Re-evaluating the Qualifying/Finish Relationship in Formula One: A Replication and Correction of Prior Findings

Joshua Weissbock
Carleton University
joshweissbock@cmail.carleton.ca

Dr. Shirley Mills
Carleton University
shirleymills@cunet.carleton.ca

October 31, 2025

Abstract

Prior academic research in Formula One, most notably Mühlbauer (2010), concluded that starting grid positions were the strongest determinants of race outcomes while only examining the top eight finishers (fewer than 40%) of the competitors over a shortened sample of 4 seasons (2006–2009). This truncated sample approach limited the generalizability of its findings and likely affected the observed relationships. This study attempts to correct these findings by using a much more comprehensive dataset of over 7,800 driver-weekend pairings, spanning nearly two decades of sporting activity and multiple technical eras, while including the entire racing field. We extended the original analysis, by applying contingency coefficients, as well as rank-order correlations and ordinal logistic regression, to quantify the strength of association between sessions in a race weekend (practices, qualification, starting grid, and finishing position). This extended work indicates that qualifying performance, rather than the starting grid (often altered by post-qualification adjustments due to penalties), exhibits the strongest and most consistent association with outcomes. By re-evaluating this start-finish relationship with more complete data and transparent methods, this study corrects earlier misinterpretations and reinforces that qualifying is the most accurate indicator of the underlying driver and car performance across a weekend, season, and regulatory eras.

Keywords: Formula One; association; contingency coefficient; ordinal logistic association; replication; correction

1 Introduction

1.1 Background

Formula One (F1) is the highest class of international auto racing within the category of open-wheel (exposed wheels), single-seat (one occupant) formula racing cars sanctioned by the Fédération Internationale de l'Automobile (FIA) (Fédération Internationale de l'Automobile (FIA), 2024). As the pinnacle of motorsport, F1 incorporates the latest advancements in all areas of science and engineering. Governed by the FIA, F1 runs 20 to 25 Grand Prix races over a calendar year, leading up to both the Driver's and Constructor's World Championships. Each race weekend generally follows a structured format of 2 practice sessions on a Friday (Practice 1 (P1) and Practice 2 (P2)) to allow teams to collect data and refine their technical setups; a third practice session (Practice 3 (P3)), a qualifying session on Saturday (used to determine the fastest to slowest cars to set the starting grid in a single-lap time trial), and a Grand Prix on Sunday, a race of just over 300 km.

F1 is a sport that leads in cutting-edge technology, from engine development to aerodynamics to the employment of various sensors, all of which make data essential for competitiveness. Teams analyze all sorts of race data, including telemetry, tyre degradation, weather, altitude, sector and mini-sector execution, and competitors' performance, to optimize their decisions regarding race strategy and execution. This immense volume and speed of data processing required for real-time in-race decision-making exemplify the advanced integration of state-of-the-art data science needed for modern-day sports. From race-day tactics to season-long strategy, data-driven insights are vital for success within the sport (Sports, 2023) This highlights the importance of capturing and maximizing every weekend variable, from practice sessions on Friday through to the Sunday race execution itself, to find competitive advantages wherever possible.

While Formula One is a data-rich sport, teams privately analyze these relationships (and in much greater depth) using proprietary data and telemetry collected from an array of sensors. However, public and academic research on performance analysis in this sport is much more limited. One of the few published studies, Mühlbauer (2010), examined the link between qualifying and finishing positions using only the top eight starters from the 2006 to 2009 seasons, which is less than half of the competitive field. This restricted design limited generalizability and may have overstated the importance of starting grid position relative to qualifying performance.

1.2 Prior Work and Motivation

The relationship between starting-grid and finishing positions in Formula One has long been of interest to teams, analysts, and researchers. The most notable published academic analysis (Mühlbauer (2010)) concluded that starting grid positions were the strongest determinants of race outcomes. However, that study relied on a truncated sample; the top eight finishers across only four seasons (2006–2009), representing roughly 40% of the field. Such restrictions in sample design can distort statistical relationships, particularly when grid positions themselves are influenced by post-qualifying penalties.

To test whether this earlier conclusion was an artifact of limited sampling, the present study replicates that analysis using the same metrics before extending it to nearly two decades of data (2006–2023) and the complete grid of competitors. We then quantify the statistical relationships between practice, qualifying, grid, and finishing positions using contingency coefficients, rank correlations, and ordinal logistic regression to provide an updated, transparent evaluation of these associations across eras and throughout the race weekend. By addressing these shortcomings, this paper refines the empirical understanding of how different race-weekend sessions evolve and relate to the final outcomes in a Formula One Grand Prix.

1.3 Scope

Rather than constructing prediction models, an effort notable for future work as it is lacking within the academic space, this study focuses on quantifying the statistical relationships among a race weekend's sessions. By emphasizing the association rather than prediction, the results aim to clarify which performance measures most accurately reflect underlying car and driver capability, correcting prior published interpretations and extending the understanding across multiple technical eras.

1.4 Literature Review

Historical research within F1 has focused on the technical, physiological, and structural factors influencing motorsport performance. Hughes (1968) reviewed the importance of motor racing within the context of evaluating driver skill and car performance and highlighted the significance of measurement systems. Schwaberger (1987) analyzed the physiological demands on an F1 driver required to manage the intense forces experienced during a race, emphasizing the high fitness levels required. Bisagni and Terletti (2008) looked at the structural optimization of composites in F1 cars, demonstrating the critical role of engineering in competitive performance.

Pivoting to an analysis of the relationship between the starting grid and race outcomes has led to various studies emerging in the past 15 years. Mühlbauer (2010) provided an early investigation of this topic, examining 70 races from 2006 to 2009. This study serves as the inspiration for this paper while it found a strong correlation between the starting grid and finishing positions, it has many limitations. It suffers from a small sample size, both in drivers (the paper only examined the top 8 per race, while there are normally 20+ per race) and in seasons (just 4), all of which provide a limited scope of analysis. Additionally, the authors fail to address that the starting grids are already established by F1's qualifying process, a time-trial that allows cars to determine which is the fastest in a single lap ahead of the race; intuitively, the fastest car on a Saturday should likely be the fastest car on a Sunday, which potentially overstates the importance of grid positions.

In addition to this keystone study, McCarthy and Rotthoff (2013) analyzed Formula One's 2012 season to determine whether certain starting-grid positions were more prone to first-corner contact, which would likely negatively impact their finishing positions. Using a probit regression model on pre-racing starting and finishing data matched with crash reports, the authors find a non-linear relationship. Mid-grid positions, particularly those in position 10 and the last position, had the highest probability of collision, while the pole position and the last position faced the least risk. Their work uniquely introduced the concept of strategic qualifying behaviour, suggesting that, under certain conditions, drivers might intentionally avoid high-risk grid spots to maximize their race finishing positions.

Sobrie (2020) developed a predictive modelling framework for F1 race outcomes, employing various machine learning algorithms (including Decision Trees, Random Forests, Adaboost, Gradient Boosting Machines, and XGBoost). They compared class-imbalance techniques and evaluated performance using accuracy, AUC, F1, and lift metrics to predict whether a driver finishes in the top 3. Their results showed that XGBoost achieved the highest performance and handled class imbalance the best without degrading precision. This thesis also examined model regularization and the effects of overfitting specific to F1 race data, demonstrating how ensemble methods can capture competitive dynamics more effectively than simpler models.

Sicoie (2022) developed a machine learning framework to predict F1 race winners and championship standings using historical data from 2014 to 2021. Their study combined API and web-scraped datasets enriched with weather, circuit, and driver attributes. Applying Random Forest, Gradient Boosting, and Support Vector Regression models with cross-validation, the results showed high rank correlations between the predicted and actual standings and identified constructor and qualifying performance as dominant predictive features. This work stands out for integrating heterogeneous race, environmental, and driver data into a single supervised learning pipeline for season-long outcome predictions.

Patil et al. (2022) analyzed five seasons of Formula One data (from 2015 to 2019) to identify which technical and strategic factors most affect total championship points, which are derived from final race positions. Using correlation analysis, principal component analysis (PCA), and linear regression on 21 car and race variables, such as tyre usage, pit stops, penalties, and laps led, they were able to reduce the dataset to four principal components that explain 70% of the variance, which are linked to key factors such as average pole position, tyre choice, and race completion rate to points scored. Their results showed that the starting position and tyre strategy dominate season outcomes, offering a dimensional reduction framework for interpreting race performance.

van Kesteren and Bergkamp (2022) advanced the field by using a Bayesian multilevel rank-ordered logit model to separate driver skill from car performance, utilizing results from the 2014 to 2021 seasons. Using hierarchical random effects for drivers, constructors, and season-specific form, they estimated that roughly 88% of the variance in race outcomes is attributable to constructor performance, with the remainder due to drivers. Their approaches uniquely allow for the comparison of probabilities of drivers and teams on an Elo-like scale and demonstrate that constructor effects dominate results while enabling counterfactual inference, such as how a driver would perform in a different car.

Padilla (2023) looked into whether F1 teams achieve better performance by employing rookie or experienced drivers, using a casual-comparative qualitative analysis of race results, salaries, and points from the 2005 to 2019 seasons. Using ANOVA and regression analysis on driver experience groups (rookie, experienced, and veteran), the study found that while experienced drivers score more points on average, rookies outperformed 30% of experienced and 62% of veteran drivers at far lower costs, implying potential overpayment to the tune of $360 million over 15 years. This study uniquely introduces a cost-per-point efficiency metric to evaluate driver hiring decisions, rather than relying solely on raw performance alone.

Collectively, these contributions highlight progress in Formula One analytics but also leave unresolved questions about the fundamental relationship between qualifying, grid, and race performance that earlier work characterized incorrectly.

1.5 Objective and Contributions

Building on the limitations identified in Mühlbauer (2010), this study expands both the scope and rigour of analysis to improve our understanding of performance relationships across the entire Formula One race weekend. While earlier studies were valuable starting points, they were heavily restricted by the small sample size of their dataset selection (i.e., top eight starters only, four seasons, etc.) and by a narrow focus on the starting grid, which is already self-sorted by the results of the prior day's qualifying session.

This work contributes to this domain in three main ways:

1. Firstly, it replicates the prior work to demonstrate how this truncation of the sample influences the statistical associations between racing sessions.
2. Secondly, it extends the analysis and dataset to include all competitors on the track over a span of two decades, enabling a longitudinal view of how session relationships have evolved under changing technical and sporting regulations.
3. Finally, it applies multiple complementary approaches to contingency coefficients, such as rank-order correlations and ordinal logistic regression, to further quantify the relative association of weekend sessions with the final race results.

Together, these contributions correct earlier misinterpretations and offer a transparent, data-driven foundation for future research on race performance dynamics, providing an empirical benchmark for future studies of race-weekend performance relationships.

2 Materials and Methods

2.1 Analytical Framework

This study follows a two-stage approach. First, we replicate the work that was previously conducted by Mühlbauer (2010) using the same variables, seasons (2006 to 2009 inclusive), and the truncated top-eight starter sample to verify their reported associations and ensure that we are aligned in our approach. Second, we extend this analysis of the full grid of competitors, including nearly two decades of races (2006 through 2023 inclusive), and evaluate whether these relationships hold under complete data and modern conditions. Throughout, the focus is on quantifying associations among race-weekend sessions, practice, qualifying, starting grid, and finishing positions, rather than constructing predictive or forecasting models.

Beyond expanding the temporal and sample scope of prior work, this study also deepens the analytical framework. Whereas Mühlbauer (2010) relied solely on the contingency coefficient (C) to describe the nominal association between the starting grid and finishing position, we introduce additional statistics that capture both nominal and ordinal relationships. Specifically, we compute:

- the **Relative Frequency (RF)** statistic to evaluate session-to-session positional consistency;
- **Spearman's rank correlation coefficient** (ρ) measures the monotonic association between ranked sessions and race results; and

- an **Ordinal Logistic Regression (OLR)** model to compare standardized association strengths across sessions.

Together, these measures provide a multi-level assessment of how practice, qualifying, and grid results relate to race outcomes while maintaining methodological continuity with the original publication.

2.2 Data Preparation and Cleaning

Data for this study was collected by scraping official race results from the Formula 1 website (`http://www.Formula1.com`). The dataset includes every Grand Prix from 2006 through 2023 (as scraping was completed before the end of the 2024 season). In total, more than 400 races were captured, yielding a comprehensive record of driver and team performance.

Exclusions and data completeness. To ensure comparability across events, several types of weekends were excluded. Races with incomplete or disrupted session data were removed; for example, the 2023 Emilia Romagna Grand Prix (cancelled due to flooding) and the 2017 Chinese Grand Prix (practice sessions cancelled due to fog). Any weekend (driver specific) where a single chassis was driven by multiple drivers, such as short-term illness substitutions or rookie test opportunities, was also excluded.

Beginning in 2021, Formula 1 introduced "Sprint" weekends featuring a short Saturday race in place of the standard three-practice format; these were excluded because their condensed schedule and altered sequence of sessions make them non-comparable to traditional weekends. Key variables include:

- **Season**: year of the race.
- **GP**: name of the Grand Prix.
- **gp_code**: unique identifier for the race weekend.
- **No**: driver's number.
- **Driver**: driver's name.
- **Car**: team or constructor represented by the driver.
- **Gap**: time gap to the session leader.
- **Laps**: number of laps completed in the session.
- **Pos**: position ranking within the session.
- **Time**: best lap time achieved.
- **Qualifying Time**: best time across qualifying stages.
- **Race Start Position**: grid position after penalties or disqualifications.
- **Race Finish Position**: final classified race result.
- **Status**: completion status or DNF reason from FIA timing sheets.

- **Points**: points awarded for the finishing position.

A snapshot of representative rows is provided in Table 1.

Season	GP	Driver	Car	Pos (Qualy)	Race Start	Race Finish
2006	Australia	M. Schumacher	Ferrari	11	10	NC
2006	Australia	F. Massa	Ferrari	16	15	NC
2006	Australia	R. Barrichello	Honda	17	16	7
2006	Australia	J. Button	Honda	1	1	10
2006	Australia	A. Davidson	Honda	NaN	NaN	NaN

Table 1: Sample of the dataset showing key columns and entries.

Data cleaning and structure. The remaining events were combined into a single analytic dataset and pivoted so that each row represented one complete driver–weekend record. Any event in which a driver failed to participate in all three practice sessions due to weather, illness, or driver changes (e.g., United States: 2015 rain; Germany: 2020 fog; Russia: 2021 rain) was excluded. These criteria kept the driver–car pairing constant and preserved full three-day weekend data for every observation. Non-finishers (DNFs) were retained and coded using their classified finishing positions, as listed in FIA results, preserving the race order at the time of retirement. Penalties affecting grid positions were maintained, as they represent the official starting order used in each Grand Prix. Each observation, therefore, included all session results (Practices 1–3, Qualifying, Race) and related metadata.

Final dataset. The cleaned dataset comprises over 7,800 driver–race observations, each representing a single driver–car pairing across a full weekend. The full dataset used for this analysis is available upon request. By standardizing inclusion criteria and weekend structure, the dataset enables a consistent comparison of performance evolution from practice through qualifying to race results.

2.3 Experiment, Analysis, and Evaluation Methods

Overview and Analytical Approach. As described in section 1.4, this study validates the work done by Mühlbauer (Mühlbauer, 2010) and adds to their methodology and results. The relationship between the starting grid position and the finishing position was analyzed using the contingency coefficient (C), as in the original work. The contingency coefficient quantifies the strength of association between two nominal variables by adjusting the chi-squared statistic for sample size (Agresti, 2010; Goodman and Kruskal, 1954).

Contingency Coefficient (C). For a $k \times k$ table of observed counts O_{ij} and expected counts E_{ij}, the coefficient is defined as:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}, \qquad \text{where} \quad \chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \tag{1}$$

and n is the total number of observations. Values of C range from 0 (no association) to a theoretical maximum of less than 1, which depends on k.

Interpretation and Rationale. The contingency coefficient C is derived directly from the chi-squared (χ^2) test of independence, which evaluates whether two categorical variables are statistically associated. In our work, χ^2 measures the deviation of the observed cell count in the position-position contingency table from those expected under independence, and C rescales this statistic by the sample size (n) to yield a bounded measure of association. This formulation of C ensures that it remains interpretable across datasets of different sizes; values near 0 indicate little association between the session and race outcomes, while values approaching 1 indicate a stronger dependence. Unlike χ^2, C is descriptive rather than inferential and is not used here for hypothesis testing; rather, it is used as a comparative measure of association strength across seasons and sessions. This method is retained to ensure methodological continuity with the prior study. C is used as it was specifically designed to measure the strength of the association between two nominal variables (e.g., comparing starting position and finishing position) and is well suited for comparing categorical variables. By using C, the original authors were able to account for the association between these two variables with a limited number of categories.

Relative Frequency (Rel Freq) Statistic. Following Mühlbauer (2010), we also compute the Relative Frequency (RF) statistic to capture the probability that a driver's rank in one session equals their rank in another. Formally (Sidney, 1957):

$$RF = \frac{1}{k} \sum_{r=1}^{k} \Pr(\text{Finish} = r \mid \text{Session} = r), \tag{2}$$

where k is the number of classified finishing positions. Higher values indicate greater consistency between session and race rankings.

Rank-Order Correlation (Spearman's ρ). We also included an additional metric in our analysis with the introduction of Spearman's ρ aimed at measuring rank-order correlations within the practice sessions. While C captures the strength of nominal associations, Spearman's ρ provides an alternative by assessing monotonic relationships between ranked variables. Spearman's ρ was calculated between each session's position ranks (e.g., Practice 1, Practice 2, Practice 3, qualifying, and starting grid placements) and the final race finishing positions to assess the rank-order consistency across all phases of a Grand Prix weekend. This produced one correlation per session-finish pair for each season, with the summarized results presented in Table 2 alongside their corresponding C and Rel. Frequency statistics. Observations containing missing session data (as described in Section 2.2) were excluded listwise to ensure that only complete driver-weekend combinations were included in the correlation analysis. This combination of nominal and ordinal association metrics allows for a consistent comparison of strength across different types of relationships observed within the dataset and provides us with another perspective to understand the trends within the sport.

Replication of Prior Study. Replicating Mühlbauer (2010)'s method, we first limited the analysis to the 2006 to 2009 seasons (inclusive), focusing on the top 8 positions in each race. This allowed us to validate their findings and create a baseline from which we can further explore the other sessions in a Grand Prix weekend, as well as fully understand the effectiveness of their approach. The replication followed the same inclusion criteria, using only the top eight finishers per Grand Prix, and computing contingency coefficients and relative frequency values between the starting grid and finishing positions for each race. Missing drivers (DNF or unclassified) were still excluded from the dataset (as described in Section 2.2) to preserve a consistent $k \times k$ structure, ensuring direct comparability with the original methodology and results.

Extended Dataset and Seasonal Analysis. The next step we took was to fully understand the impact of the race weekend evolution and to expand upon the initial work. To ensure consistency across races, we included the full dataset to ensure that all starting and finishing positions from each race were reviewed (most races consist of 20 drivers in this era). This approach ensures that all races include the full competitive field (a limitation noted in prior work). We also examined the years from 2010 to 2023, electing to start in 2010, as this season saw major technical regulation changes, as well as the introduction of the current points system. For the extended analysis, we computed the same association measures separately for each season and then reported both annual and pooled means to account for within-season variability. Each Grand Prix contributed one observation per driver, yielding approximately 7,800 driver–race instances in total. This approach ensured consistent weighting across seasons, despite variations in the number of events per year.

Building on the expanded dataset, we analyzed the relationship between the starting grid and finishing results, as well as the results of each practice session (P1, P2, and P3), and the qualifying results, to understand the statistical reliability of these sessions. Session results (P1, P2, P3, qualifying, and race) were aligned chronologically within each Grand Prix and merged by driver name and unique weekend code. Only drivers with valid entries for all sessions were retained to ensure consistent pairwise comparisons between session and race results. This alignment preserved the temporal progression of the weekend and prevented mismatched driver-session combinations. Further analyzing these year over year trends, we can see if the strength of the relationship has changed over time, highlighting any potential changes in the reliability of the grid or changes within the sport.

To account for the potential variability in field strength across seasons, reflecting the differences in car competitiveness, team dominance, and driver talent, the analysis was conducted separately by season and aggregated only after the within-season estimates were computed. This approach attempts to minimize bias arising from unequal competitive strengths between eras and years.

Ordinal Logistic Regression (OLR) Modelling. In our final step within this work, we employed an Ordinal Logistic Regression (OLR) model to quantify the relationship between practice and qualifying sessions with the final race performance The predictor variables $(x_{P1}, x_{P2}, x_{P3}, x_Q, x_G)$ were coded as integer rank positions (i.e, 1 = best, 2 = second best, 3 = third best, etc), preserving their ordinal nature. The dependent variable for the OLR Y represents the driver's final classified finishing position at the end of the race (also expressed as an ordered integer) No standardization or transformation was applied, as OLR estimates relationships based on the rank ordering rather than numerical distances between categories.

The ordinal logistic (proportional-odds) model expresses the cumulative log-odds of finishing position Y at or above a given threshold j (McCullagh, 1980; UCLA Statistical Consulting Group, nd; Singh, 2016) as:

$$\log\left[\frac{\Pr(Y \leq j)}{\Pr(Y > j)}\right] = \alpha_j + \beta_{P1}x_{P1} + \beta_{P2}x_{P2} + \beta_{P3}x_{P3} + \beta_Q x_Q + \beta_G x_G, \tag{3}$$

where x_{P1}, x_{P2}, x_{P3} are practice positions, x_Q is the qualifying position, and x_G is the grid position. The coefficients *beta* were standardized to enable a comparison of their relative association strengths among GP sessions. The resulting coefficients are interpreted as comparative measures of association strength rather than as predictive weights that align with the descriptive objectives of this study. This method is well suited for our target variable (race finishing position), which is ordinal with a natural ranking but lacks equal spacing between categories. Unlike linear regression, which assumes continuous and interval-scaled outcomes, OLR models the cumulative log-odds of finishing position thresholds, assuming proportional effects of predictors across outcome levels. This structure, known as the promotional-odds model, is appropriate for motorsport results, where the finishing ranks are ordered but position gaps may vary (i.e., 1st and 2nd finishing 3 seconds apart vs 30 seconds) due to incident-driven dynamics, race strategy, or other external factors.

Model Implementation and Robustness Checks. The ordered logit model was implemented using the `OrderedModel` class in `statsmodels` (distribution = logit) and estimated via maximum likelihood. The dependent variable represents each driver's final classified finishing position, while the predictors include the three practice session ranks (P1, P2, P3), the qualifying rank, and, where relevant, the starting grid position. All predictors were coded as integer ranks (i.e., 1 = best, 20 = worst) and entered without transformation to preserve their ordinal interpretation. This cumulative logit framework enables interpretable estimation of how each session contributes to a driver's likelihood of achieving a higher finishing position (UCLA Statistical Consulting Group, nd; Singh, 2016).

To ensure the robustness of the OLR modelling on this data, we conducted a multicollinearity check (e.g., variance inflation factors) to ensure variables were not overly correlated, standardized regression coefficients (β values) to assess the relative importance of each race session, a model fit evaluation (McFadden's R^2) to understand the explanation of variance in race finishing positions, and significance testing (p-values) to determine which factors had a meaningful impact on race outcomes. By using OLR, we aim to provide a much deeper understanding of how the different phases of an F1 weekend contribute to final race performance. This allows us to assess the strengths of these relationships and provide further statistical insight into weekend dynamics that can be applied within the sport. All modelling steps were evaluated for robustness using multicollinearity diagnostics (variance inflation factors) and significance testing. The analyses are strictly descriptive in scope, aiming to quantify relative association strengths among race-weekend sessions rather than to forecast future race results. All analysis was conducted in Python 3.12 using Pandas, statsmodel and Matplotlib in Python Notebooks. The combination of replication, expanded metrics, and regression modelling ensures a comprehensive evaluation of the relationship between race weekend variables and final results.

3 Results

3.1 Replication of Original Work (2006-2009, Top Eight)

We first replicated the original analysis of Mühlbauer (2010) using the same sample restrictions: the 2006–2009 seasons and the top eight finishers in each race. This replication produces comparable results ($C = 0.628$, Spearman's $\rho = 0.589$, $p < 0.001$), confirming the previously reported association between starting and finishing positions. These results validate the original calculations but also demonstrate the effect of sample truncation: when extended to include all 20 drivers, the association coefficients increase substantially ($C = 0.768$, $\rho = 0.741$, $p < 0.001$). The relative frequency (Rel Freq) of maintaining starting positions to the end of the race in the top 8 was 0.558 which decreases to 0.543 when the full grid is included. These highlight the sensitivity of earlier conclusions to restricted sampling.

The results are summarized in Table 2, which outlines the contingency coefficients (C), Spearman's ρ values, and Relative Frequency for various rank comparisons across the analyzed seasons and permutations.

Start	End	Top n	Rank 1	Rank 2	C	Spearman Rho	Rel Freq
2006	2009	8	Race Starting Pos	Race Finish Pos	0.628	0.589	0.558
2006	2009	20	Race Starting Pos	Race Finish Pos	0.768	0.741	0.543
2010	2023	20	Race Starting Pos	Race Finish Pos	0.734	0.748	0.516
2010	2023	20	Qualification Pos	Race Finish Pos	0.741	0.763	0.518
2010	2023	20	P3 Position	Race Finish Pos	0.674	0.676	0.350
2010	2023	20	P2 Position	Race Finish Pos	0.664	0.674	0.324
2010	2023	20	P1 Position	Race Finish Pos	0.627	0.629	0.300
2010	2023	20	P3 Position	Qualification Pos	0.742	0.757	0.390

Table 2: Replication of Mühlbauer (2010) (2006-2009, top eight) and extended analysis including the full grid (top 20). Values show contingency coefficients (C), Spearman's ρ (-1 to 1), and Relative Frequency (0-1), where higher values indicate stronger association between sessions.

3.2 Extended Association Analysis (Full Grid, 2006–2023)

Extending the analysis to the full grid of competitors and additional seasons (2010–2023) reveals that the strength of association between starting position and final race result remains robust, though slightly reduced compared to 2006–2009. Across this modern period, $C = 0.734$ and Spearman's $\rho = 0.748$ ($p < 0.001$), the Relative Frequency declined modestly to 0.516. These trends likely reflect both greater parity among teams and the impact of regulatory and strategic variability introduced after 2010.

Post 2010, following regulatory changes in F1 (such as the removal of in-race refuelling and the expansion of points), the correlation between starting position and race finishing has declined. For the top 20 positions from 2010 to 2023, C was 0.734 and Spearman's ρ was 0.748 ($p < 0.001$). The Rel Freq in this period was 0.516, which suggests a small reduction in consistency from the 2006-2009 eras.

When the qualifying position is used instead of the starting grid position, associations strengthen ($C = 0.741$, $\rho = 0.763$), confirming that qualifying offers a more direct measure of performance unaffected by post-session penalties. The Rel Freq of all positions during this era was 0.518 continuing to reflect the critical importance of qualifying success.

We hypothesize that qualification positions are more strongly associated with race results compared to the initial starting grid due to the fact that the qualification process sorts cars from the fastest to the slowest. The starting grid adjusts these results due to the application of penalties (e.g., replacing power units, exceeding gearbox allocations, or infractions during practice), which gives a distorted view of the fastest to slowest cars. This can artificially inflate slower cars, dampen faster cars, and distort the observed predictability of starting positions. By contrast, these qualification results are more likely to be unadjusted outcomes of outright single-lap pace and performance for all cars and drivers under the same conditions, making them a more reliable indicator of race outcomes. Teams are encouraged to optimize qualification setup for raw speed, unencumbered by strategic influences, and to have their drivers start in an optimal position given the difficulty of passing in modern F1.

3.3 Session Level Association

The association between early weekend performance and race outcomes varies by session. Practice 3 (the session right before qualifying) presents the strongest association with race results ($C = 0.674$, $\rho = 0.676$, $p < 0.001$), followed by Practice 2 ($C = 0.664$, $\rho = 0.674$) and Practice 1 ($C = 0.627$, $\rho = 0.629$). Relative frequencies similarly increase from P1 (0.300) to P3 (0.350), reflecting how teams progressively optimize setups closer to qualifying.

These results are summarized in Table 2, which outlines the contingency coefficients (C), Spearman's ρ, and Relative Frequency for each session comparison.

3.4 Visualizing Association Strength

Figure 1 presents a heatmap demonstrating the relationship between qualifying positions (Pos_q) and race positions (Pos_r) for the 2010-2023 season. There is a strong diagonal trend indicating that drivers starting closer to the front tend to finish towards the front at the end of the race. This trend is supported by the data presented above in Table 2.

Interestingly, this relationship tends to break down for the lower-ranked qualifying cars, suggesting a less consistent relationship at the back of the grid. While pole position (first on the starting grid, usually first in qualifying but not always due to post-qualifying penalties) overwhelmingly translates to top race finishes, we see this distribution among the mid-field and back-markers on the grid. This is likely due to the greater susceptibility of drivers caused by first-corner incidents, strategic constraints, overtaking challenges related to DRS (Drag-Reduction System), and moving up in positions when cars ahead of them do not finish due to incidents and reliability.

Overall, this suggests that qualifying positions, being unaffected by race-day penalties, are better indicators of race performance than starting grid positions.

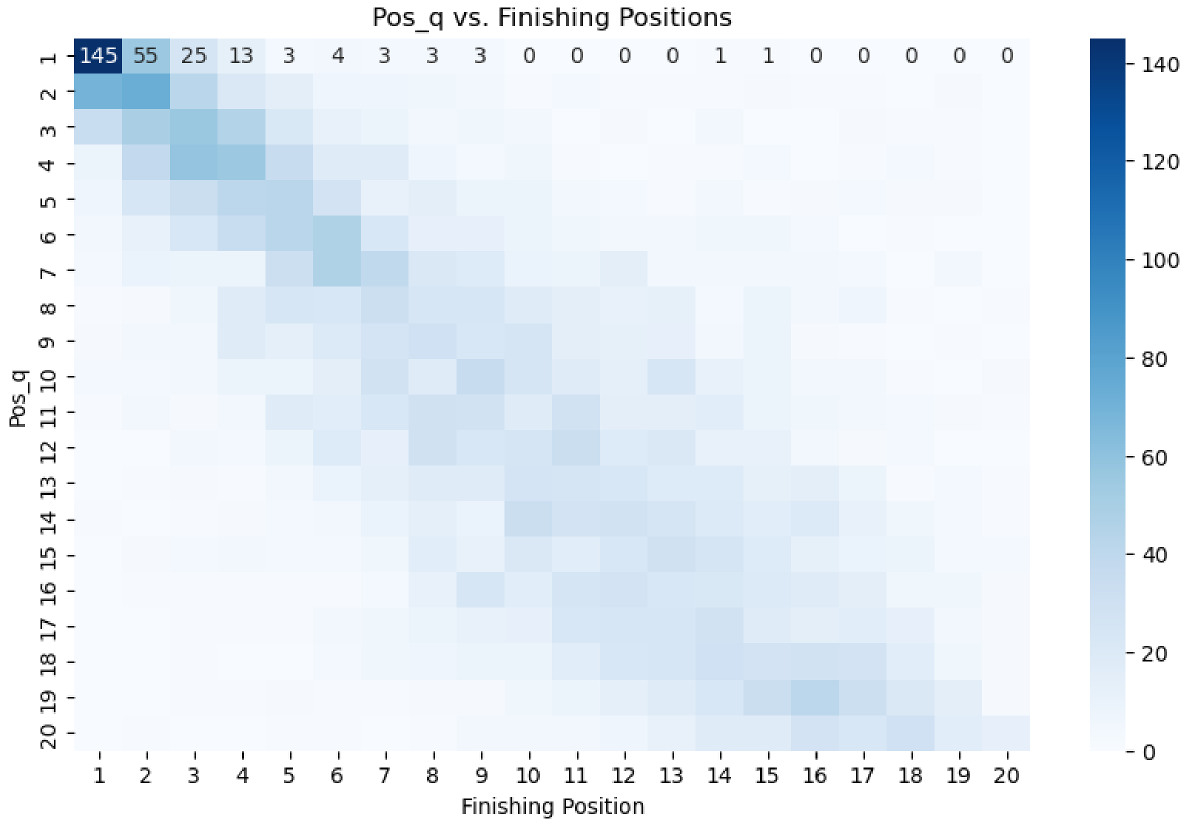

Figure 1: Heatmap showing the relationship between qualifying positions (y-axis, Pos_q) and race finishing positions (x-axis, Pos_r) for the 2010–2023 seasons. Darker cells represent higher frequencies of drivers finishing in the same or nearby positions. Axes range from positions 1 (best) to 20 (worst).

3.5 Analysis by Technical Eras

When we examine F1 from the perspective of their major technical eras (as presented in Table 3), to see how changes in rules might impact this relationship, the relationship between qualifying and finishing positions shows a continued trend. From 2006 to 2008, during the V8 era, the correlation was moderate ($C = 0.628$, $\rho = 0.589$, $p < 0.001$) with a Rel Freq of 0.557. This improved in the 2009-2013 era, during which cars used V8s with KERS (Kinetic Energy Recovery System), leading to a stronger association ($C = 0.744$, $\rho = 0.745$), however, Rel Freq dropped to 0.478, likely due to the reduced reliability of new technologies.

The 2014-2022 turbo-hybrid era demonstrated the strongest correlation between qualifying positions and race finishes ($C = 0.756$, $\rho = 0.777$, $p < 0.001$) with a relative frequency of 0.519. The introduction of hybrid power units further reinforced the importance of qualifying and engineering reliability, as power unit efficiency and durability became key factors in race performance. These trends underscore how F1's evolving technology has shaped the strength of the race session relationship and performance outcomes, with qualifying continuing to be the most important moment of performance throughout the weekend.

Start	**End**	**Era**	**Rank 1**	**Rank 2**	**C**	**Spearman's ρ**	**Rel. Freq**
2006	2008	V8 Engines	Qual. Pos.	Finish Pos.	0.628	0.589	0.558
2009	2013	V8 + KERS	Qual. Pos.	Finish Pos.	0.744	0.745	0.478
2014	2022	Turbo-Hybrid	Qual. Pos.	Finish Pos.	0.756	0.777	0.519

Table 3: Association between qualifying and race finishing positions across F1 technical eras. Values show contingency coefficients (C), Spearman's ρ (-1 to 1), and Relative Frequency (0-1).

3.6 Season-by-Season Variation

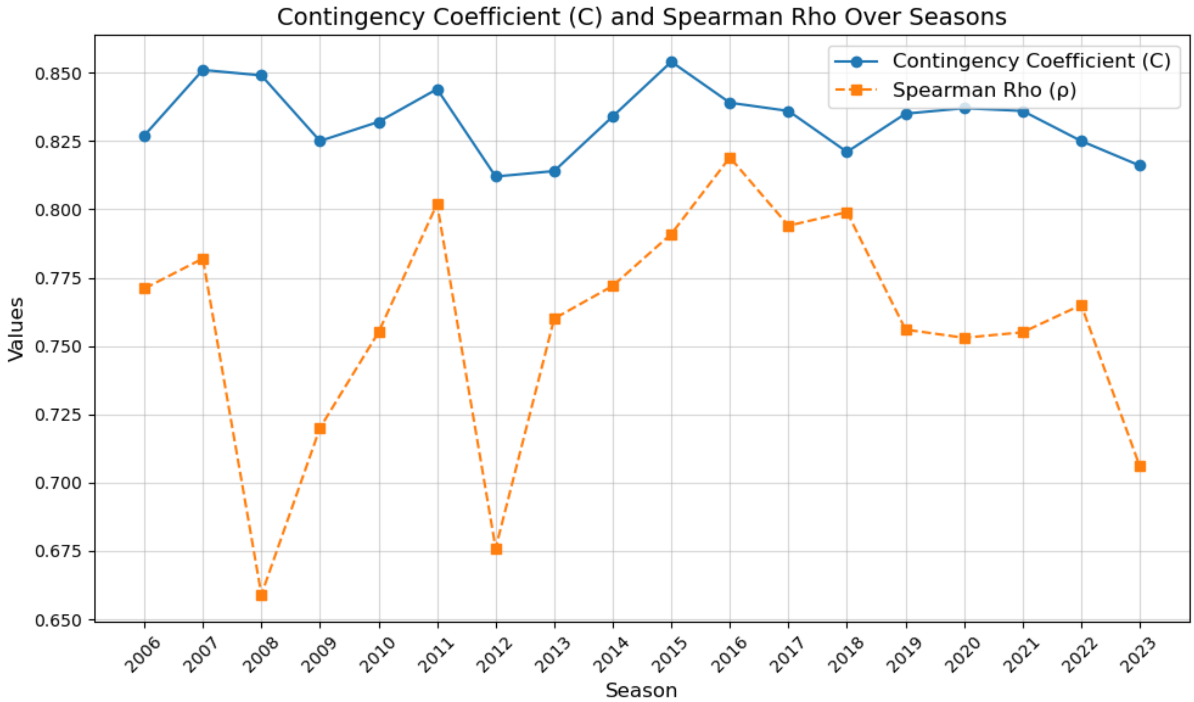

Figure 2: Season-by-season analysis of the contingency coefficient (C) and Spearman's ρ for qualifying positions versus race finishing positions (2006-2023). Higher values indicate stronger association between qualifying and finishing performance. Y-axis truncated at 0.65 for visual clarity; correlations range from -1 to 1.

Figure 2 compares the contingency coefficient (C) and Spearman's ρ between qualifying results and race finishes, season by season. Both metrics show consistently strong associations (with C values generally exceeding 0.81 and ρ over 0.65), confirming that qualifying performance remains a reliable indicator of race outcomes across all years. However, their trajectories do not always align perfectly, as C and ρ capture different statistical properties of the same relationship.

The contingency coefficient (C) reflects categorical concentration, indicating how tightly finishing positions cluster around specific grid slots, while Spearman's ρ measures the monotonic consistency of ranks across the entire field. As a result, C can remain high in seasons dominated by a few top-performing teams (e.g., Ferrari dominating in the mid/late 2000s, Redbull Racing dominating from 2010 to 2013 and in 2023, and

Mercedes dominating from 2014 to 2020), even if mid-field ordering varies substantially, which lowers ρ. Conversely, when performance is more evenly distributed while rank order remains stable, ρ may exceed C. Using both statistics provides complementary insight, as C captures categorical concentration while ρ reflects rank-order consistency, together offering a fuller picture of competitive balance across seasons.

This divergence highlights how different competitive dynamics influence the two statistics. In eras of strong team dominance, grid positions translate into race outcomes more categorically (high C), whereas, in more competitive seasons, positional order is preserved across the field but is less tightly clustered (high ρ). These complementary metrics together show that qualifying remains a strong predictor of race outcomes, though the exact pattern of association varies with the competitive balance and technical regulations of each season. While this work does not explicitly control for variation in field strengths across different seasons, these differences in competitive parity provide useful context for interpreting the fluctuations in both association measures.

3.7 Ordinal Logistic Regression

3.7.1 Results and Analysis

An ordinal logistic regression (OLR) model (as described in Equation 3) was applied to quantify the relative association strength between session rankings and the final race outcome. The OLR model included the independent variables of the three practice session rankings (Pos_p1, Pos_p2, Pos_p3) as well as the qualification placement (Pos_q). Table 4 presents the outputs from the OLR, highlighting the estimated coefficients (β), standardized regression coefficients (β), and the statistical significance (p) of each predictor.

Predictor	Coefficient (β)	Standardized β	p-value
Pos_q (Qualification Position)	0.2545	1.558	< 0.001
Pos_p3 (Practice 3 Position)	0.0610	0.379	< 0.001
Pos_p2 (Practice 2 Position)	0.0576	0.368	< 0.001
Pos_p1 (Practice 1 Position)	0.0463	0.291	< 0.001

Table 4: Ordinal Logistic Regression coefficients (β) and standardized β values for session rankings predicting race finish. Higher standardized β values indicate stronger association with finishing position.

The results of the OLR suggest that the Qualification position shows the strongest association with final race placement, with a coefficient of 0.2545 This implies that for every position improved in Qualification, the log-odds of achieving a better finishing position in the race increase by 28.9% ($e^{0.2545} = 1.289$). Among the practice sessions, Practice 3 exhibits the highest association in the final race standings, followed by Practice 2, and then Practice 1. This suggests that Practice 3 is the most relevant session for race outcomes (among practice sessions) and reinforces the importance of teams using this time to optimize car setup ahead of qualifying. In contrast, Practice 1 has the smallest effect and confirms that early-weekend sessions are better exploratory sessions for data collection or potentially new driver testing.

While raw coefficients provide some insight into the direct relationship between these variables and the final race result, they cannot be directly compared due to the scale of the predictors. To address this, we calculated the standardized regression coefficients (also reported in Table 4). These standardized coefficients present results similar to the relative association strength among sessions. In addition, they confirm that qualification has an impact that is about four-times larger than any of the practice sessions. This pattern empirically supports the strategic emphasis that teams place on qualifying performance. Among the practice sessions, Practice 3 is the most important and confirms the significance of establishing an optimal setup within this window ahead of the qualification session. While teams already use Practice 3 to fine-tune setups, this work provides empirical evidence that performance in P3 significantly predicts race outcomes, reinforcing the importance of this session.

3.7.2 Threshold Estimates and Performance Barriers

We present the threshold values from the OLR model in Figure 3. These thresholds define the log-odds cutoffs between ranking categories, which provide us with insight into the difficulty of progressing through the grid. Alternatively, this can be stated as follows: these cutoffs represent the log-odds of a driver finishing at or above a given position, the more negative the value, the greater the difficulty in surpassing that position, while smaller values suggest smoother transitions.

The 1.0/2.0 threshold (-0.1957) is one of the least negative values of all thresholds, which suggests that the pole position is more vulnerable to being lost than most other transitions (one cannot finish higher than first). This reflects real world patterns where P1 can be seen to lose the lead in the opening laps due to first-corner incidents, poor starts, slip-streaming, and strategies. Pole sitters should focus not just on qualification but also on their race start procedures.

The 2.0/3.0 threshold is the most unique threshold, as it is the only positive value (0.0441) that indicates finishing in P2 versus P3 is more evenly distributed compared to other transitions In other words, once drivers are able to reach the podium ranks, it is much more fluid to swap and finish between P2 and P3. This aligns with real world trends, where close margins, race strategy, pit stop execution, and team orders can all have an impact.

Conversely, with the 3.0/4.0 threshold, the negative threshold value (-0.3036) indicates a key difficulty in standing on the podium. While P4 is often within reach of a podium finish, P3 needs to maintain a significant performance gap over P4, further emphasizing all the external factors that can affect final race standings.

The most negative threshold value comes from the 10.0/11.0 threshold (-0.7979), which indicates that breaking into the Top 10 is the hardest transition. This aligns with and reflects Formula 1's competitive structure, where only the top 10 positions in a race earn points. For backmarker teams, single digit points can make the difference in their final season standings and, thus, the earnings they can bring home at the end of the season. Drivers running in P11 need to consider greater strategic risks to attempt to secure a points finish.

The remainder of the threshold values, in the midfield positions, tend to indicate a smoother transition, indicating that these positions on the field likely experience the most frequent position changes. This can arise from cars ahead not being able to finish and missing the opportunity to move up in the ranking. It also implies that these teams have more opportunities to try to move closer to final positions with a point finish.

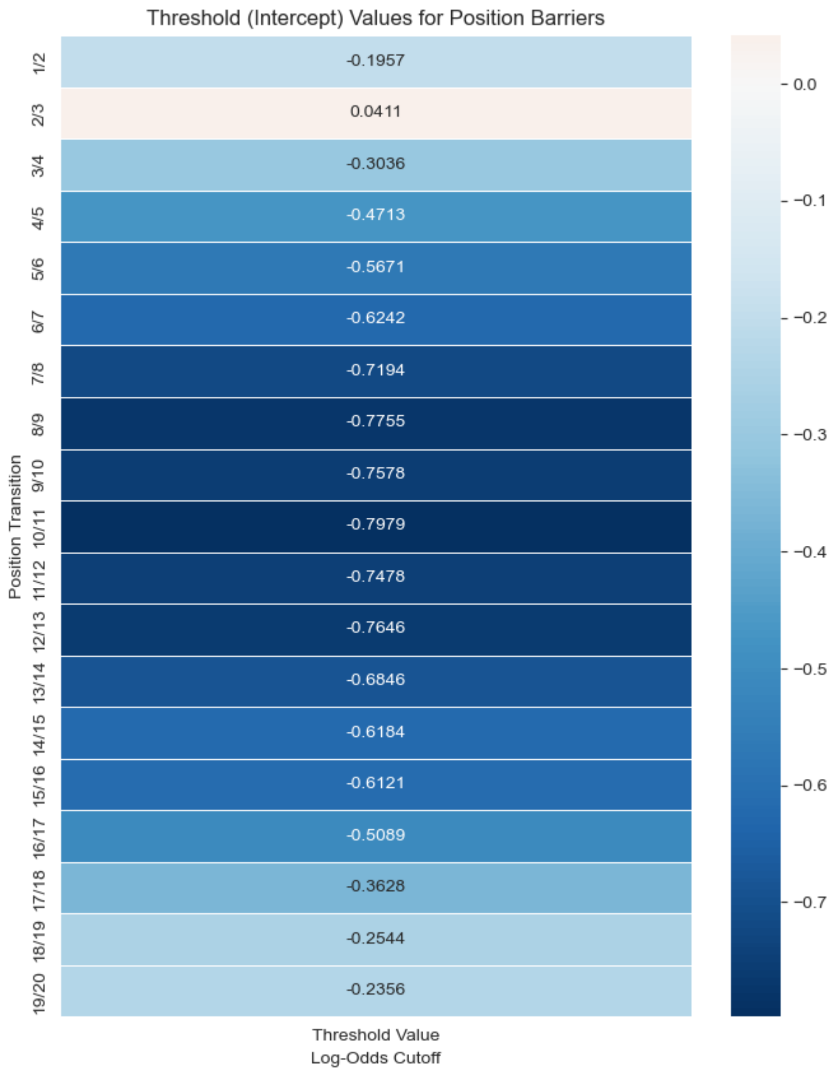

Figure 3: Threshold Estimates (Intercepts) from Ordinal Logistic Regression. X-axis = finish-position thresholds; Y-axis = cumulative log-odds intercepts. Thresholds represent the cumulative log-odds cutoffs between ranking categories; more negative values indicate greater difficulty in surpassing that position.

3.7.3 Model Evaluation and Fit

To assess our association model's fit, we first computed McFadden's pseudo-R^2, which compares the log-likelihood of the fitted model against a null (intercept-only) model. We obtained a pseudo-$R^2 = 0.0166$ (low R^2 is normal for descriptive ordinal models), indicating that while our model captures directional relationships between session results and final positions, the majority of the variance in race outcomes remains unexplained. This reflects the inherently unpredictable nature of Formula One, where race-day incidents, weather, and strategy often override qualifying indicators. The model's Akaike Information Criterion (AIC) of 25,657.28 provides a useful benchmark for comparing competing models. However, this pseudo-R^2, as used in ordinal logistic regression, is widely criticized for offering limited insight into the overall suitability of the model (Fagerland and Hosmer, 2016, 2017).

To address this concern, we also implemented a Pearson chi-square goodness-of-fit test that compares observed and expected frequencies across the predicted outcome categories. This test yielded a statistic of $\chi^2 = 6,828.25$ with 414 degrees of freedom ($p < 0.001$), which indicates a statistically significant deviation between the predicted and observed outcomes. While this result suggests a lack of fit, such chi-square statistics are highly sensitive to large sample sizes and sparse contingency tables, which could inflate Type I errors (Fagerland and Hosmer, 2016; Lipsitz et al., 1996).

Other options were considered for goodness-of-fit tests, such as the C_g (Hosmer-Lemeshow type), the Lipsitz test, and the Pulkstenis–Robinson (PR) test. Given their difficulty in implementation with standard Python libraries, we opted for the Pearson chi-square goodness-of-fit; the others are recommended in the literature for their sensitivity to different forms of model misspecification. As an example, the C_g and Lipsitz tests are more effective at identifying issues related to continuous predictors, while the PR tests are better suited to models with categorical predictors (Fagerland and Hosmer, 2017; Lipsitz et al., 1996; Gertheiss et al., 2023). Future work could incorporate these more targeted methods.

Despite the model's low overall explanatory power, the qualifying position remains the most influential predictor among all weekend sessions. Rather than forecasting exact outcomes, the model quantifies the relative association strength of each session with final performance. All together, while the model indicates directional trends, the residual lack of fit may indicate that important sources of variation (e.g., weather, pit strategy, and race incidents) are not fully accounted for in this current model. Future refinement and the incorporation of additional covariates are likely necessary to improve calibration.

3.7.4 Multicollinearity

To assess the robustness of our OLR model, we evaluated the presence of multicollinearity among the predictor values using Variance Inflation Factors (VIFs). High multicollinearity can distort coefficient estimates, making it difficult to determine the true influence of each predictor on race outcomes.

We first computed the VIF scores for all predictor variables in the OLR, which were the three practice sessions (Pos_p1, Pos_p2, Pos_p3) and the qualification position (Pos_q). The results can be seen in Table 5 and reveal significant multicollinearity, particularly between qualification positions and the later practice sessions.

Using a baseline of a VIF score of 5 to indicate moderate collinearity, and 10 or higher suggesting severe collinearity, we can see from Table 5 that both Pos_q and Pos_p2

Predictor	VIF Score
Pos_q (Qualification Position)	12.110
Pos_p3 (Practice 3 Position)	9.608
Pos_p2 (Practice 2 Position)	10.443
Pos_p1 (Practice 1 Position)	7.951

Table 5: Variance Inflation Factor (VIF) Scores (Initial Model). Higher scores indicate greater collinearity; values above 10 suggest severe multicollinearity.

exhibit substantial multicollinearity, with Pos_p3 right behind. This suggests that there is a strong relationship between a driver's late weekend performance and their qualifying results, which aligns with the strategy whereby teams use these later sessions to fine-tune car setup and maximize qualifying performance.

Given the high collinearity among these sessions, we tested two approaches to mitigate this. First, we iteratively removed the highest VIF contributors; first Pos_p2, then Pos_p1. However, this still left a substantial correlation between Pos_p3 and Pos_q (VIF of 8.396279 for both). We then used Principal Component Analysis (PCA) to preserve informational variance by combining the three practice sessions into a single composite variable termed "Practice Performance Score," dramatically lowering VIF to near 1.000.

While PCA was mathematically effective in reducing multicollinearity, it came at a significant cost to the explanatory power and usefulness of our model. After applying OLR, we found that while the log-likelihood improved slightly (-12,802 to -12,286) and the AIC dropped slightly (25,467 to 24,621), McFadden's R^2 fell to 0.00, suggesting that the model lost almost all explanatory power, and the Practice Performance Score became statistically insignificant ($p = 0.590$).

Altogether, these results confirm that qualification is the strongest individual predictor of race outcomes; however, they also emphasize that Practice 3 plays a crucial role in preparing for qualification. The strong collinearity between these sessions suggests that teams optimizing for performance in Practice 3 are more likely to achieve a strong qualifying result and thus dictate a strong race outcome. Despite multicollinearity among sessions, the qualification position remains the most robust explanatory variable, reinforcing its dominant statistical association with race outcomes. Teams should evaluate progression across the entire weekend; however, their combined impact cannot be captured by a single aggregate score.

4 Discussion and Conclusion

4.1 Summary and Correction of Prior Findings

The replication of prior findings presented in this work demonstrates that the conclusions of Mühlbauer (2010) were likely largely influenced by their decision to truncate the sample size and focus on reviewing only the top eight starting positions within a race. By restricting the analysis to only these top positions, this data selection understates the strength of the relationship between the starting position and their final race outcome. By using a full-grid dataset spanning two decades, our findings show that qualifying performance (measured before any penalties or grid starting position adjustments) is the most reliable indicator of race results of any session over a race weekend. The previously reported dominance of starting grid position likely reflects an artifact of the limited sampling in the study rather than the genuinely strongest relationship.

4.2 Interpretation of Results

The dominance of qualifying results having the strongest relationship with finishing position in the race, across all statistical measures and weekend sessions, reflects the nature of F1 as a sport itself: one-lap performance, under controlled conditions, is the clearest demonstration of a car and driver's capability. By contrast, when you focus on just the starting-grid position, which has already incorporated post-session penalties (i.e., component changes, interference with other competitors, or other infractions), that only disguises the true competitive nature of all entries. This helps hypothesize why qualifying, and not grid placement itself, is the strongest and most stable association with race outcomes across a weekend.

Practice sessions help demonstrate the evolution of a car weekend, but they display weaker associations (vs qualifying) that continue to improve over the weekend (from P1 through P3). This pattern reflects how teams tend to focus on iteratively optimizing their car's setups (i.e., making mechanical and aerodynamic adjustments to a car to improve performance for the specific track conditions), with the final practice session, P3, serving as a bridge to qualifying. Although these sessions are primarily used for data collection and driver preparations, their statistical link to final race results underscores that weekend preparation still carries measurable performance implications.

4.3 Limitations

This study deliberately focuses on the observable session-level variables to isolate their statistical association with race results in order to compare and challenge existing published work. As such, this analysis does not incorporate the contextual or stochastic factors that can also impact the final placement of a race, such as weather variability, track type, pit strategy, tyre strategy, tyre degradation, safety-car interventions, in-race incidents, or the separation of driver and car skills. The OLR model explains a small proportion of total variance, consistent with the inherent stochasticity of F1 racing, but still indicates that important explanatory variables remain unmodelled. Additionally, this study assumes uniform competitive strength across seasons; however, in reality, Formula One varies in parity by year, with some seasons seeing one team dominate, while in other years, there is a more competitive balance. Eras of team dominance inflate session-race associations, while the more balanced grids introduce additional variability in results.

Additionally, this dataset selection excludes sprint weekends (which continue to grow over the years) and races with incomplete session data, which may slightly limit generalizability to these evolving race formats. While our approach addresses a key sampling flaw in prior work, it remains more of a correlational framework than a predictive or causal one. Despite these choices in this study, this work establishes a large-scale and statistically transparent replication of F1 performance associations using open-source data that can serve as a framework for future research.

4.4 Future Work

The findings presented in this work aim to establish and correct an empirical baseline for understanding session-to-race relationships in a Formula One weekend. By revisiting and extending earlier analysis with a full-grid, multi-season dataset, this study provides a more accurate foundation for quantifying how performance evolves across practice, qualifying, and race sessions.

Future research should continue to expand this framework by integrating additional covariates that capture the complex dynamics influencing race outcomes. Many of these are suggested in Section 4.3 and include weather conditions, track characteristics, tyre choice and strategies, reliability events, and pit-stop execution metrics. This effort can be complemented with session-level indicators such as in-race telemetry, sector and micro-sector performance, driver physiological data, and car-specific performance parameters to develop a more comprehensive explanatory framework extending toward predictive modelling.

To strengthen this progression toward predictive applications, future analysis should incorporate indicators of competitive balance within each season—such as point variance, team win or podium shares, or Elo-like strength ratings—to explicitly model field strength and adjust for variability in competitive parity across seasons. Predictive models developed on these enhanced datasets should also include alternative benchmarks, such as betting-market odds and out-of-sample validation, to ensure robustness and real-world calibration.

The introduction of sprint weekends since 2021 provides a valuable natural experiment to examine how reduced preparation time affects weekend performance dynamics. All of these richer datasets can help identify which factors most strongly determine whether faster cars successfully convert qualifying pace into race success and begin to separate the relative contributions of driver skill and car performance. By correcting the empirical foundation of earlier work, this paper establishes a benchmark for evaluating how new rules, technologies, and driver–team interactions shape competitive outcomes in Formula One, especially given the limited scope of prior academic research in this domain.

4.5 Conclusion

This study aims to replicate and extend prior research on Formula One performance and correct earlier misinterpretations arising from an intentionally truncated sample size, both in terms of the time span of the data and the participants in the races. Across nearly two decades of races and multiple technological eras, qualifying performance consistently emerges as the variable most strongly associated with final race outcomes, confirming it as the most accurate reflection of underlying car and driver capability ahead of the starting flag on race day.

By contrast, grid position, which is adjusted for penalties, dilutes this relationship and should be treated as the primary independent predictor of race performance. This work aims to shift the discussion from prediction to association, providing a statistically sound foundation for future modelling efforts to incorporate richer contextual variables and causal mechanics.

References

Agresti, A. (2010). *Analysis of ordinal categorical data*. John Wiley & Sons.

Bisagni, C. and Terletti, D. (2008). Structural optimisation of composite elements of a formula one racing car. *Journal of Composite Materials*, 42(5):471–490. Accessed January 25, 2025.

Fagerland, M. W. and Hosmer, D. W. (2016). Tests for goodness of fit in ordinal logistic regression models. *Journal of Statistical Computation and Simulation*, 86(15):3191–3208.

Fagerland, M. W. and Hosmer, D. W. (2017). How to test for goodness of fit in ordinal logistic regression models. *Stata Journal*, 17(3):668–686.

Fédération Internationale de l'Automobile (FIA) (2024). Sporting Regulations for Formula One. Accessed January 24, 2025.

Gertheiss et al. (2023). *gofcat: Goodness-of-Fit Tests for Categorical Response Regression Models*. R package version 0.1.0.

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American statistical association*, 49(268):732–764.

Hughes, D. E. (1968). Measurement in motor racing performance. *International Journal of Sports Science*, 8:285–292.

Lipsitz, S. R., Fitzmaurice, G. M., and Molenberghs, G. (1996). Goodness-of-fit tests for ordinal response regression models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(2):175–190.

McCarthy, L. M. and Rotthoff, K. W. (2013). Incentives on the starting grid in formula one racing. *The Journal of SPORT*, 2(2):175–184. Accessed January 25, 2025.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.

Mühlbauer, T. (2010). Relationship between starting and finishing position in formula one car races. *International Journal of Performance Analysis in Sport*, 10(2):99–105.

Padilla, N. (2023). A research study on formula one: Determining the effectiveness of rookie drivers. Accessed January 25, 2025.

Patil, A., Jain, N., Agrahari, R., Hossari, M., Orlandi, F., and Dev, S. (2022). A data-driven analysis of formula 1 car races outcome. In *Irish Conference on Artificial Intelligence and Cognitive Science*, pages 134–146. Springer.

Schwaberger, G. (1987). Physiological demands in formula one racing: A review. *British Journal of Sports Medicine*, 40(7):655–656.

Sicoie, H. (2022). Machine learning framework for formula 1 race winner and championship standings predictor. *Bachelor's thesis, Tilburg University, School of Humanities and Digital Sciences.*

Sidney, S. (1957). Nonparametric statistics for the behavioral sciences. *The Journal of Nervous and Mental Disease*, 125(3):497.

Singh, H. (2016). Multinomial and ordinal logistic regression in r. Accessed: 2025-05-04.

Sobrie, L. (2020). Sifting through the noise in formula one: Predictive performance of tree-based models. Master's thesis, Ghent University, Ghent, Belgium.

Sports, C. (2023). Formula 1 race strategy analysis: The role of data in securing victory. Accessed January 24, 2025.

UCLA Statistical Consulting Group (n.d.). R data analysis examples: Ordinal logistic regression. Accessed: 2025-05-04.

van Kesteren, E.-J. and Bergkamp, T. (2022). Bayesian analysis of formula one race results: Disentangling driver skill and constructor advantage. *Journal of Quantitative Analysis in Sports*, 18(2):85–104.