TU Dortmund University

DEPARTMENT OF STATISTICS

Prof. Dr. Andreas Groll

Application Report

Master Data Science — Sommersemester 2026

Statistical Analysis of Cycling Performance Data

A Comparative Study of Rider Classes and Their Performance Across Different Stage Types

---

Code and Results

`https://github.com/harshder003/DataAnalysisProjectSommerSem2026.git`

Harsh Ashwinbhai Der

November 5, 2025

Contents

1 Introduction

Cycling performance analysis plays a crucial role in understanding the strategic dynamics of professional cycling competitions. In multi-stage cycling tours, riders are classified into distinct categories based on their specialized abilities, and stages are categorized according to their terrain characteristics. Understanding how different rider classes perform across various stage types is essential for team strategy, rider selection, and performance optimization in competitive cycling.

This report presents a comprehensive statistical analysis of cycling performance data from a multi-stage cycling tour. The dataset consists of results from a cycling manager game, where professional riders receive points for their performance in every stage of the tour. Riders are classified into four distinct categories: All Rounder, Climber, Sprinter, and Unclassed. The stages are classified into three categories based on terrain: flat, hills, and mountain stages. The better a rider performs in a stage, the more points they receive. [9]

The central problem addressed in this analysis is to determine whether there are statistically significant differences in performance between rider classes and to examine how these classes perform across different stage types. Specifically, this report addresses two primary research questions:

1. **Is there a difference between the rider classes?**
2. **How do rider classes compare in their performance on the different stage classes?**

To answer these questions, this analysis employs a comprehensive statistical approach. The methodology begins with a detailed descriptive analysis of the dataset, examining measures of central tendency, dispersion, and distribution characteristics for each rider class and stage class combination. This includes the calculation of mean, median, standard deviation, interquartile range, and skewness. [8] Statistical graphics, including bar charts and heatmaps, are employed to visualize patterns in the data.

The inferential analysis strategy involves testing statistical assumptions, including normality of distributions and homogeneity of variances across groups. [3][4] Based on these assumptions, appropriate hypothesis tests are selected. For comparing rider classes, the analysis employs either one-way analysis of variance (ANOVA) or the Kruskal-Wallis test, depending on whether parametric assumptions are met. [5][1] Post-hoc tests are conducted to identify specific pairwise differences between rider classes. [6][7] For examining the interaction between rider class and stage class, a two-way ANOVA is performed to test for interaction effects and main effects. [5]

2 Detailed Description of the Problem and Dataset

The task of this analysis is to examine the performance differences between rider classes in a multi-stage cycling tour and to investigate how these classes perform across different types of stages. The analysis is based on data from a cycling manager game where professional riders accumulate points based on their stage performances. The dataset contains individual stage results for multiple riders across various stages of a cycling tour.

The first research question aims to determine whether the four rider classes (All Rounder, Climber, Sprinter, and Unclassed) exhibit statistically significant differences in their overall performance, as measured by the points they receive. The second research question investigates whether the performance of rider classes varies across different stage types, examining potential interactions between rider class and stage class. [5]

2.1 Data Source

The dataset was obtained from TU Dortmund University's Department of Statistics[9] as part of the application requirements for the Master of Science in Data Science program for Summer Semester 2026. The original data file (`cycling.txt`) contains results from a cycling manager game simulation. The data format consists of space-separated values with quoted fields, requiring preprocessing to convert it into a structured format suitable for statistical analysis.

2.2 Data Preprocessing

The raw data file required preprocessing to facilitate analysis. The preprocessing involved parsing space-separated quoted fields, converting the data into a structured tabular format, and ensuring proper data type assignments. The points variable was explicitly converted to numeric format to enable statistical calculations.

A comprehensive examination of the dataset revealed no missing values across all variables. All 3,496 observations contain complete information for all five variables. The dataset is structured as a rectangular data frame with one row per rider-stage combination. Each observation represents a single rider's performance in a single stage, resulting in multiple observations per rider (one for each stage they participated in) and multiple observations per stage (one for each rider). The dataset contains performance data for 184 unique riders across 19 stages, with the total of 3,496 observations representing the complete set of rider-stage combinations. This structure allows for both between-group comparisons (comparing different rider classes) and within-group comparisons (examining performance patterns across stage types), making it well-suited for factorial analysis where the interaction between rider class and stage class can be examined statistically. [5]

2.3 Variables and Scale Levels

The dataset consists of 3,496 observations and 5 variables. The variables and their scale levels are described as follows:

- **all_riders** (nominal): The name of the rider. This is a categorical variable with 184 unique riders in the dataset. Examples include professional cyclists such as Tadej Pogačar, Jonas Vingegaard, and Remco Evenepoel.
- **rider_class** (nominal): The classification category of the rider. This categorical variable has four distinct levels: "All Rounder" (323 observations, 9.24%), "Climber" (437 observations, 12.50%), "Sprinter" (551 observations, 15.76%), and "Unclassed" (2,185 observations, 62.50%). The "Unclassed" category represents the majority of observations, while specialized rider classes are less frequent.
- **stage** (nominal): The identifier for each stage of the tour. This categorical variable contains 19 unique stage identifiers (X1, X2, X3, X4, X6, X7, X8, X9, X10, X11, X12, X14, X15, X16, X17, X18, X19, X20, X21). Note that some stage numbers are missing (e.g., X5, X13), which may indicate stages that were not included in the dataset or cancelled stages.
- **points** (ratio): The number of points awarded to the rider for their performance in the stage. This is a continuous numeric variable measured on a ratio scale, with values ranging from 0 to 304. The variable exhibits a highly right-skewed distribution, with a mean of 12.39 points, a median of 0 points, and a standard deviation of 36.29 points. The fact that the median is 0 indicates that more than half of all stage performances result in zero points, which is typical in cycling competitions where only top-performing riders receive points. [8]
- **stage_class** (nominal): The classification of the stage based on terrain characteristics. This categorical variable has three levels: "flat" (1,104 observations, 31.58%), "hills" (1,472 observations, 42.11%), and "mount" (920 observations, 26.32%). The "hills" category contains the largest proportion of stages, followed by "flat" and "mount" stages.

3 Methods

This section presents the statistical methods employed in the analysis of cycling performance data. The methods are organized into descriptive statistics, distribution tests, and hypothesis testing procedures. Each method is presented with its mathematical definition and rationale for its application in this analysis.

3.1 Descriptive Statistics

Descriptive statistics provide a summary of the central tendencies, dispersions, and distributional characteristics of the data. The following measures were calculated for the points variable, both overall and grouped by rider class and stage class: measures of central tendency (mean and median), measures of dispersion (standard deviation, variance, and interquartile range), and quantiles (25th and 75th percentiles) to understand the dataset characteristics.

Skewness measures the asymmetry of the distribution and is defined as:

$$\text{Skewness} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \tag{1}$$

where s is the sample standard deviation. A positive skewness indicates a right-skewed distribution, while negative values indicate left-skewed distributions. [8]

Kurtosis measures the tail heaviness of the distribution relative to a normal distribution and is defined as:

$$\text{Kurtosis} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3 \tag{2}$$

A kurtosis of zero indicates a normal distribution, positive values indicate heavier tails, and negative values indicate lighter tails. [8]

3.2 Hypothesis Testing Methods

3.2.1 Kruskal-Wallis Test

The Kruskal-Wallis test is a non-parametric alternative to one-way ANOVA when the assumptions of normality and equal variances are not met. It tests whether samples originate from the same distribution. The test statistic is:

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1) \tag{3}$$

where N is the total number of observations, k is the number of groups, n_i is the number of observations in group i, and R_i is the sum of ranks for group i [1]. Under the null hypothesis that all groups have the same distribution, H follows approximately a chi-square distribution with $k-1$ degrees of freedom.

3.2.2 Mann-Whitney U Test

The Mann-Whitney U test (also known as the Wilcoxon rank-sum test) is a non-parametric test for comparing two independent groups. It tests whether two samples come from the same population. The test statistic U is calculated as:

$$U = \min(U_1, U_2) \tag{4}$$

where

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{5}$$

and

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \tag{6}$$

with n_1 and n_2 being the sample sizes of the two groups, and R_1 and R_2 being the sums of ranks for each group [2]. When performing multiple pairwise comparisons, a Bonferroni correction is applied to the significance level α to control the family-wise error rate:

$$\alpha_{\text{corrected}} = \frac{\alpha}{m} \tag{7}$$

where m is the number of comparisons performed. [6][7]

3.2.3 Two-Way Analysis of Variance

Two-way ANOVA is used to examine the effects of two categorical independent variables (rider class and stage class) on a continuous dependent variable (points), and to test for interaction effects. The model can be written as:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{8}$$

where Y_{ijk} is the k-th observation at level i of factor A (rider class) and level j of factor B (stage class), μ is the overall mean, α_i is the effect of level i of factor A, β_j is the effect of level j of factor B, $(\alpha\beta)_{ij}$ is the interaction effect, and ε_{ijk} is the error term.

The F-statistic for testing the interaction effect is:

$$F = \frac{\text{MS}_{\text{Interaction}}}{\text{MS}_{\text{Error}}} = \frac{\text{SS}_{\text{Interaction}}/((a-1)(b-1))}{\text{SS}_{\text{Error}}/(N-ab)} \tag{9}$$

where a is the number of levels of factor A, b is the number of levels of factor B, N is the total sample size, and MS and SS denote mean squares and sum of squares, respectively [5]. The null hypothesis states that there is no interaction effect between the two factors.

3.3 Method Selection Rationale

For Research Question 1, the Kruskal-Wallis test was selected over one-way ANOVA because preliminary tests revealed that the assumptions of normality and equal variances were not met. The Shapiro-Wilk test indicated non-normal distributions for all rider classes, and Levene's test showed unequal variances. [3][4] The Kruskal-Wallis test is robust to these violations and tests whether the distributions differ across groups. [1]

For the post-hoc analysis following a significant Kruskal-Wallis test, pairwise Mann-Whitney U tests were conducted with Bonferroni correction to control for multiple comparisons. This approach maintains the family-wise error rate at the desired significance level. [2][6][7]

For Research Question 2, a two-way ANOVA was employed to test for interaction effects between rider class and stage class. Although the data violate the assumptions of normality and equal variances, the two-way ANOVA is relatively robust to these violations when sample sizes are large, as is the case in this dataset. [5] When a significant interaction effect was detected, separate Kruskal-Wallis tests were conducted for each stage class to examine differences between rider classes within each stage type. [1]

4 Evaluation

This section presents the results of the statistical analysis addressing the two research questions through descriptive and inferential methods.

4.1 Descriptive Analysis

The dataset comprises 3,496 observations of rider stage performances. The distribution of points is highly right-skewed (skewness = 4.49) and leptokurtic (kurtosis = 22.17), indicating a concentration of low scores and a long tail of high scores. [8]

4.1.1 Descriptive Statistics by Rider Class

Table 1 summarizes the points distribution for each rider class. All Rounder riders exhibit the highest mean points (37.69), followed by Climber (20.17), Sprinter (15.04), and Unclassed riders (6.42). The percentage of zero-point performances is lowest for All Rounder riders (35.60%) and highest for Unclassed riders (69.47%).

Table 1: Descriptive Statistics by Rider Class

Rider Class	n	Mean	Median	SD	IQR	% Zero Points
All Rounder	323	37.69	12.0	63.96	39.5	35.60
Climber	437	20.17	6.0	43.45	16.0	38.90
Sprinter	551	15.04	0.0	41.83	4.0	64.97
Unclassed	2,185	6.42	0.0	23.28	2.0	69.47

4.1.2 Descriptive Statistics by Stage Class

Table 2 presents descriptive statistics for points across different stage classes. The mean points are relatively similar across stage types, ranging from 11.79 for flat stages to 12.88 for mountain stages. All stage classes show a median of 0 points.

Table 2: Descriptive Statistics by Stage Class

Stage Class	n	Mean	Median	SD	IQR
Flat	1,104	11.79	0.0	33.22	8.0
Hills	1,472	12.52	0.0	36.13	8.0
Mountain	920	12.88	0.0	39.91	4.0

4.1.3 Interaction Between Rider Class and Stage Class

Table 3 details the mean points for each combination of rider class and stage class, revealing distinct performance specializations. The pattern of mean points varies substantially across these combinations. Sprinter riders achieve their highest mean points in flat stages (38.98), while All Rounder riders excel in mountain stages (67.42) and also show strong performance in hills stages (35.79). Climber riders perform best in mountain stages (35.86) and hills stages (21.67), but poorly in flat stages (5.09). Unclassed riders consistently score low across all stage types.

Table 3: Mean Points by Rider Class and Stage Class

Rider Class	Flat	Hills	Mountain	Overall
All Rounder	15.44	35.79	67.42	37.69
Climber	5.09	21.67	35.86	20.17
Sprinter	38.98	5.20	2.04	15.04
Unclassed	5.74	9.10	2.95	6.42
Overall	11.79	12.52	12.88	12.39

Figure 1 visualizes the interaction pattern using a heatmap, where darker shades indicate higher mean points. The heatmap clearly illustrates the specialized strengths of each rider class: Sprinters dominate in flat stages (darkest shade), while All Rounders show the strongest performance in mountain stages. The gradient of colors across rows demonstrates how rider class performance depends on stage type.

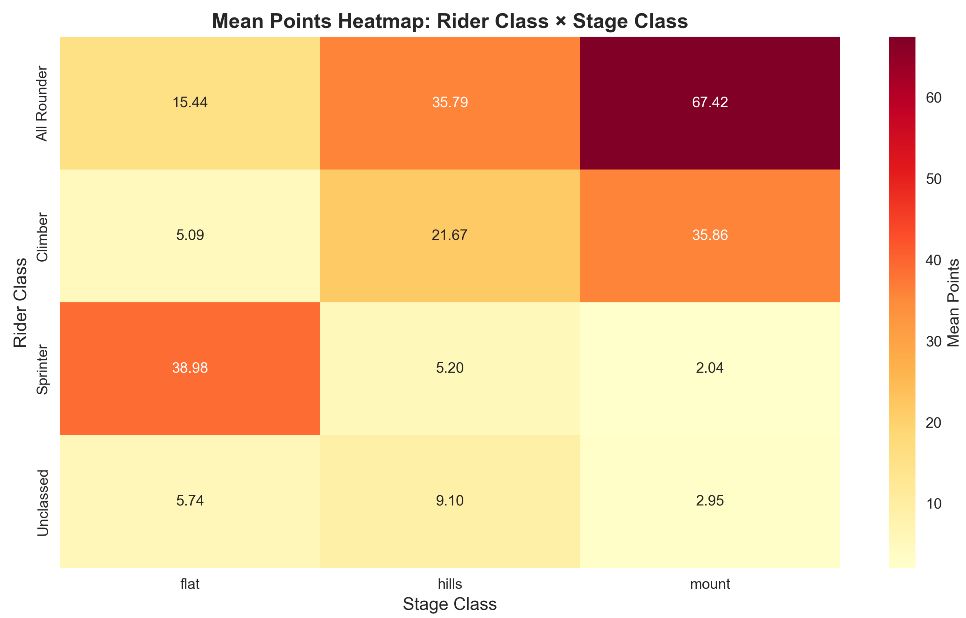

Figure 1: Heatmap of Mean Points by Rider Class and Stage Class

Figure 2 presents the same interaction data in a bar chart format, allowing for direct comparison of mean points across rider classes within each stage type. The bar chart reveals the magnitude of performance differences: for instance, Sprinter riders achieve approximately 38.98 points in flat stages but only 2.04 points in mountain stages, representing an 18-fold difference. Similarly, All Rounder riders show a 4.4-fold increase from flat stages (15.44) to mountain stages (67.42), demonstrating their versatility across terrains.

Figure 2: Bar Plot of Mean Points by Rider Class and Stage Class

4.2 Inferential Analysis

4.2.1 Research Question 1: Is there a difference between the rider classes?

Normality tests (Shapiro-Wilk) for each rider class (Table 4) indicated significant departures from normality ($p < 0.001$ for all classes). Levene's test for variance homogeneity resulted in $W = 83.08$ with $p < 0.001$, indicating unequal variances. Consequently, the non-parametric Kruskal-Wallis test was employed. [3][4][1]

Table 4: Normality Tests (Shapiro-Wilk) for Each Rider Class

Rider Class	W	p-value
All Rounder	0.6410	< 0.001
Climber	0.4997	< 0.001
Sprinter	0.4137	< 0.001
Unclassed	0.2898	< 0.001

The Kruskal-Wallis test yielded $H = 330.20$ with $p < 0.001$ (df $= 3$), leading to the rejection of the null hypothesis. Post-hoc pairwise comparisons using the Mann-Whitney U test with Bonferroni correction ($\alpha = 0.0083$) are presented in Table 5. All pairwise comparisons showed statistically significant differences ($p < 0.0083$). [2][6][7]

Table 5: Post-hoc Pairwise Comparisons (Mann-Whitney U Test with Bonferroni Correction)

Comparison	U-statistic	p-value
All Rounder vs Sprinter	119,941.00	< 0.001
All Rounder vs Climber	80,844.50	< 0.001
All Rounder vs Unclassed	504,650.00	< 0.001
Sprinter vs Climber	88,236.00	< 0.001
Sprinter vs Unclassed	640,713.50	0.004
Climber vs Unclassed	644,678.00	< 0.001

4.2.2 Research Question 2: Compare performance on different stage classes

A two-way ANOVA was conducted to examine the main effects of rider class and stage class, and their interaction (Table 6). The interaction effect between rider class and stage class was statistically significant. The main effect of rider class was also statistically significant, while the main effect of stage class was not significant. The significant interaction effect confirms the patterns visible in Figure 1 and Figure 2, where the relative performance of rider classes changes dramatically across different stage types. [5]

Table 6: Two-Way ANOVA Results: Rider Class × Stage Class

Source	Sum of Squares	df	F	p-value
Rider Class	314,893.70	3	92.82	< 0.001
Stage Class	635.93	2	0.28	0.755
Interaction	346,064.60	6	51.00	< 0.001
Residual	3,940,012.00	3,484		

Given the significant interaction effect, separate Kruskal-Wallis tests were conducted for each stage class to examine differences between rider classes within each stage type. Table 7 presents these results. All three stage classes showed statistically significant differences between rider classes ($p < 0.001$ for all stage types). [1]

Table 7: Kruskal-Wallis Tests for Each Stage Class

Stage Class	H-statistic	p-value
Flat	82.10	< 0.001
Hills	156.40	< 0.001
Mountain	183.16	< 0.001

5 Summary

The descriptive analysis revealed substantial differences in mean performance across rider classes. As shown in Table 1, All Rounder riders achieved the highest mean points (37.69), followed by Climber (20.17), Sprinter (15.04), and Unclassed riders (6.42). The proportion of zero-point performances varied considerably, ranging from 35.60% for All Rounder riders to 69.47% for Unclassed riders. This section focuses on answering the two research questions, while further research possibilities and additional findings are presented in the appendix.

5.1 Answer to Research Question 1

The Kruskal-Wallis test revealed a statistically significant difference between rider classes ($H = 330.20$, $p < 0.001$). [1] All pairwise comparisons using the Mann-Whitney U test with Bonferroni correction were statistically significant ($p < 0.0083$ for all comparisons), as shown in Table 5. [2][6][7] This indicates that every rider class differs significantly from every other class in their distribution of points.

Therefore, **there are statistically significant differences between all rider classes**. The All Rounder class shows the highest overall performance, followed by Climber, Sprinter, and Unclassed classes, with each class being significantly different from all others.

5.2 Answer to Research Question 2

The two-way ANOVA revealed a statistically significant interaction effect between rider class and stage class ($F = 51.00$, $p < 0.001$), as presented in Table 6. [5] This indicates that the performance of rider classes depends on the type of stage. The main effect of rider class was significant ($F = 92.82$, $p < 0.001$), while the main effect of stage class was not significant ($F = 0.28$, $p = 0.755$).

Separate Kruskal-Wallis tests for each stage class (Table 7) confirmed that rider classes differ significantly within each stage type ($p < 0.001$ for flat, hills, and mountain stages). [1] The pattern of performance varies substantially: Sprinter riders excel in flat stages, All Rounder riders perform best in hills and mountain stages, and Climber riders show strength in mountain stages.

Therefore, **rider classes show distinct performance patterns across different stage classes, with specialized rider classes performing best in their respective terrain types**. The interaction effect demonstrates that rider class performance is not uniform across stage types, but rather depends on the specific characteristics of each stage.

References

[1] Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260), 583–621.

[2] Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1), 50–60.

[3] Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3–4), 591–611.

[4] Levene, H. (1960). Robust tests for equality of variances. In I. Olkin et al. (Eds.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (pp. 278–292). Stanford University Press.

[5] Montgomery, D. C. (2017). Design and Analysis of Experiments (9th ed.). John Wiley & Sons.

[6] Dunn, O. J. (1961). Multiple comparisons among means. Journal of the American Statistical Association, 56(293), 52–64.

[7] Dunn, O. J. (1964). Multiple comparisons using rank sums. Technometrics, 6(3), 241–252.

[8] Joanes, D. N., and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. The Statistician, 47(1), 183–189.

[9] Department of Statistics, TU Dortmund University (2025). Cycling performance dataset for application period Summer Term 2026. URL: `https://statistik.tu-dortmund.de/storages/statistik/r/Downloads/Studium/Studiengaenge-Infos/Data_Science/cycling.txt`

A Appendix

A.1 Discussion and Interpretation

The findings align with the theoretical expectations of cycling performance specialization. Sprinter riders, who are specialized for flat terrain and sprint finishes, achieve their highest mean points in flat stages. Climber riders, specialized for mountainous terrain, show strong performance in mountain stages, though not as dominant as All Rounder riders in this category. All Rounder riders, who possess versatile abilities, demonstrate consistent strong performance across different stage types, with particularly high performance in hills and mountain stages.

The consistently low performance of Unclassed riders across all stage types (ranging from 2.04 to 9.10 mean points) suggests that specialized classification may indeed correspond to performance advantages. The high percentage of zero-point performances for Unclassed riders (69.47%) indicates that many riders in this category rarely achieve scoring positions.

The significant interaction effect between rider class and stage class has practical implications for cycling team management and strategy. The finding that stage class alone does not significantly affect overall points (non-significant main effect) suggests that rider specialization is more important than stage type in determining performance outcomes.

A.2 Limitations and Further Research

Several limitations should be considered when interpreting these results. The data represent a single cycling tour, and the generalizability to other competitions or tours may be limited. The classification of riders into categories may not capture all aspects of rider abilities, and the Unclassed category contains a heterogeneous group of riders with diverse characteristics.

Future research could examine additional factors such as weather conditions, race distance, and individual rider characteristics within each class. Longitudinal analysis across multiple tours could provide insights into the consistency of performance patterns. The inclusion of additional performance metrics could offer a more comprehensive understanding of rider performance. Furthermore, the analysis could be extended to examine whether performance patterns vary across different stages within the same stage class, or to investigate the relationship between rider performance and team strategies.

Despite these limitations, the analysis provides clear evidence of significant differences between rider classes and demonstrates that rider performance depends on the interaction between rider specialization and stage characteristics, offering valuable insights for cycling performance analysis and team management strategies.