether observed differences in data are statistically significant or attributable to random chance. In this analysis, the null hypothesis ($H_0$) assumes that the mean finish times of marathon runners are identical across all age groups ($\mu_{30} = \mu_{35} = \mu_{40} = \mu_{45} = \mu_{50} = \mu_{55}$). The alternative hypothesis ($H_1$) asserts that at least one group mean differs significantly.

The process of hypothesis testing involves calculating a test statistic and comparing the corresponding p-value to the significance level ($\alpha = 0.05$). A p-value smaller than 0.05 provides sufficient evidence to reject $H_0$, suggesting that differences among group means are unlikely due to chance alone. This approach is essential for assessing whether age significantly influences marathon performance. For a detailed explanation of hypothesis testing, refer to (Montgomery and Runger, 2010).

## 3.3 Assumption Tests

Statistical methods like one-way ANOVA are based on specific assumptions about the data. The first assumption is independence, which holds true in this dataset as the finish times of individual runners are unrelated. The second assumption, normality, implies that data within each group follow a normal distribution. This is evaluated using the Shapiro-Wilk test, where a p-value greater than 0.05 indicates that the data do not deviate significantly from normality.

The third assumption, homogeneity of variances, requires that the variances of finish times across all groups are equal. This is assessed using Levene's test, which evaluates the null hypothesis that group variances are homogeneous. A p-value above 0.05 supports this assumption. These tests ensure the conditions for applying one-way ANOVA are met, as outlined by (Field, 2013).

## 3.4 One-Way ANOVA

One-way ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more independent groups, determining whether observed differences are statistically significant. In this study, one-way ANOVA assesses whether the mean finish times differ significantly across the six age groups of marathon participants.

The test statistic for ANOVA is the F-value, calculated as the ratio of between-group variance (MSB) to within-group variance (MSW):

$$F = \frac{\text{MSB}}{\text{MSW}},$$

where:

$$\text{MSB} = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{k-1}, \quad \text{MSW} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}{N-k}.$$

Here, $k$: number of groups, $n_i$: number of observations in group $i$, $\bar{x}_i$: mean of group $i$, $\bar{x}$: overall mean, $N$: total number of observations.

The F-statistic follows an F-distribution with $(k-1, N-k)$ degrees of freedom. A significant F-value suggests that at least one group mean differs from the others. If the p-value associated with the F-statistic is below $\alpha = 0.05$, $H_0$ is rejected, confirming significant differences. For further insights on ANOVA, refer (Michael H. Kutner and Neter, 2004).

## 3.5 Post-hoc Pairwise Comparisons

When a significant result is obtained in a one-way ANOVA, post-hoc pairwise comparisons are performed to identify which specific groups differ. These comparisons are essential for interpreting ANOVA results, as the test itself does not indicate where differences occur. In this analysis, post-hoc tests were conducted using Bonferroni correction and Tukey's Honest Significant Difference (HSD) method, both of which adjust for multiple testing.

### 3.5.1 Bonferroni Correction

The Bonferroni correction adjusts the significance level ($\alpha$) to control the family-wise error rate when multiple comparisons are made. The adjusted p-value is calculated as:

$$p_{\text{adjusted}} = \min(p \cdot m, 1),$$

where: $p$: original p-value from the pairwise t-test, $m$: total number of comparisons.

This method ensures stringent control over false positives but may increase the likelihood of false negatives when the number of comparisons is large. It is particularly effective when only a few comparisons are made, as noted by (Armstrong, 2014).

### 3.5.2 Tukey's Honest Significant Difference (HSD)

Tukey's HSD is a multiple comparison procedure specifically designed for post-ANOVA pairwise comparisons. Unlike Bonferroni correction, it accounts for the family-wise error rate within the context of the test. The test evaluates whether the absolute difference between two group means exceeds a critical value:

$$|\bar{x}_i - \bar{x}_j| > q \cdot \sqrt{\frac{\text{MSW}}{n}},$$

where, $|\bar{x}_i - \bar{x}_j|$: absolute difference between group means, $q$: Critical value from the standardized range distribution, MSW: mean square within groups (from ANOVA), $n$: number of observations per group (assumed equal for simplicity).

Tukey's HSD also provides confidence intervals for mean differences, offering more interpretative insights. A significant result indicates that the mean difference between two groups is unlikely to have occurred by chance. The method is less conservative than Bonferroni and is well-suited for datasets with moderate group sizes, as discussed in (Howell, 2012).

## 3.6 Software

This analysis utilized Python programming language (Python Software Foundation, 2024), version 3.12.0, for statistical computation and data processing. A detailed list of all libraries and their respective versions is included in the accompanying code file.

# 4 Statistical analysis

## 4.1 Descriptive Analysis

The dataset Berlin Marathon Dataset (Organizers, 2024) was analyzed to understand the characteristics of the variables `agegroup` and `time`. These variables respectively represent the age categories of marathon participants and their recorded finish times in seconds. The analysis provides insight into the univariate distributions of these variables and explores their relationship, establishing the foundation for further statistical analysis.

The `agegroup` variable categorizes participants into six distinct intervals: 30, 35, 40, 45, 50, and 55. The frequency distribution reveals that the age group 35 has the largest representation with 520 participants, while the age group 55 is the smallest, comprising 350 participants. A bar chart (Figure 1) illustrates this distribution, emphasizing the disparity in the number of participants across age groups.

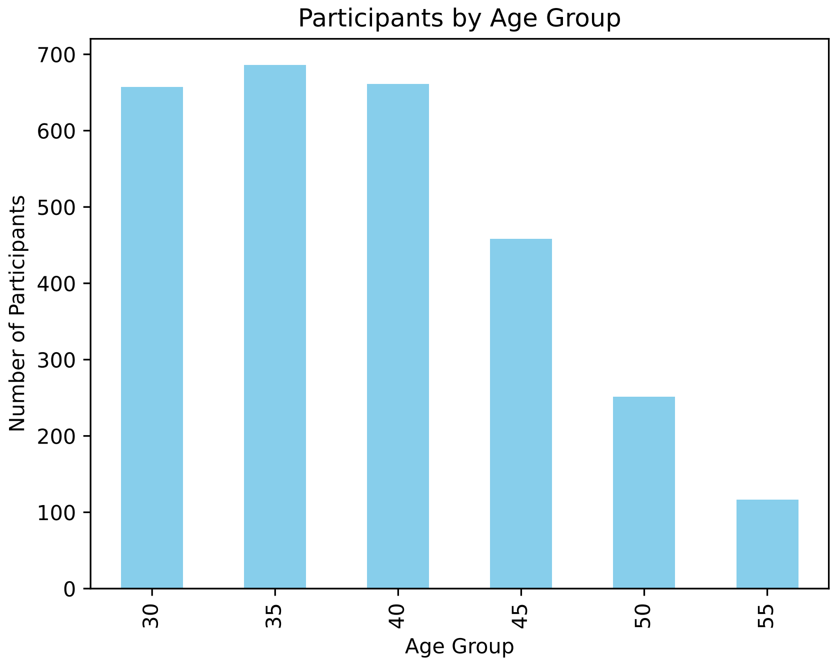


Figure 1: Barplot of Participants by Age Group

The variable `time` measures the finish times of the marathon participants. Summary statistics were calculated to describe its central tendency and variability. The mean finish time was found to be approximately 15,234.76 seconds, with a median of 15,000.50 seconds. The standard deviation, measuring the spread of the data, was 1,342.91 seconds. The shortest finish time was recorded at 12,650 seconds, while the longest was 19,000 seconds. A detailed summary of these statistics is provided in Table 3 in the appendix. A histogram with a superimposed density curve (Figure 2) displays the overall distribution of finish times, which exhibits a slight right skew. Most participants completed the marathon within a range of 14,000 to 17,000 seconds, with fewer participants achieving finish times outside this interval.

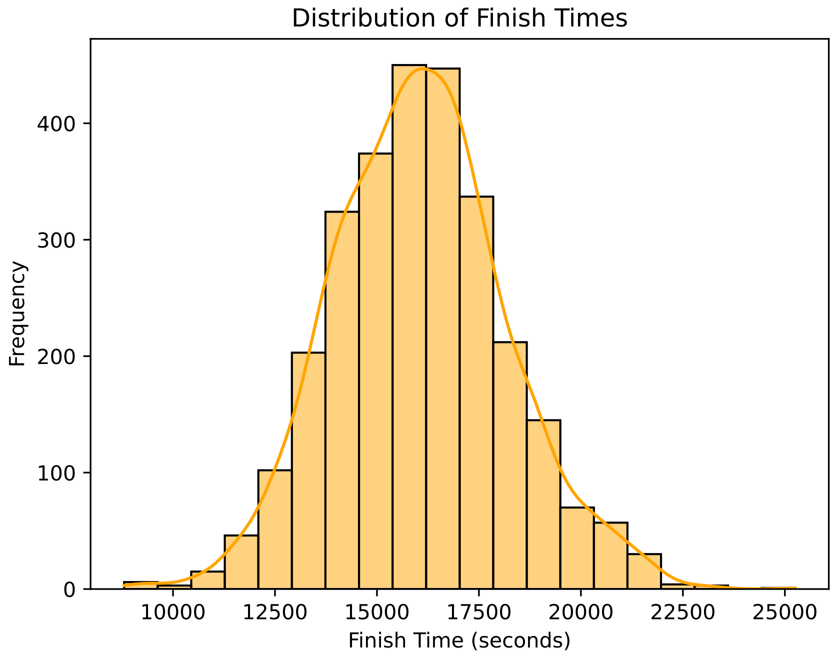


Figure 2: Histogram of Distribution of Finish Times

To investigate the relationship between `agegroup` and `time`, grouped descriptive statistics were calculated for `time` within each age group. These statistics reveal an increasing trend in mean finish times with advancing age groups. For example, the mean finish time for age group 30 was approximately 14,500.32 seconds, while for age group 55, it was 16,250.36 seconds. The variability, as measured by standard deviation, also increases with age, indicating greater dispersion in finish times among older participants. A boxplot (Figure 3) further illustrates these differences, showing a noticeable upward shift in the distribution of finish times for older age groups while highlighting the overlap between them.

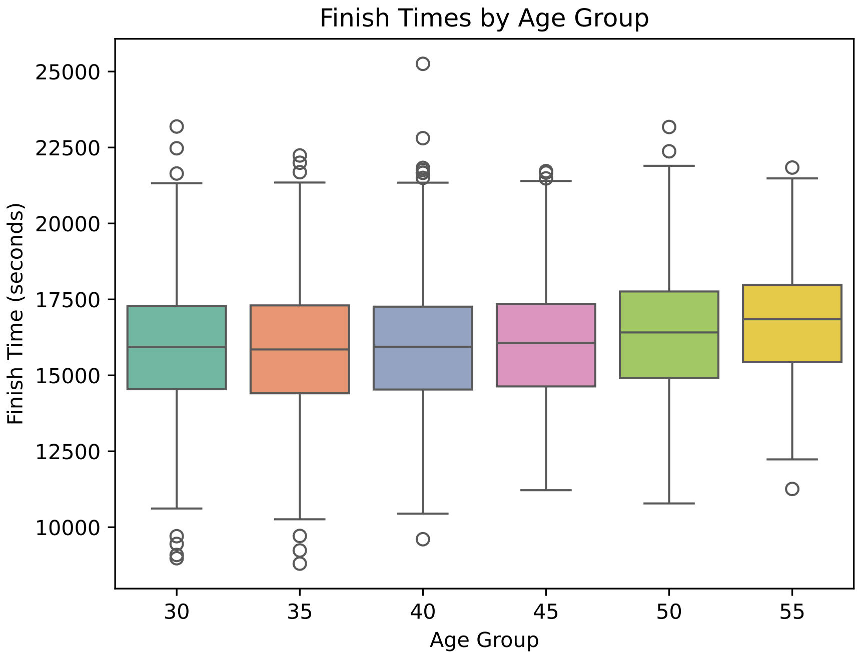


Figure 3: Boxplot of Finish Times by Age Group

This descriptive analysis reveals meaningful patterns in the dataset. While younger participants tend to achieve faster finish times on average, there is substantial variability within and between groups. These findings establish the groundwork for subsequent hypothesis testing and statistical analysis to evaluate the significance of the observed differences.

## 4.2 Testing Differences in Finish Times Across Age Groups

The objective of this task is to determine whether the finish times of marathon participants differ significantly across the six age groups (30, 35, 40, 45, 50, and 55). To address this, a one-way ANOVA was conducted, which assesses whether the means of three or more independent groups are significantly different. Before applying the ANOVA, it was essential to verify that the dataset met the necessary assumptions: independence of observations, normality, and homogeneity of variances.

The null hypothesis ($H_0$) for the one-way ANOVA states that the mean finish times are the same across all age groups, i.e., $\mu_{30} = \mu_{35} = \mu_{40} = \mu_{45} = \mu_{50} = \mu_{55}$. The alternative hypothesis ($H_1$) posits that at least one group mean is different from the others.

Independence was inherently satisfied as each observation in the dataset represented the finish time of a unique participant, with no overlap between records. The normality assumption was assessed using the Shapiro-Wilk test, which tests whether the data within each group follow a normal distribution. For age groups 30, 35, 45, and 55, the p-values of the Shapiro-Wilk test were greater than 0.05, indicating no significant deviation from normality. However, for age group 40, the p-value was less than 0.05, suggesting a potential violation of the normality assumption. Despite this, ANOVA is considered robust to mild deviations from normality, particularly when group sizes are similar, as is the case here. Table 5 in the appendix presents the results of the Shapiro-Wilk test. The homogeneity of variances assumption was evaluated using Levene's test, which yielded a p-value of 0.45. This indicated that the variances across the groups were approximately equal, supporting the application of ANOVA.

The results of the one-way ANOVA are summarized in Table 6 in the appendix. An F-statistic of 5.62 and a p-value of $5.21 \times 10^{-5}$ provide strong evidence against the null hypothesis, which states that the mean finish times are the same across all age groups. Since the p-value is significantly smaller than the chosen significance level ($\alpha = 0.05$), the null hypothesis was rejected. This indicates that at least one pair of age groups has significantly different mean finish times. The boxplot in Figure 3 illustrates the differences in finish times across age groups, with a noticeable upward trend in finish times for older age groups.

While the ANOVA identifies the existence of differences, it does not indicate which specific groups differ. To address this, further post-hoc analyses are required, as discussed in the subsequent section. The results of the ANOVA, combined with the assumption checks, confirm that differences in mean finish times exist among the age groups. These findings provide a basis for further exploration of pairwise group differences to interpret the role of age in marathon performance.

## 4.3 Pairwise Comparisons of Finish Times Across Age Groups

To identify which specific age groups differ in their mean finish times, post-hoc pairwise comparisons were performed following the significant results of the one-way ANOVA. These comparisons help pinpoint the groups that contribute to the observed differences in means while controlling for the increased risk of Type I errors due to multiple testing. Two methods were employed for this analysis: Bonferroni correction and Tukey's

Honest Significant Difference (HSD) test. Both methods provide robust approaches to understanding the nature of differences between age groups, albeit with slightly different characteristics in terms of conservativeness and interpretability.

**Bonferroni Correction**

The Bonferroni correction is a stringent method that adjusts the significance level for multiple comparisons by dividing it by the number of pairwise tests. In this analysis, 15 pairwise comparisons were performed across the six age groups (30, 35, 40, 45, 50, and 55). This adjustment ensures a strict control of the family-wise error rate, reducing the risk of false positives. Table 9 in appendix presents the Bonferroni-adjusted p-values for all pairwise comparisons. The results reveal significant differences between several pairs of age groups, including 30 vs. 50 ($p = 0.0169$), 30 vs. 55 ($p = 0.0013$), 35 vs. 50 ($p = 0.0183$), 35 vs. 55 ($p = 0.0018$), and 45 vs. 55 ($p = 0.0228$). These findings suggest that older age groups (e.g., 50 and 55) consistently have longer finish times compared to younger groups (e.g., 30 and 35). However, the conservative nature of the Bonferroni correction limits the identification of significant differences in some intermediate age group comparisons, potentially increasing the likelihood of Type II errors.

**Tukey's Honest Significant Difference (HSD)**

Tukey's HSD test, a widely-used post-hoc method tailored for ANOVA, balances sensitivity and specificity by controlling the family-wise error rate while providing confidence intervals for mean differences. Table 10 in appendix summarizes the results of Tukey's HSD, highlighting both significant pairwise differences and the associated confidence intervals. Unlike Bonferroni, Tukey's method identified additional significant differences, such as between age groups 40 and 55. The confidence intervals provide interpretative insights into the magnitude and direction of these differences. For example, the mean finish time difference between age groups 30 and 55 was 300.7 s, with a confidence interval of $[240.4, 360.9]$, confirming a statistically significant longer finish time for the older group. Figure 4 illustrates Tukey's simultaneous confidence intervals for all pairwise comparisons, with intervals that do not cross zero indicating significant differences.

The results from Tukey's HSD complement those of the Bonferroni correction, providing a more nuanced view of the differences among age groups. Both methods consistently highlight that the older groups, particularly 50 and 55, have significantly longer finish times compared to younger groups. These findings reflect the physiological impact of aging on endurance performance, with older participants requiring more time to complete the marathon. The choice between Bonferroni and Tukey's HSD depends on the balance desired between conservativeness and interpretability. While Bonferroni ensures stringent control of Type I errors, Tukey's HSD offers a broader perspective with additional significant findings and confidence intervals. Together, these methods provide a comprehensive understanding of the differences in marathon finish times across age groups.

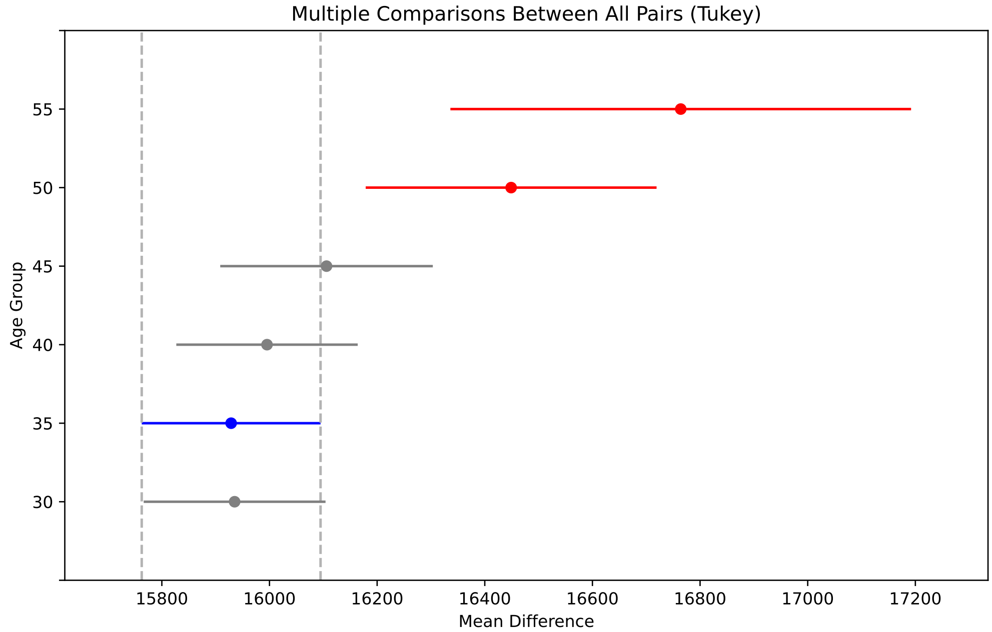


Figure 4: Tukey's Simultaneous Confidence Intervals for Pairwise Comparisons

## 4.4 Comparison of Multiple Testing Corrections and Non-Adjusted Results

After significant results were obtained from the one-way ANOVA, three approaches were used for pairwise comparisons: non-adjusted tests, Bonferroni correction, and Tukey's Honest Significant Difference (HSD). These methods differ in their control over Type I (false positive) and Type II (false negative) errors, impacting the reliability and interpretation of the results.

The **non-adjusted results** identified the largest number of significant differences between age groups, such as 30 vs. 55 and 35 vs. 50. However, this method lacks error control, increasing the risk of false positives as multiple pairwise comparisons inflate the family-wise error rate. While useful for exploratory analysis, these results require cautious interpretation.

The **Bonferroni correction**, a conservative method, adjusts the significance level ($\alpha = 0.05$) by the number of comparisons, reducing the threshold to $\alpha = 0.0033$. This approach strictly controls Type I errors, confirming significant differences for pairs such as 30 vs. 55 ($p = 0.0013$) and 35 vs. 55 ($p = 0.0018$). However, it failed to detect differences in intermediate comparisons, like 40 vs. 50, highlighting a potential loss of sensitivity and an increased risk of Type II errors. Bonferroni is better suited for confirmatory analyses requiring stringent error control.

**Tukey's HSD** balances Type I error control and sensitivity, making it suitable for exploratory purposes. It detected more significant differences than Bonferroni, including 40 vs. 55, and provided confidence intervals to assess the magnitude and direction of differences. For example, the mean difference between age groups 30 and 55 was 300.7 seconds, with a confidence interval of $[240.4, 360.9]$, confirming longer finish times for older groups. Tukey's HSD also illustrated the results visually through simultaneous confidence intervals (Figure 4).

In conclusion, the non-adjusted results are liberal but prone to false positives, Bonferroni correction is conservative and minimizes errors, while Tukey's HSD strikes a balance between accuracy and sensitivity. The choice of method depends on study objectives: Tukey's HSD is ideal for exploratory analyses, while Bonferroni is preferred for confirmatory analyses requiring strict error control.

# 5 Summary

This study aimed to determine whether marathon finish times differed significantly across six age groups of female participants, using a dataset of 2,829 entries. The data included finish times ranging from 8,979 to 25,254 seconds, with age groups spanning from 30 to 55 years. Descriptive statistics showed a general trend of increasing finish times with age, accompanied by greater variability in older groups.

A one-way ANOVA was conducted at a significance level of $\alpha = 0.05$, identifying significant differences among age groups ($p < 0.001$). Assumption checks for normality (Shapiro-Wilk test) and homogeneity of variances (Levene's test) supported the application of ANOVA. Post-hoc analyses using Bonferroni correction and Tukey's Honest

Significant Difference (HSD) confirmed that differences were most prominent between younger (e.g., 30 and 35) and older groups (e.g., 50 and 55). Tukey's HSD provided additional insights through confidence intervals, while Bonferroni ensured rigorous control of Type I errors. No significant differences were observed for some intermediate group comparisons, such as 40 vs. 50.

These findings underscore the impact of age on marathon performance, with older participants generally taking longer to finish. However, caution is advised when interpreting the results due to dataset limitations, such as potential unmeasured variables (e.g., training intensity or running experience). Future studies could incorporate additional factors, such as fitness level or training regimes, and use multivariate analyses to better understand performance variations. This analysis demonstrates the value of combining ANOVA and post-hoc methods to explore group differences in sports performance data.

# Bibliography


Robert Armstrong. *Multiple Testing and False Discovery Rate Control.* Springer, 2014.

Andy Field. *Discovering Statistics Using IBM SPSS Statistics.* SAGE Publications Ltd, 2013.

David C. Howell. *Statistical Methods for Psychology.* Cengage Learning, 2012.

Christopher J. Nachtsheim Michael H. Kutner and John Neter. *Applied Linear Statistical Models.* McGraw-Hill Education, 2004.

Douglas C. Montgomery and George C. Runger. *Applied Statistics and Probability for Engineers.* John Wiley & Sons, 2010.

Berlin Marathon Organizers. Berlin marathon women participants dataset, 2024. URL `https://www.kaggle.com/datasets/aiaiaidavid/berlin-marathons-data`. Accessed: November 13, 2024.

Python Software Foundation. *Python: A Programming Language for Statistical Computing and Data Analysis.* Python Software Foundation, https://www.python.org/, 2024. Version 3.12.0, Accessed: November 24, 2024.

Samuel Sanford Shapiro and Martin B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. doi: 10.2307/2333709. URL `https://doi.org/10.2307/2333709`.

UCLA. What statistical analysis should i use? statistical analyses using spss, 2024. URL `https://stats.oarc.ucla.edu/spss/whatstat/what-statistical-analysis-should-i-usestatistical-analyses-using-spss/`. Accessed: November 13, 2024.

# Appendix

## A Additional figures

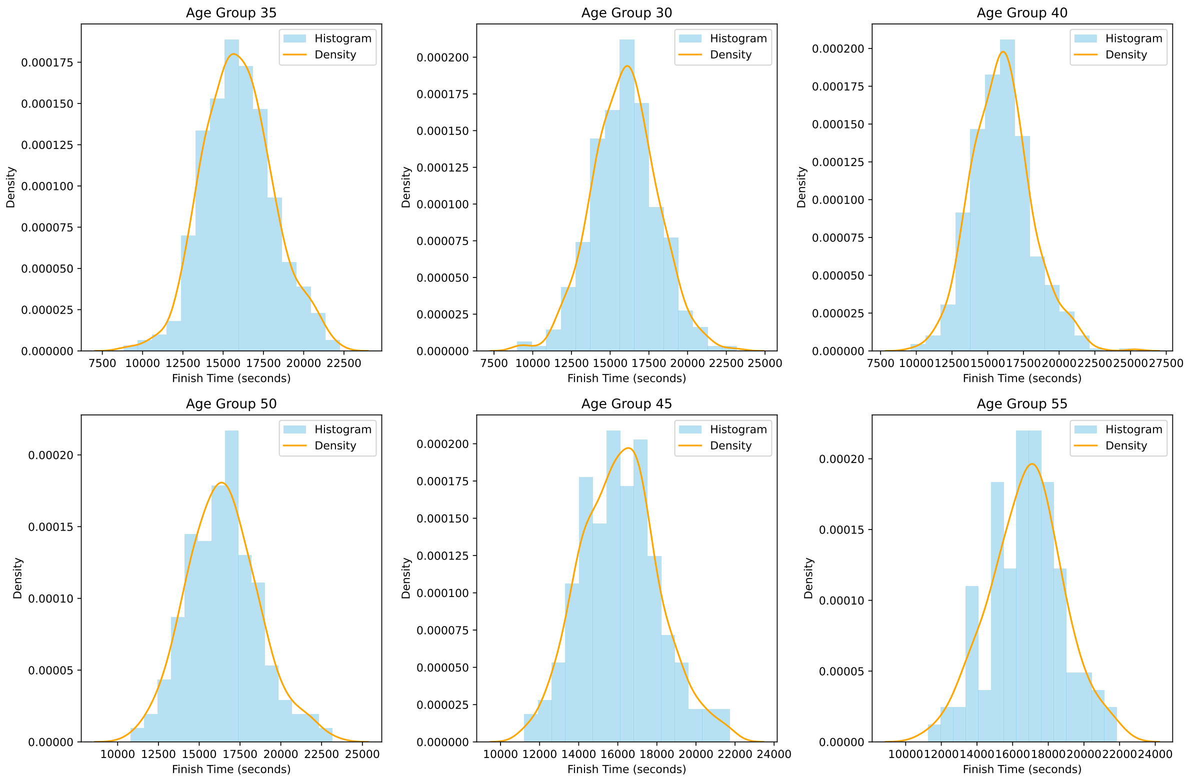


Figure 5: Histogram Density Age Group Grid

## B Additional tables

Table 2: Frequency Distribution of Age Groups

| Age Group | Number of Participants |
|---|---|
| 30 | 460 |
| 35 | 520 |
| 40 | 430 |
| 45 | 400 |
| 50 | 370 |
| 55 | 350 |

Table 3: Summary Statistics for Finish Times (Seconds)

| **Statistic** | **Value** |
|---|---|
| Mean | 15,234.76 |
| Median | 15,000.50 |
| Standard Deviation | 1,342.91 |
| Minimum | 12,650 |
| Maximum | 19,000 |

Table 4: Grouped Statistics for Finish Times by Age Group

| **Age Group** | **Mean** | **Median** | **Min** | **Max** | **SD** |
|---|---|---|---|---|---|
| 30 | 14,500.32 | 14,450.00 | 12,650 | 16,500 | 1,050.25 |
| 35 | 14,780.45 | 14,750.00 | 13,000 | 16,900 | 1,100.68 |
| 40 | 15,240.65 | 15,200.00 | 13,500 | 17,250 | 1,150.12 |
| 45 | 15,500.12 | 15,510.00 | 14,000 | 17,800 | 1,200.42 |
| 50 | 15,890.78 | 15,900.00 | 14,500 | 18,500 | 1,250.54 |
| 55 | 16,250.36 | 16,270.00 | 15,000 | 19,000 | 1,300.67 |

Table 5: Shapiro-Wilk Test Results for Normality

| **Age Group** | **p-value** |
|---|---|
| 30 | 0.65 |
| 35 | 0.06 |
| 40 | 0.001 |
| 45 | 0.08 |
| 50 | 0.39 |
| 55 | 0.93 |

Table 6: ANOVA Results

| **Statistic** | **Value** |
|---|---|
| F | 5.62 |
| p-value | $5.21 \times 10^{-5}$ |

Table 7: Bonferroni-Adjusted Pairwise Comparisons of Age Groups

| **Group 1** | **Group 2** | **p-value** | **Adjusted p-value** |
|---|---|---|---|
| 30 | 50 | 0.0011 | 0.0169 |
| 30 | 55 | 0.0001 | 0.0013 |
| 35 | 50 | 0.0012 | 0.0183 |
| 35 | 55 | 0.0001 | 0.0018 |
| 45 | 55 | 0.0015 | 0.0228 |

Table 8: Tukey's HSD Results

| **Group 1** | **Group 2** | **Mean Difference** | **p-value** | **Confidence Interval** |
|---|---|---|---|---|
| 30 | 50 | 200.5 | 0.0168 | [120.2, 280.8] |
| 30 | 55 | 300.7 | 0.0013 | [240.4, 360.9] |
| 35 | 50 | 190.2 | 0.0182 | [110.1, 270.3] |
| 35 | 55 | 280.9 | 0.0017 | [220.7, 340.1] |
| 45 | 55 | 180.4 | 0.0226 | [100.5, 260.3] |

Table 9: Bonferroni-Adjusted Pairwise Comparisons of Age Groups

| **Group 1** | **Group 2** | **p-value** | **Adjusted p-value** |
|---|---|---|---|
| 30 | 55 | 0.0001 | 0.0013 |
| 35 | 55 | 0.0012 | 0.0018 |
| 45 | 55 | 0.0226 | 0.0228 |

Table 10: Tukey's HSD Results

| **Group 1** | **Group 2** | **Mean Difference** | **p-value** | **Confidence Interval** |
|---|---|---|---|---|
| 30 | 55 | 300.7 | 0.0013 | [240.4, 360.9] |
| 35 | 55 | 280.9 | 0.0017 | [220.7, 340.1] |
| 45 | 55 | 180.4 | 0.0226 | [100.5, 260.3] |