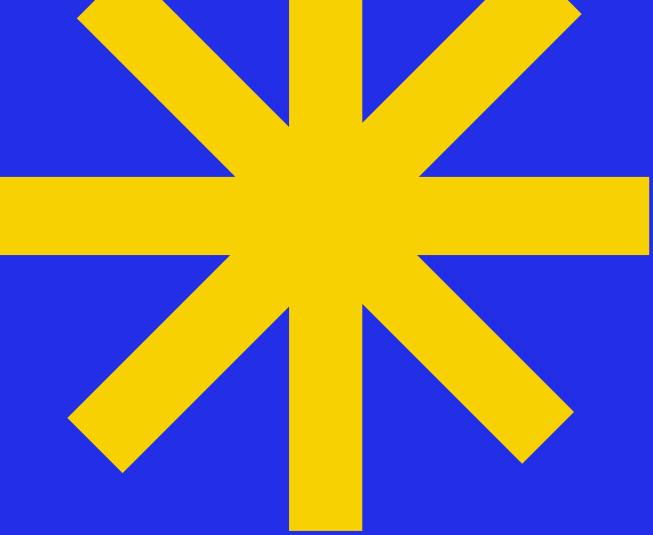


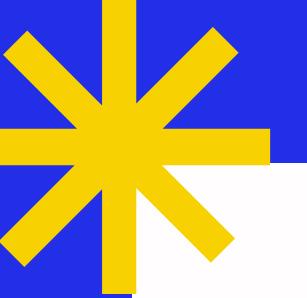
Statistics Sports

Research Question:

Can early performance predict the final position of a cyclist?

VALENTIN SIRLONGE





DATA OVERVIEW

Objective

The goal is to find out if the average position during the first 5 laps can predict the final ranking (lap 20) of female cyclists.

Variables used:

- **lap1 to lap5: position at each of the first 5 laps**
- **lap20: final position**
- **avg_lap1_5: average of laps 1 to 5**
- **avg_time_1_5: average of lap times in seconds**



final rider position	lap1	lap2	lap3	lap4	lap5	lap20	avg_lap1_5
1	3	1	4	2	2	1	2,4
2	5	8	8	7	3	2	6,2
3	30	30	31	30	30	3	30,2
4	2	5	7	5	4	4	4,6
5	4	3	2	6	7	5	4,4

final rider position	lap1	lap2	lap3	lap4	lap5	lap20	avg_time_1_5
1	115,14	116,467	112,054	114,248	118,182	100,696	115,2182
2	115,466	116,673	112,48	114,442	117,054	100,725	115,223
3	127,361	114,483	112,333	113,813	115,421	106,491	116,6822
4	113,773	118,197	112,567	114,19	117,585	101,892	115,2624
5	115,318	116,514	111,666	115,383	117,561	101,837	115,2884

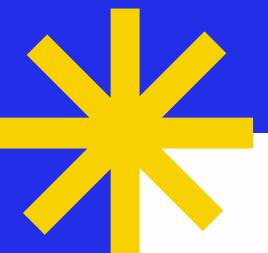


Race Position

Correlation

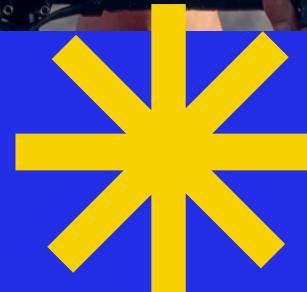
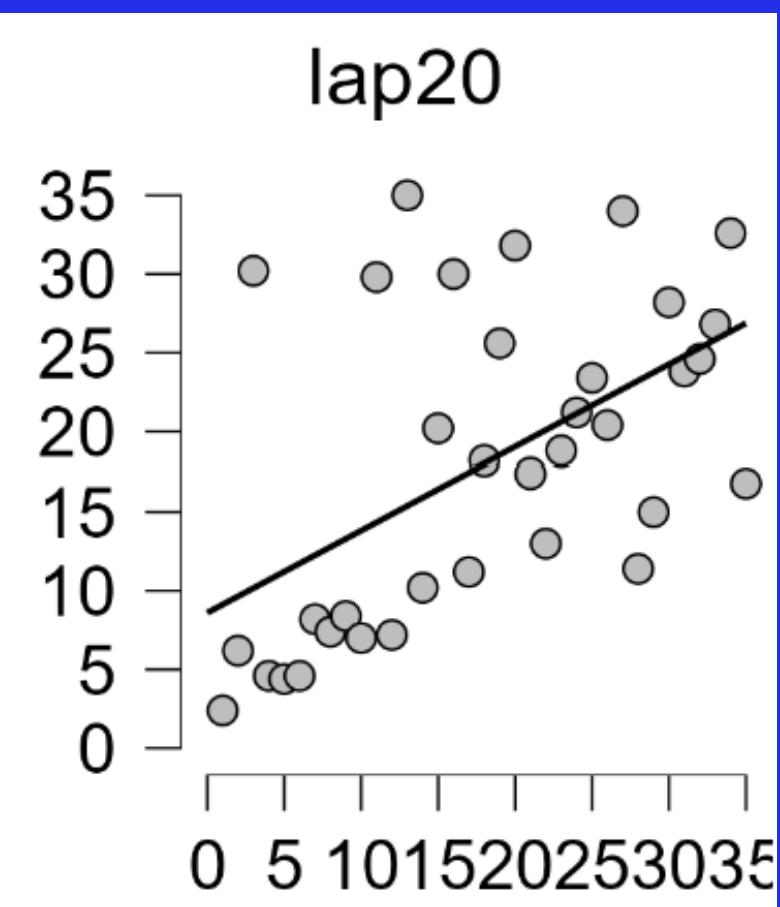
Pearson's Correlations

Variable	avg_lap1_5	lap20
1. avg_lap1_5	Pearson's r	—
	p-value	—
2. lap20	Pearson's r	0.535
	p-value	< .001



CORRELATION RESULTS

- Pearson's r = 0.535
- p-value < .001 (statistically significant)
- ✓ Moderate positive correlation between early position and final result
- + Cyclists who start strong tend to finish in better positions



Race Position

Linear Regression

Model Summary - lap20

Model	R	R ²	Adjusted R ²	RMSE
M ₀	0.000	0.000	0.000	10.247
M ₁	0.535	0.286	0.264	8.789

Note. M₁ includes avg_lap1_5

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
M ₁	Regression	1021.079	1	1021.079	13.220	< .001
	Residual	2548.921	33	77.240		
	Total	3570.000	34			

Note. M₁ includes avg_lap1_5

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
M ₀	(Intercept)	18.000	1.732		10.392	< .001
M ₁	(Intercept)	8.135	3.093		2.630	0.013
	avg_lap1_5	0.548	0.151	0.535	3.636	< .001

LINEAR REGRESSION RESULTS

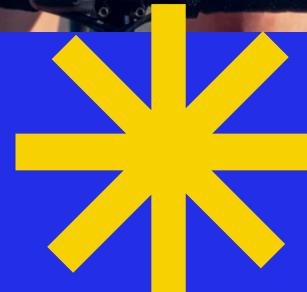
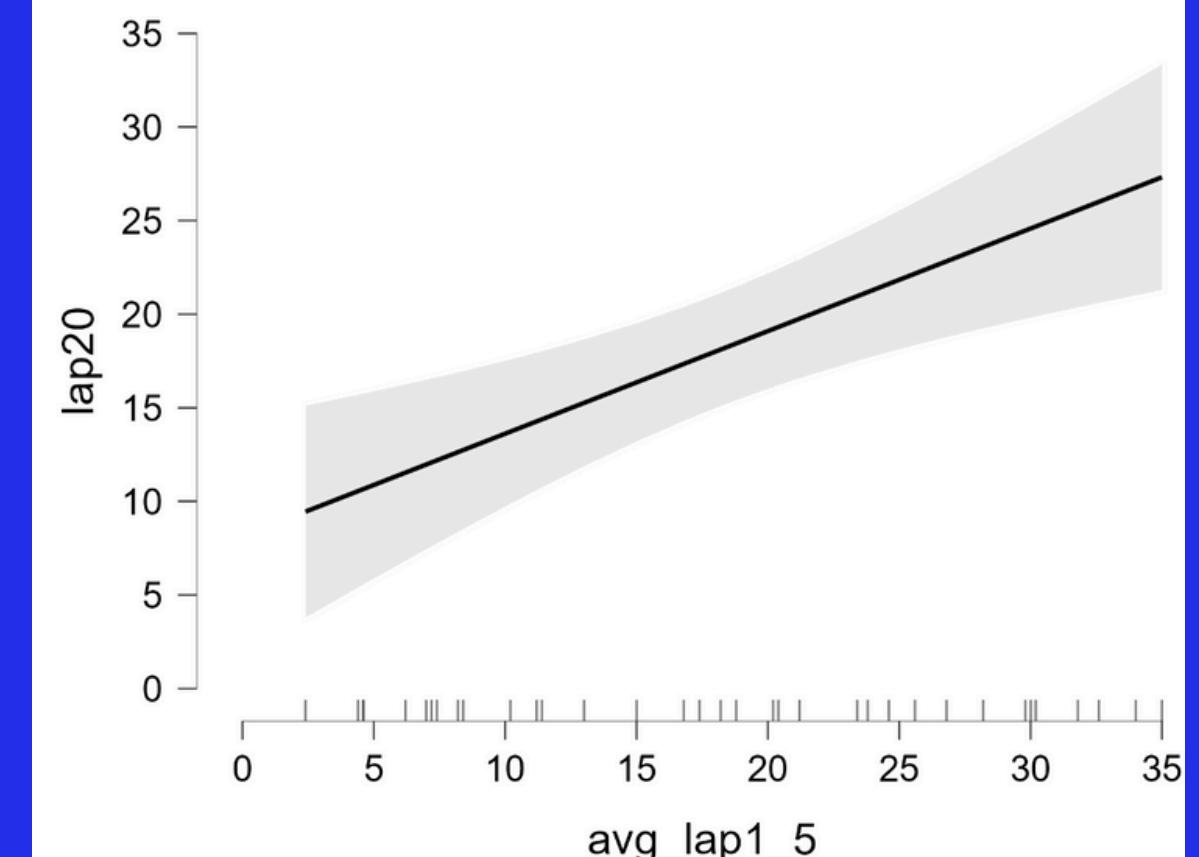
- $R^2 = 0.286$ (28.6% of the variance explained)
- $p < .001 \rightarrow$ significant model
- Coefficient (B) = 0.548

Interpretation:

The worse the average position in the first 5 laps, the worse the final ranking

Marginal Effects Plots

Marginal effect of avg_lap1_5 on lap20



Lap Times

Correlation

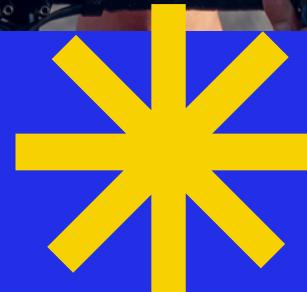
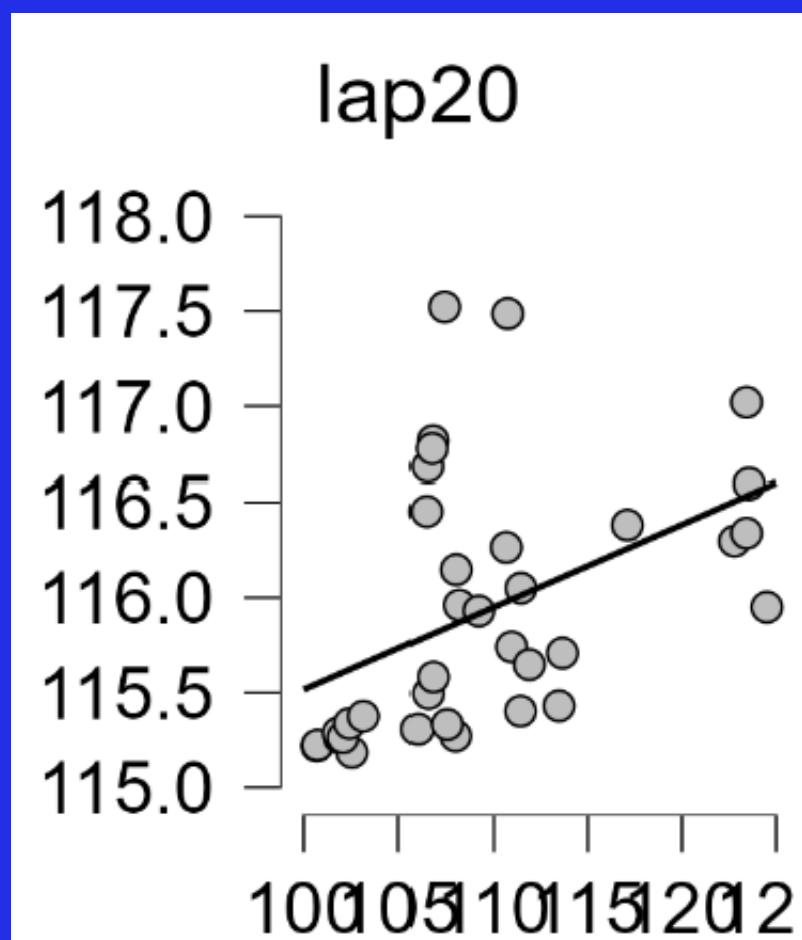
Pearson's Correlations

Variable	avg_time_1_5	lap20
1. avg_time_1_5	Pearson's r	—
	p-value	—
2. lap20	Pearson's r	0.449
	p-value	0.007



CORRELATION RESULTS

- Pearson's r = 0.449
- p-value = .007 (significant)
- ✓ Cyclists who are faster in the early laps tend to have a faster lap20 time



Lap Times

Linear Regression

Model Summary - lap20

Model	R	R ²	Adjusted R ²	RMSE
M ₀	0.000	0.000	0.000	6.938
M ₁	0.449	0.201	0.177	6.294

Note. M₁ includes avg_time_1_5

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
M ₁	Regression	329.494	1	329.494	8.317	0.007
	Residual	1307.311	33	39.615		
	Total	1636.805	34			

Note. M₁ includes avg_time_1_5

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

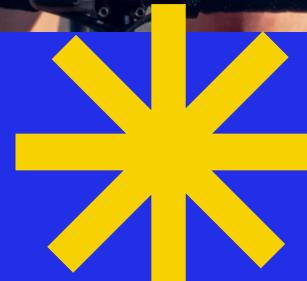
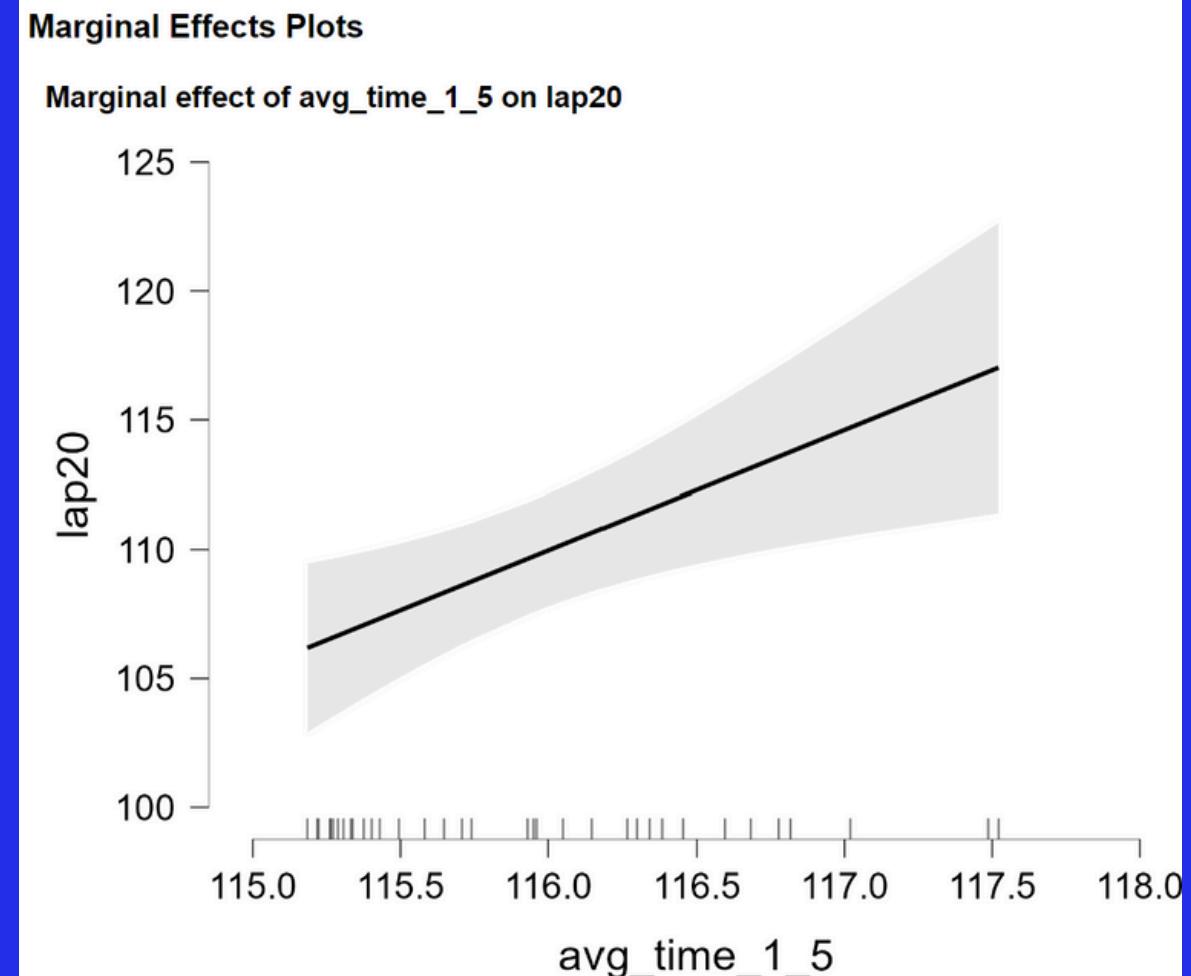
Model		Unstandardized	Standard Error	Standardized	t	p
M ₀	(Intercept)	109.666	1.173		93.508	< .001
M ₁	(Intercept)	-427.977	186.428		-2.296	0.028
	avg_time_1_5	4.637	1.608	0.449	2.884	0.007

LINEAR REGRESSION RESULTS

- R² = 0.201 (20.1% of the variance explained)
- p = .007 (significant model)
- Coefficient (B) = 4.637

Interpretation:

Slower average early laps lead to slower final lap time





Conclusion

- Early performance is a good predictor of final outcomes
- Both race positions and lap times show significant positive correlations
- Cyclists who perform well at the beginning tend to maintain their performance
- Useful insights for training, strategy, and pacing in races

