Comparison of Rider and Stage Classes in a Cycling Manager Dataset

Application Report – Master Data Science, Summer Semester 2026

Tallal Eatazaz

TU Dortmund University, Department of Statistics

November 2025

Contents

1 Introduction

Professional road cycling divides riders into roles such as all rounders, climbers and sprinters. Each role suits a different type of stage. For example, sprinters tend to do well on flat stages, while climbers do well in the mountains. Team managers must match riders to stages and plan their tactics around these strengths.

This report analyses data from a cycling manager game. The data contain stage results for many riders across several stage types. For each rider and stage, the dataset records the rider class, the stage class and the points gained in that stage. Points measure the success of a rider in a stage and serve as the performance variable in this analysis.

The main goal is to compare the performance of rider classes across stage classes. The report addresses the following research questions:

- RQ1: Do rider classes differ in their average points?
- RQ2: Do stage classes differ in average points?
- RQ3: Does the effect of rider class on points depend on the stage class?

To answer these questions, the report uses descriptive statistics and a two-way analysis of variance (ANOVA) with interaction. The descriptive part summarises points for rider and stage classes and uses boxplots to show the distribution of points. The inferential part fits a two-way ANOVA model and tests for main and interaction effects. It also checks standard model assumptions using residual plots and a normality test.

The rest of the report is organised as follows. Section 2 describes the problem in more detail and gives an overview of the dataset. Section 3 explains the statistical methods used in the analysis. Section 4 presents the descriptive results, the ANOVA results and the model checks. Section 5 summarises the main findings and discusses limitations and possible extensions.

2 Problem and Data Description

The dataset comes from a simulated cycling tour used in a manager-style game. Each observation corresponds to a combination of one rider and one stage. The goal is to understand how rider classes perform across different stage types, using the points assigned by the game as a performance measure.

The analysis uses the following variables:

- `all_riders`: rider name (categorical, identifier).
- `rider_class`: rider class with four levels: All Rounder, Climber, Sprinter and Unclassed (categorical, explanatory).

- `stage`: stage identifier (categorical, not used as a predictor).
- `points`: points gained by the rider in that stage (numeric, response).
- `stage_class`: stage class with three levels: flat, hills and mount (categorical, explanatory).

There are 3,496 observations and 5 variables. The response variable `points` has a strong spike at zero and a long right tail. The sample mean is about 12.39 points, with a standard deviation of about 36.29 points. The median is 0 points, and the maximum observed value is 304 points. Many rider–stage combinations therefore earn no points, while a few combinations earn very high scores.

Four rider classes appear in the data. All Rounders, Climbers and Sprinters represent the typical specialist roles in stage races. The Unclassed group collects riders who do not fall into the three main roles. The three stage classes reflect different terrain: flat stages, hilly stages and mountain stages.

There are no missing values in the variables used for the analysis. All 3,496 observations are kept. The focus of the analysis is on how the expected points depend on the rider class, the stage class and their interaction.

3 Methods

This section describes the statistical methods used in the analysis. Section 3.1 introduces the descriptive statistics and graphical tools. Section 3.2 presents the two-way ANOVA model and the associated hypothesis tests.

3.1 Descriptive statistics

Let $x_1, \ldots, x_n$ denote a sample of observed values of the points variable in a given group. The sample mean is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

which measures the average performance in that group. The sample variance is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

and the sample standard deviation is $s = \sqrt{s^2}$. The variance and standard deviation describe variability around the mean.

To describe the distribution of points in more detail, the analysis also uses quantiles. The empirical p-quantile q_p is a value such that a proportion p of the observations is less than or equal to q_p. In particular, the median is the 0.5-quantile, and the first and third quartiles are the 0.25- and 0.75-quantiles.

Boxplots display the median, the quartiles and the overall spread of the data in each group. In the standard form used here, the box shows the interquartile range, the line in the box marks the median, and the whiskers extend to the most extreme points that are not classified as outliers. Individual points beyond the whiskers are shown as separate markers. Boxplots are useful to compare the central tendency and spread of points between rider classes and between stage classes.

The descriptive part of the analysis reports the sample size, mean, standard deviation, median and quartiles of points by rider class and by stage class. It also computes the mean points for each combination of rider class and stage class. These summaries give a first view of how performance differs across groups.

3.2 Two-way analysis of variance

To address the research questions in a formal way, the analysis uses a two-way ANOVA model with interaction. The response variable is the number of points in a rider–stage combination.

Let Y_{ijk} denote the points for the k-th observation in rider class i and stage class j. The model assumes

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

where

- μ is the overall mean level of points,
- α_i is the effect of the i-th rider class ($i = 1, \dots, 4$),
- β_j is the effect of the j-th stage class ($j = 1, \dots, 3$),
- $(\alpha\beta)_{ij}$ is the interaction effect between rider class i and stage class j,
- ε_{ijk} is a random error term.

The error terms are assumed to be independent and identically distributed with mean zero and variance σ^2. The model implies that the expected points depend on the rider class, the stage class and their interaction.

The following null hypotheses are of interest:

- H_0^A: there is no rider class effect, that is $\alpha_1 = \cdots = \alpha_4 = 0$.
- H_0^B: there is no stage class effect, that is $\beta_1 = \cdots = \beta_3 = 0$.
- H_0^{AB}: there is no interaction effect, that is $(\alpha\beta)_{ij} = 0$ for all i, j.

The ANOVA splits the total sum of squares of Y into components due to rider class, stage class, interaction and residual variation. For each effect, it computes a mean square by dividing the sum of squares by the corresponding degrees of freedom. The test statistic for an effect is an F-ratio of the form

$$F = \frac{\mathrm{MS}_{\text{effect}}}{\mathrm{MS}_{\text{residual}}},$$

where $\mathrm{MS}_{\text{residual}}$ is the mean square of the residuals. Under the respective null hypothesis, this statistic follows an F distribution with degrees of freedom given by the effect and the residual term. Large values of F provide evidence against the null hypothesis.

The model assumptions are checked by inspecting residual plots and a normal Q–Q plot. A residuals-versus-fitted plot is used to look for non-constant variance and systematic patterns. A Q–Q plot compares the empirical distribution of residuals to a normal distribution. In addition, a Shapiro–Wilk test is applied to the residuals to assess normality in a formal way.

The model is fitted using the `statsmodels` package in Python. The ANOVA table is computed with type-II sums of squares.

4 Evaluation

4.1 Descriptive analysis

Table 1 summarises the distribution of points by rider class. All Rounders and Climbers have higher mean points than Sprinters and Unclassed riders. However, the standard deviations are also large, and all groups show many zero scores.

Table 1: Summary of points by rider class.

Rider class	Count	Mean	Std. dev.	Min	Q1	Median	Q3
All Rounder	323	37.69	63.96	0	0	12	39.5
Climber	437	20.17	43.45	0	0	6	16
Sprinter	551	15.04	41.83	0	0	0	4
Unclassed	2185	6.42	23.28	0	0	0	2

Table 2 shows the same summary by stage class. Mean points are similar for flat, hilly and mountain stages. All three stage types have a median of 0 points and a wide range of positive scores.

Table 2: Summary of points by stage class.

Stage class	Count	Mean	Std. dev.	Min	Q1	Median	Q3
flat	1104	11.79	33.22	0	0	0	8
hills	1472	12.52	36.13	0	0	0	8
mount	920	12.88	39.91	0	0	0	4

To understand how rider and stage classes interact, Table 3 reports the mean points for each combination. All Rounders have the highest average points on mountain stages, and meaningful averages on flat and hilly stages as well. Climbers gain most of their points on hilly and mountain stages. Sprinters score high on flat stages but very little on hills and mountains. Unclassed riders score much less on all stage types.

Table 3: Mean points by rider and stage class.

Rider class	flat	hills	mount
All Rounder	15.44	35.79	67.42
Climber	5.09	21.67	35.86
Sprinter	38.98	5.20	2.04
Unclassed	5.74	9.10	2.95

Figure 1 shows a boxplot of points by rider class. All Rounders have the highest median and spread, followed by Climbers and Sprinters. Unclassed riders have a median of zero and a lower spread. Each group has many zero scores and some very large outliers.

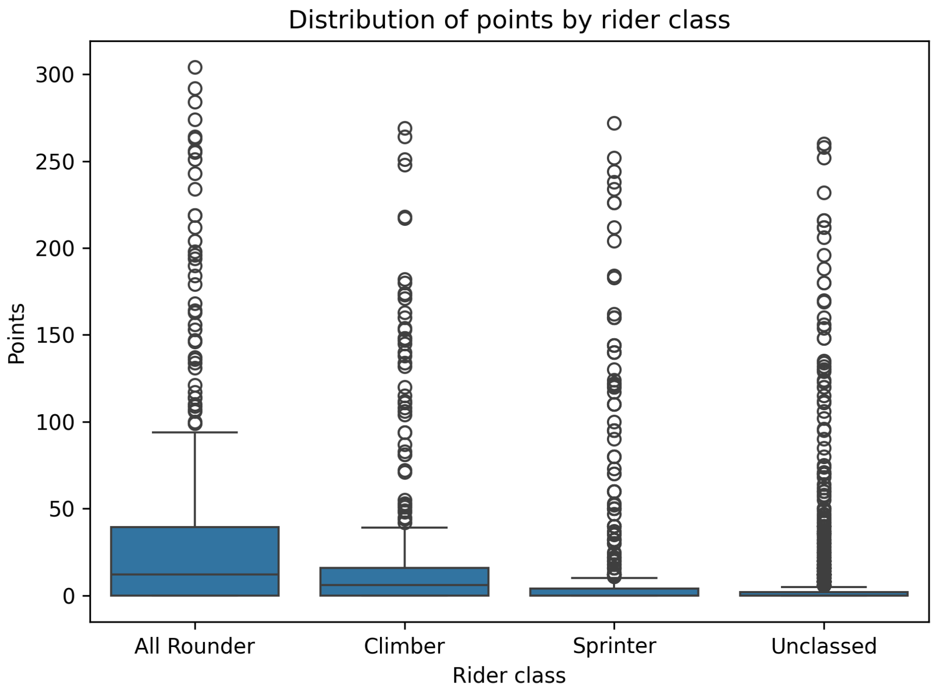

Figure 1: Distribution of points by rider class.

Figure 2 shows a boxplot of points by stage class. The three stage types look similar at this level of aggregation. This agrees with the similarity in mean points seen in Table 2.

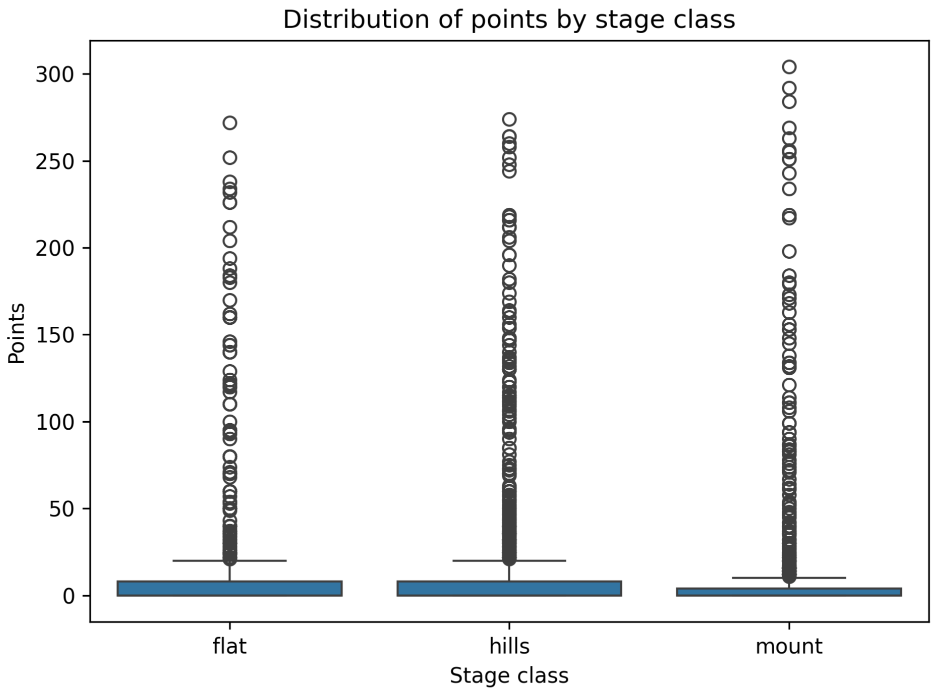

Figure 2: Distribution of points by stage class.

Overall, the descriptive analysis suggests large differences between rider classes, but not between stage classes when averaged over riders. The mean table by rider and stage classes hints at a strong interaction: the best stage type depends on the rider class.

4.2 Hypothesis tests

The two-way ANOVA model described in Section 3.2 was fitted using `statsmodels`. Table 4 reports the ANOVA table with type-II sums of squares.

Table 4: Two-way ANOVA for points with rider class and stage class.

Source	Sum Sq	df	F	p-value
Rider class	314,893.7	3	92.82	8.7×10^{-58}
Stage class	635.9	2	0.28	0.75
Rider class × stage class	346,064.6	6	51.00	2.0×10^{-60}
Residual	3,940,012.0	3484		

The rider class effect is highly significant ($F \approx 92.8$, $p < 10^{-50}$). This rejects the null hypothesis of equal mean points across rider classes. The stage class main effect is not significant ($F \approx 0.28$, $p \approx 0.75$). This means that once rider class and interaction are taken into account, the average points do not differ between stage types on their own. The interaction between rider class and stage class is strongly significant ($F \approx 51.0$, $p < 10^{-50}$). This shows that the effect of rider class on points depends on the stage class.

The interaction pattern matches the means in Table 3. Sprinters perform very well on flat stages but not on hills or mountains. Climbers and All Rounders perform better on hilly and mountain stages. Unclassed riders perform worse in all stage types.

4.3 Model checks

Model diagnostics focus on the residuals from the ANOVA model. Figure 3 shows the residuals versus the fitted values. The residuals form vertical bands because each combination of rider class and stage class has a single fitted mean. The spread of residuals is large and roughly similar across fitted values, with some very large positive outliers. There is no clear funnel shape or strong systematic pattern, but the variance is not constant in a strict sense.

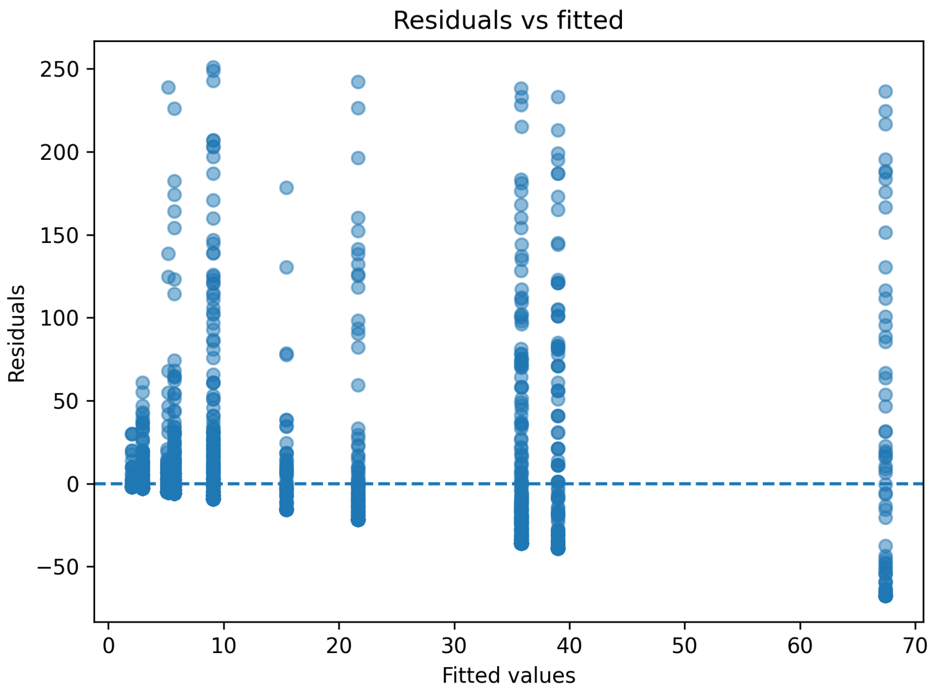

Figure 3: Residuals versus fitted values for the ANOVA model.

Figure 4 presents a normal Q–Q plot of the residuals. The residuals deviate strongly from the reference line. There is a heavy concentration around zero and a long tail of large positive residuals. This reflects the many zero scores and a few very high scores in the original data.

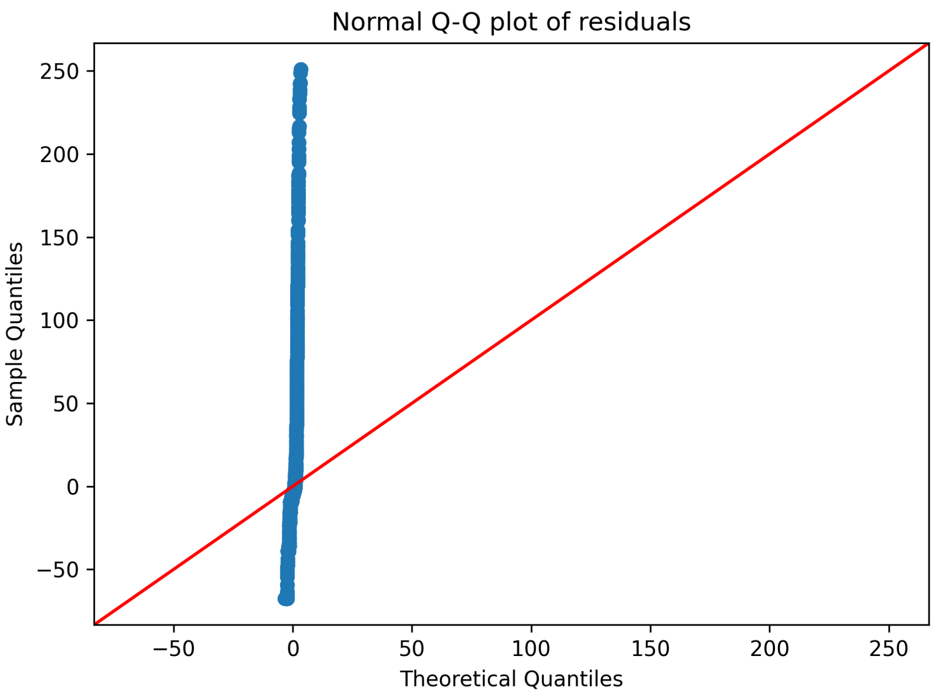

Figure 4: Normal Q–Q plot of residuals.

A Shapiro–Wilk test for normality gives a test statistic of about 0.56 and a p-value of about 2.2×10^{-69}. This strongly rejects the assumption of normal residuals. Given the large sample size, small p-values are expected for even moderate deviations from normality.

In summary, the residual analysis shows that the data deviate from the ideal ANOVA assumptions. The residuals are not normal and the variance is not constant. However, the sample size is large and the group sizes are substantial. In such settings, ANOVA F-tests are known to be robust to non-normality and mild heteroscedasticity. The very large F-values for rider class and the interaction suggest that the main conclusions about these effects are stable despite the imperfect model fit.

5 Summary and Discussion

This report studied the performance of different rider classes across stage classes in a cycling manager dataset. The response variable was the number of points a rider gained in a given stage. The main questions were whether rider classes differ in average points, whether stage classes differ, and whether the effect of rider class depends on the stage class.

The descriptive analysis showed clear differences between rider classes. All Rounders had the highest mean points, followed by Climbers and Sprinters, while Unclassed riders had much lower scores. When averaged over riders, the three stage classes had similar mean points and very similar distributions, with many zero scores and some large positive values. A table of mean points by rider and stage class revealed a strong interaction: Sprinters performed very well on flat stages but not on hills or mountains, while Climbers and All Rounders performed better on hilly and mountain stages.

The two-way ANOVA confirmed these impressions. Rider class had a strong and highly significant effect on points, while the main effect of stage class was not significant. The interaction between rider class and stage class was also highly significant. This means that the ranking of rider classes changes across stage types. Sprinters dominate on flat stages, Climbers and All Rounders do better in hills and mountains, and Unclassed riders lag behind in all conditions.

Model checks showed that the residuals deviate from normality and that the variance is not constant. These features arise because many rider–stage combinations earn no points while a few score very highly. Despite these issues, the large sample size and the strong effect sizes make the ANOVA conclusions robust in practice. Still, the violations of the assumptions should be kept in mind when interpreting the results.

There are several ways to extend this analysis. One option is to use models that handle count data or zero inflation, such as generalized linear models with a suitable link function. Another option is to model riders as random effects to capture repeated measurements of the same rider across stages. It would also be interesting to include further covariates, such as stage length or weather conditions, if available.

Overall, the analysis supports a simple but important conclusion: rider role and terrain interact in a clear way. Different rider classes are strong on different types of stages, and this pattern shapes the distribution of points in the tour.

Bibliography

References

[1] Montgomery, D. C., & Runger, G. C. (2014). *Applied Statistics and Probability for Engineers* (6th ed.). Wiley.

[2] Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference.* Springer.

[3] Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference.*