Análisis de casos en inferencia estadística.

Mitsiu Alejandro Carreño Sarabia
E23S-18014

Tire Changes, Fresh Air, and Yellow Flags: Challenges in Predictive Analytics for Professional Racing

Theja Tulabandhula and Cynthia Rudin

Objetivo

Diseñar un sistema de predicción y apoyo en la toma de decisiones en tiempo real en carreras automovilísticas profesionales.

La investigación parte de las siguientes preguntas:

1. ¿Es posible predecir el cambio de posición en carrera basado en carreras recientes del corredor?
2. ¿Es posible optimizar el cambio de neumáticos y repostaje (carga de gasolina) basado en la predicción del rendimiento del coche?
3. ¿Es posible obtener conocimiento de carreras pasadas que asistan al equipo en carreras futuras?

Contexto

Durante la carrera un coche puede **repostar y cambiar** dos, cuatro o ningún neumático, sabiendo que cambiar más neumáticos **requiere más tiempo**, pero neumaticos nuevos mejoran el rendimiento del coche.

La toma de decisiones requiere compensación y caracterización del **riesgo-beneficio.** En el que se deben tomar en cuenta factores como cambio de **neumáticos de oponentes, la disposición de la pista, temperatura, clima**, etc.

Contexto

A diferencia del análisis de otros deportes, en las carreras, la historia de la carrera **es difícilmente segmentada,** en cada punto de la carrera la historia de la carrera completa determina la posición actual del corredor, una **cadena de markov.**

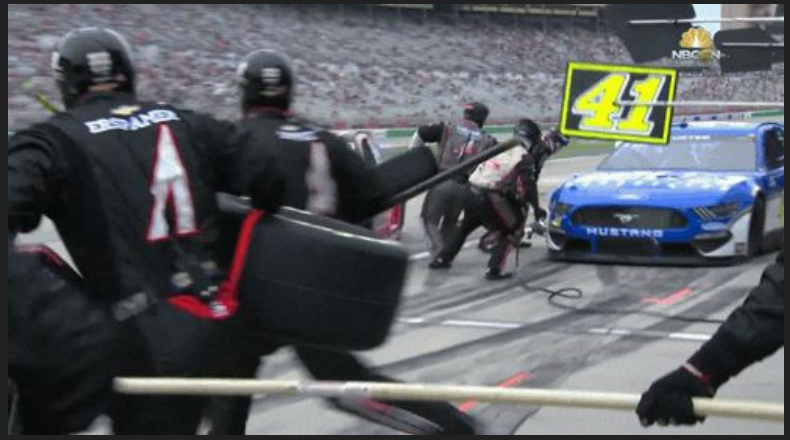
Para el estudio se evaluaron 119,178 tiempos de vuelta, incluyendo posición, estado de carrera (condición normal, precaución), 2,932 pasos por pits y cantidad de neumáticos cambiados.

Contexto

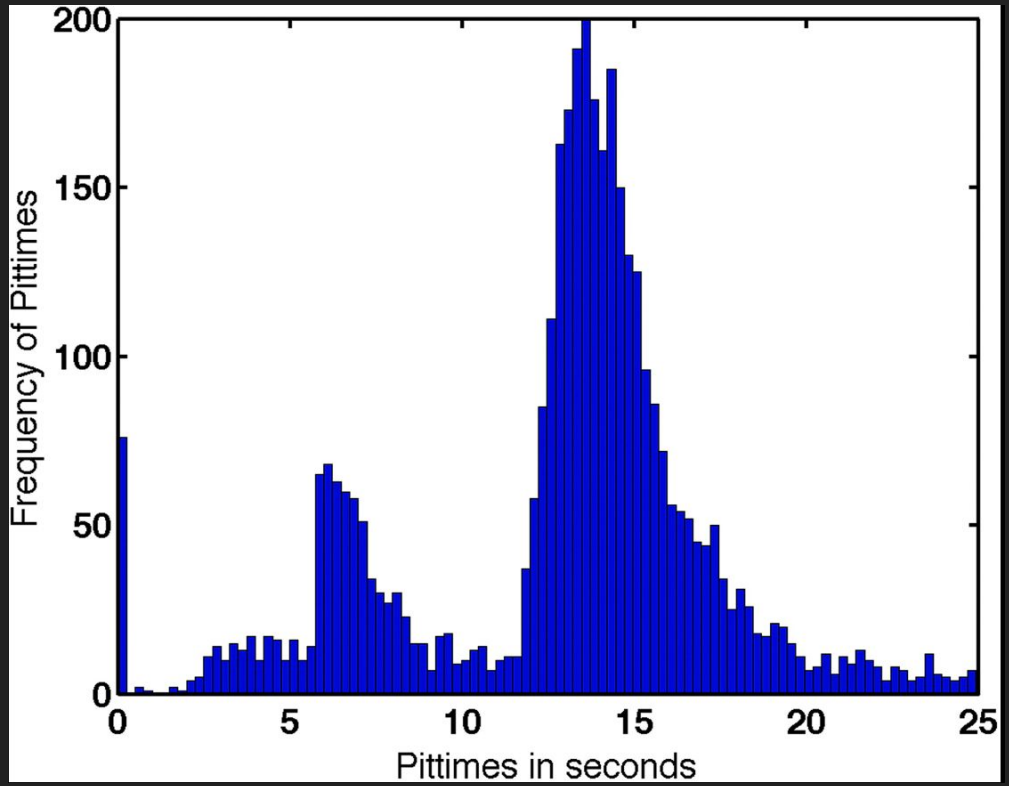
Cambiar dos neumáticos usualmente es 6 segundos más rápido que cambiar cuatro.

Al entrar a pits se tiene una velocidad limitada y la calle de pits tiene una longitud variable.

Finalmente las acciones de los competidores afectan directamente la estratégia y el tiempo de pits.



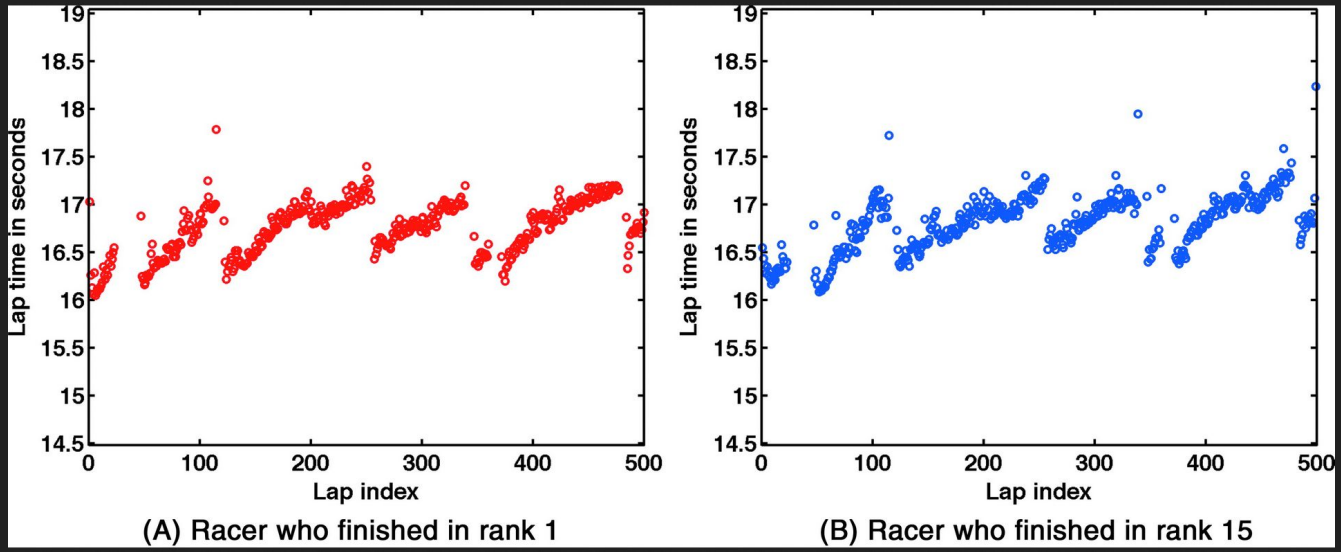
Análisis



Se muestra la distribución de tiempos en las paradas, se notan 3 picos en 0 segundos (penalizaciones), 6 segundos (2 neumáticos) y 15 segundos (4 neumaticos)

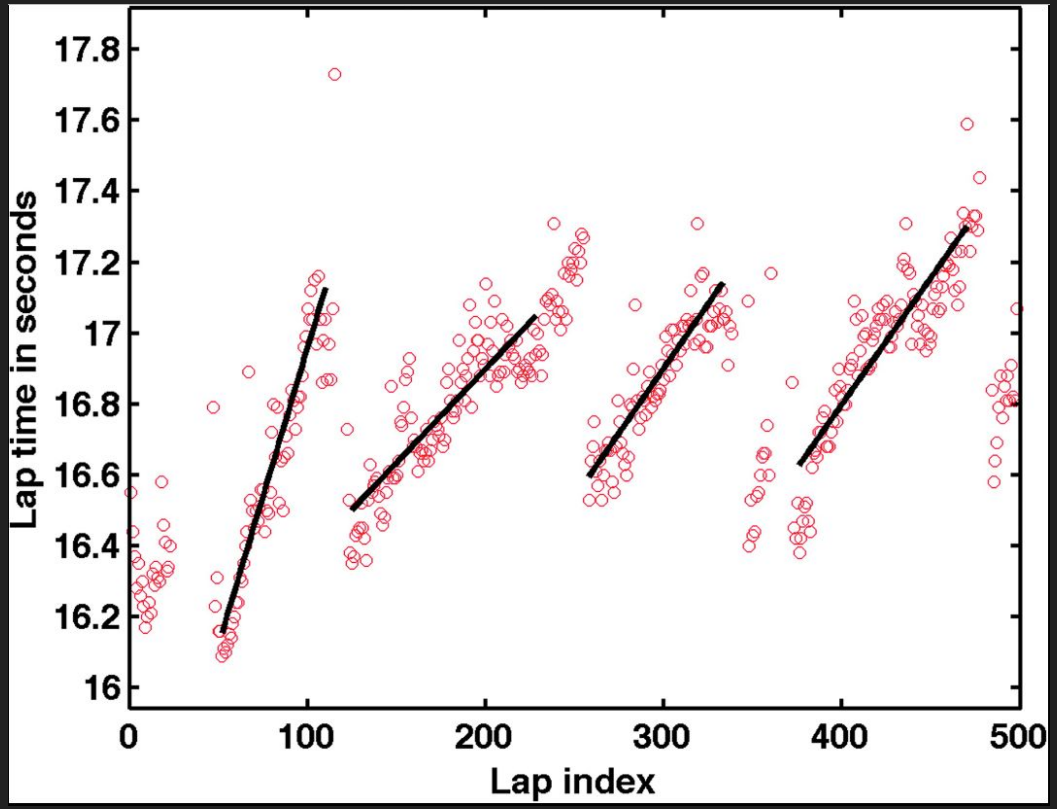
Sería interesante realizar un análisis de varianza segmentando por cantidad de neumáticos cambiados, para conocer la **distribución de probabilidad** y la tendencia a tardar más.

Análisis



Tiempo por vuelta de primer lugar (rojo) vs quinceavo lugar (azul), en el que se nota la pérdida de rendimiento relacionado al desgaste del neumático e inverso a la pérdida de combustible.

Análisis

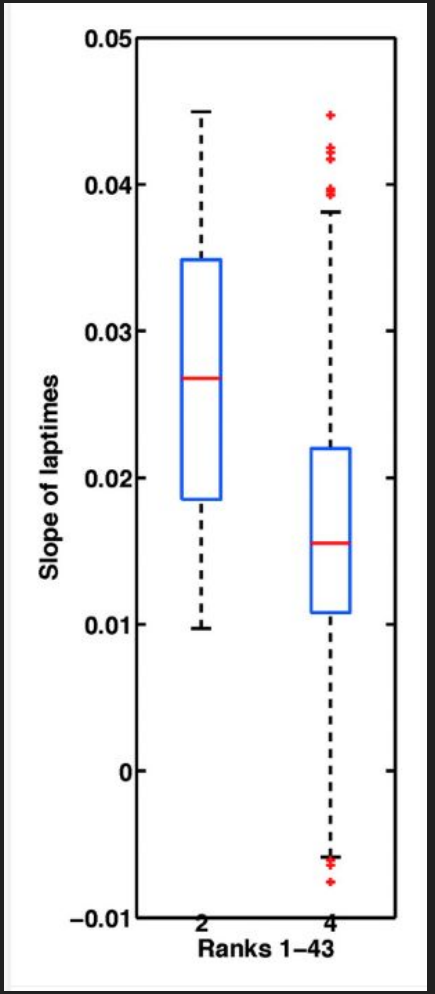


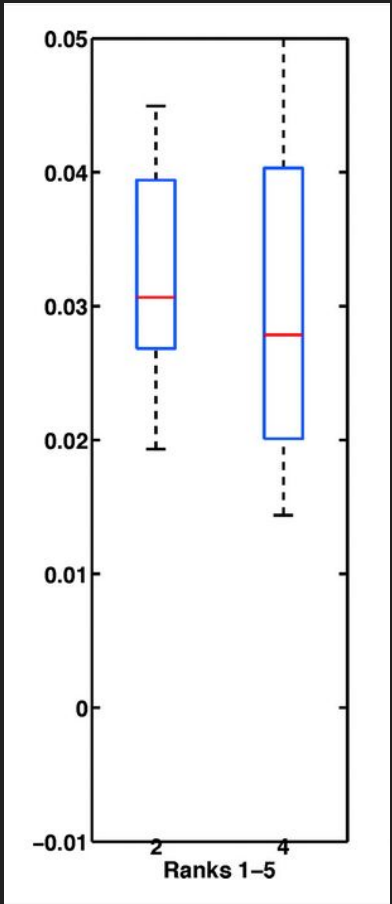
Detalle de carrera del quinceavo lugar aplicando regresión lineal simple.

Se pueden apreciar líneas más verticales, debido a que su tiempo no solo está condicionado al rendimiento del coche sino al de los coches que le preceden

Paradoja de Simpson

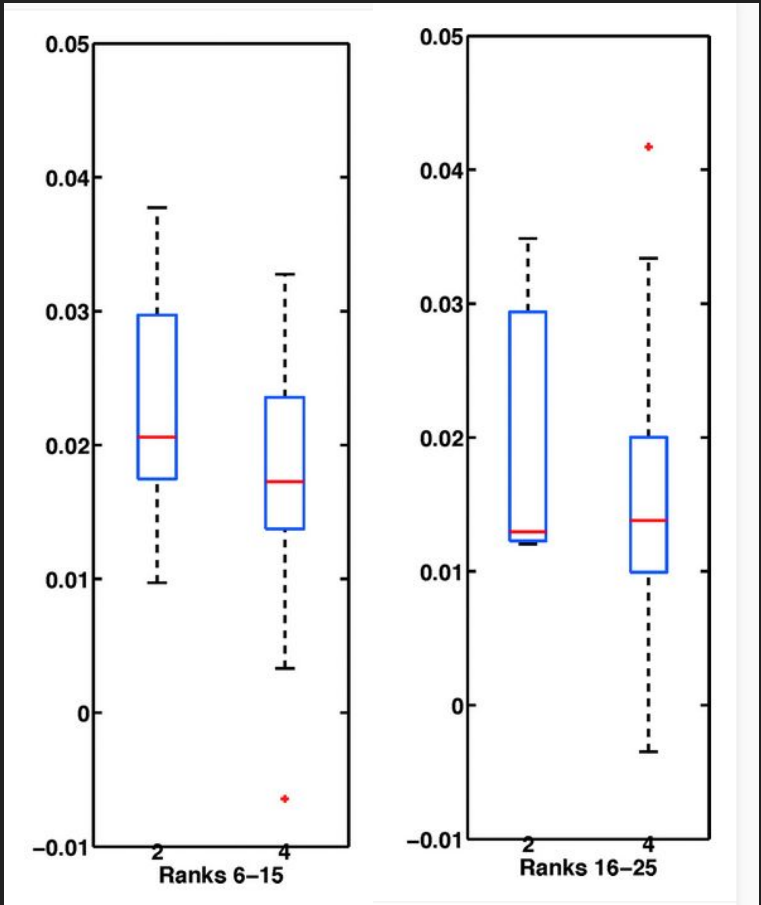
Para una carrera específica, y considerando todos los corredores, los coches que cambiaron dos neumáticos tuvieron una pérdida de rendimiento (mayor pendiente en los tiempos de vuelta) mayor, por lo tanto fue mejor estratégia cambiar 4 neumaticos (y tener menos pendiente en los tiempos de vuelta).





Paradoja de Simpson

Segmentando por el top 5, se nota que las pendientes son similares entre cambiar 2 y 4 neumáticos.



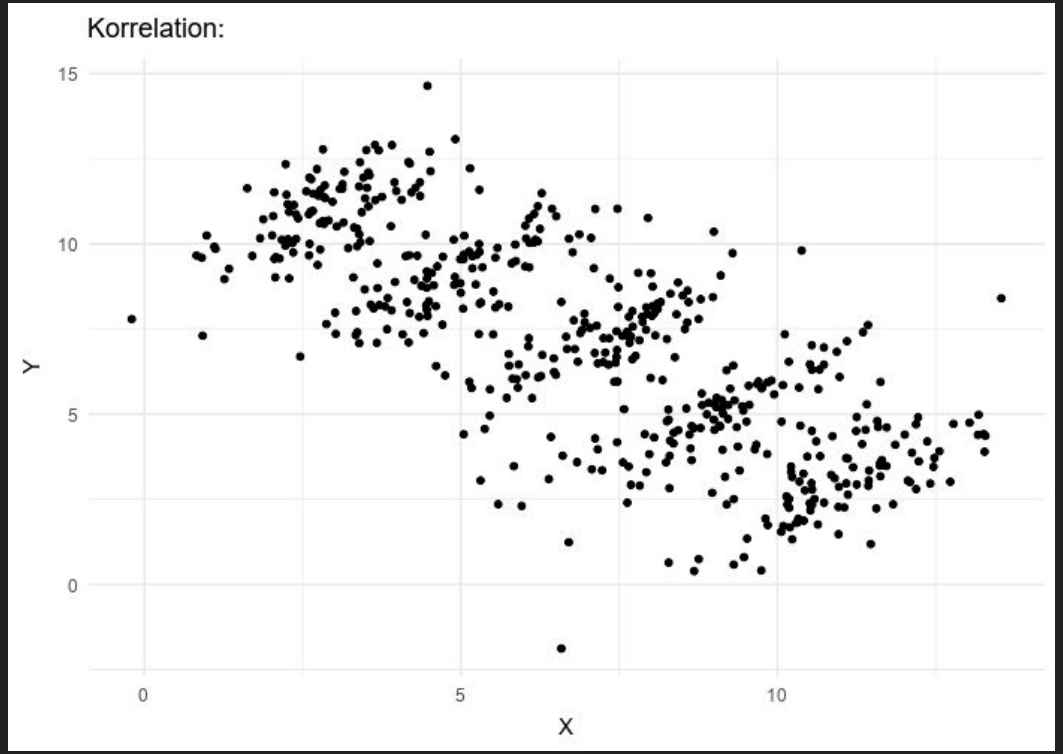
Paradoja de Simpson

Segmentando del 6 al 15, se nota que las pendientes son similares entre cambiar 2 y 4 neumáticos de nuevo.

Igual que entre el lugar 16 al 25.

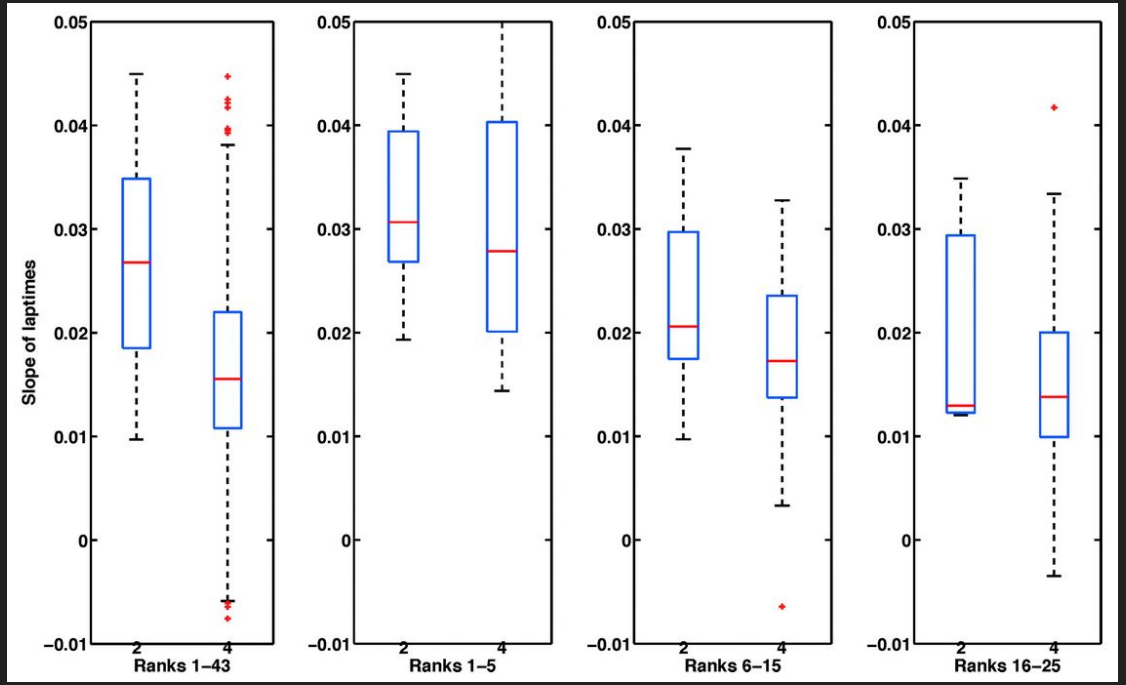
Paradoja de Simpson

Existe una tendencia aparente en un conjunto de grupos, pero desaparece o se revierte cuando se toma la totalidad de los datos.

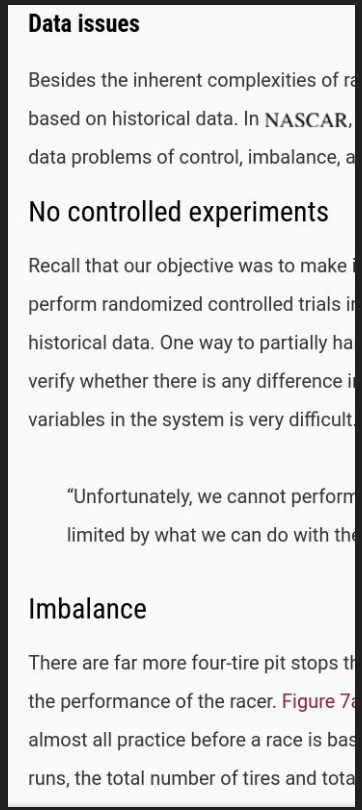


Análisis personal

El ritmo está condicionado a la posición actual y la capacidad de rebasar, si no hay capacidad de rebase, nulifica las prestaciones del neumático. (Undercut)



Análisis personal



Data issues

Besides the inherent complexities of ra
based on historical data. In NASCAR,
data problems of control, imbalance, a

No controlled experiments

Recall that our objective was to make i
perform randomized controlled trials in
historical data. One way to partially ha
verify whether there is any difference i
variables in the system is very difficult.

"Unfortunately, we cannot perform
limited by what we can do with the

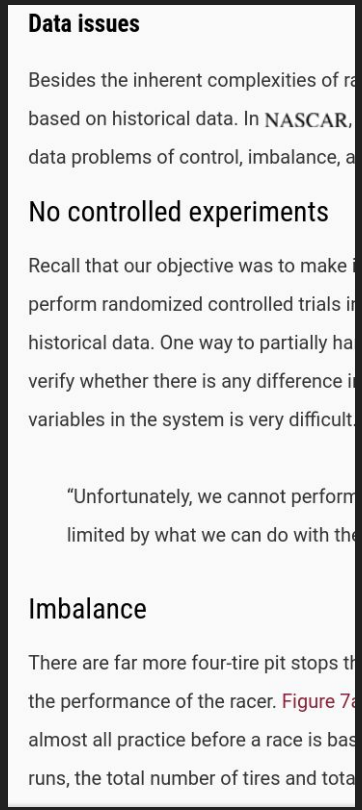
Imbalance

There are far more four-tire pit stops th
the performance of the racer. Figure 7a
almost all practice before a race is bas
runs, the total number of tires and tota

En la investigación enfatizan en tener el problema de **sólo contar con los datos históricos** y **notar tendencia en cambio de 4 neumáticos**.

Creo que este es un caso en el que aplicar t**écnicas de remuestreo** así como evaluaciones de error de muestreo, y **nivel de confianza** son útiles.

Análisis personal



Data issues

Besides the inherent complexities of ra
based on historical data. In NASCAR,
data problems of control, imbalance, a

No controlled experiments

Recall that our objective was to make i
perform randomized controlled trials in
historical data. One way to partially ha
verify whether there is any difference i
variables in the system is very difficult.

> "Unfortunately, we cannot perform
> limited by what we can do with the

Imbalance

There are far more four-tire pit stops th
the performance of the racer. Figure 7a
almost all practice before a race is bas
runs, the total number of tires and tota

En la investigación en lugar de remuestrear decidieron agrupar por circuitos con características similares, lo cual también me parece apropiado, pero abre el debate a cómo evaluar la similitud entre circuitos.

Gracias