

Estimación causal del impacto de la localía y la presencia de público en los resultados de partidos de fútbol de Argentina

El presente proyecto de investigación tiene como objetivo principal examinar el efecto causal de la localía y la presencia de público en los resultados de los partidos de fútbol. Para ello, utilizaremos un conjunto de datos que contiene información sobre los resultados de una liga de fútbol argentino, incluyendo el resultado de los partidos, la localidad y la fecha en la que se llevaron a cabo. El dataset incluye partidos jugados durante la pandemia de COVID-19, que se llevaron a cabo sin público así como partidos posteriores y anteriores a la pandemia, con espectadores.

Palabras clave: inferencia causal, fútbol, estadística, reproducibilidad

Conocimientos deseables

Estadística

¿Qué podría aprender quien realice esta tesis?

Durante el desarrollo de este proyecto de licenciatura en ciencia de datos, el estudiante aprenderá varias habilidades y conceptos en el campo de la investigación causal y el análisis estadístico. Aprenderá a aplicar técnicas estadísticas para controlar los sesgos y crear grupos de tratamiento y control comparables, lo que permitirá realizar una inferencia causal sólida. Además, adquirirá experiencia en el análisis descriptivo de datos, examinando las características de los partidos de fútbol en diferentes condiciones. También aprenderá a utilizar modelos de regresión adecuados, como el modelo logístico, para estimar el efecto causal de la localía y la presencia de público en los resultados de los partidos.

Dirección de la tesis

*Solovey, Guillermo
Instituto de Cálculo*

Contacto: gsolovey@gmail.com

Más información en el pdf a continuación.

Título del proyecto:**Estimación causal del impacto de la localía y la presencia de público en los resultados de partidos de fútbol de Argentina****Introducción:**

El presente proyecto de investigación tiene como objetivo principal examinar el efecto causal de la localía y la presencia de público en los resultados de los partidos de fútbol. Para ello, utilizaremos un conjunto de datos que contiene información sobre los resultados de una liga de fútbol argentino, incluyendo el resultado de los partidos, la localidad y la fecha en la que se llevaron a cabo. El dataset incluye partidos jugados durante la pandemia de COVID-19, que se llevaron a cabo sin público así como partidos posteriores y anteriores a la pandemia, con espectadores.

Metodología:

Para estimar el efecto causal de la localía y la presencia de público en la probabilidad de ganar un partido, utilizaremos el marco de la inferencia causal. Aplicaremos técnicas estadísticas para controlar los sesgos potenciales inherentes a los datos observacionales. Este enfoque nos permitirá crear grupos de tratamiento y control comparables, lo que nos permitirá realizar una inferencia causal sólida. El primer paso será realizar un análisis descriptivo de los datos, examinando las características de los partidos jugados en condiciones de localía y visitante, así como los partidos con y sin público. A continuación, crearemos un grupo de comparación adecuado, emparejando los partidos en función de variables relevantes, como la calidad de los equipos, la posición en la tabla de clasificación, entre otras. Posteriormente, estimaremos el efecto causal de la localía y la presencia de público en la probabilidad de ganar un partido utilizando modelos de regresión adecuados, como el modelo logístico. También consideraremos la inclusión de variables de control adicionales, como el desempeño previo del equipo, la diferencia de goles, entre otras, para obtener estimaciones más precisas.

Resultados esperados:

Esperamos que los resultados de este estudio proporcionen una estimación cuantitativa del impacto causal de la localía y la presencia de público en los resultados de los partidos de fútbol. Estas estimaciones podrían tener implicaciones significativas en la comprensión de los factores que influyen en el rendimiento de los equipos y en el diseño de estrategias futuras. Además, esta investigación podría contribuir al debate actual sobre la influencia de la audiencia y el entorno en el rendimiento deportivo.

Herramientas y lenguaje de programación:

Este proyecto se llevará a cabo utilizando el lenguaje de programación R, aprovechando las diversas librerías y funciones disponibles para el análisis estadístico y la construcción de modelos.

Tiempo de ejecución:

Este proyecto se llevará a cabo en un período de 6 meses, con un seguimiento continuo y reuniones regulares para supervisar el progreso y brindar asistencia a los estudiantes.

Referencias:

- Cross, J., & Uhrig, R. (2023). Do fans impact sports outcomes? A COVID-19 natural experiment. *Journal of Sports Economics*, 24(1), 3-27.
<https://doi.org/10.1177/15270025221100204>
- Cunningham, S. (2021). *Causal inference: The mixtape*. Yale university press.