

Causal Inference UConn Sports Analytics Symposium

Kevin Cummiskey

2024-04-12

For this handout and some R code, please go here:



github user: kfcaby repo: UCSAS_causal_inference_workshop

Introduction to Causal Inference

1. What is the goal of causal inference?

2. How is causation different from correlation (i.e. an association)?

Table 1—Examples of Tasks Conducted by Data Scientists Working with Electronic Health Records

	Description	Data Science Task	
		Prediction	Causal inference
Example of scientific question	How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?	What is the probability of having a stroke next year for women with certain characteristics?	Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?
Data	<ul style="list-style-type: none"> • Eligibility criteria • Features (symptoms, clinical parameters ...) 	<ul style="list-style-type: none"> • Eligibility criteria • Output (diagnosis of stroke over the next year) • Inputs (age, blood pressure, history of stroke, diabetes at baseline) 	<ul style="list-style-type: none"> • Eligibility criteria • Outcome (diagnosis of stroke over the next year) • Treatment (initiation of statins at baseline) • Confounders • Effect modifiers (optional)
Examples of analytics	Cluster analysis ...	Regression Decision trees Random forests Support vector machines Neural networks ...	Regression Matching Inverse probability weighting G-formula G-estimation Instrumental variable estimation ...

Figure 1: Hernan, Hse, and Healy 2019

Fundamentals of Causal Inference

Traditional conditional probability notation cannot distinguish between causation and association. We need to introduce different notation to express causation.

Counterfactuals and the Potential Outcomes Framework

Confounding

Causal Diagrams

1. Confounding

2. Collider

Assumptions of Causal Inference

- SUTVA (Stable Unit Treatment Value Assumption)

- Ignorability

- Positivity

Methods for Causal Inference

Matching

Propensity Score Matching

Stratification on the Propensity Score

Inverse Probability of Treatment Weighting

Covariate Adjustment Using the Propensity Score

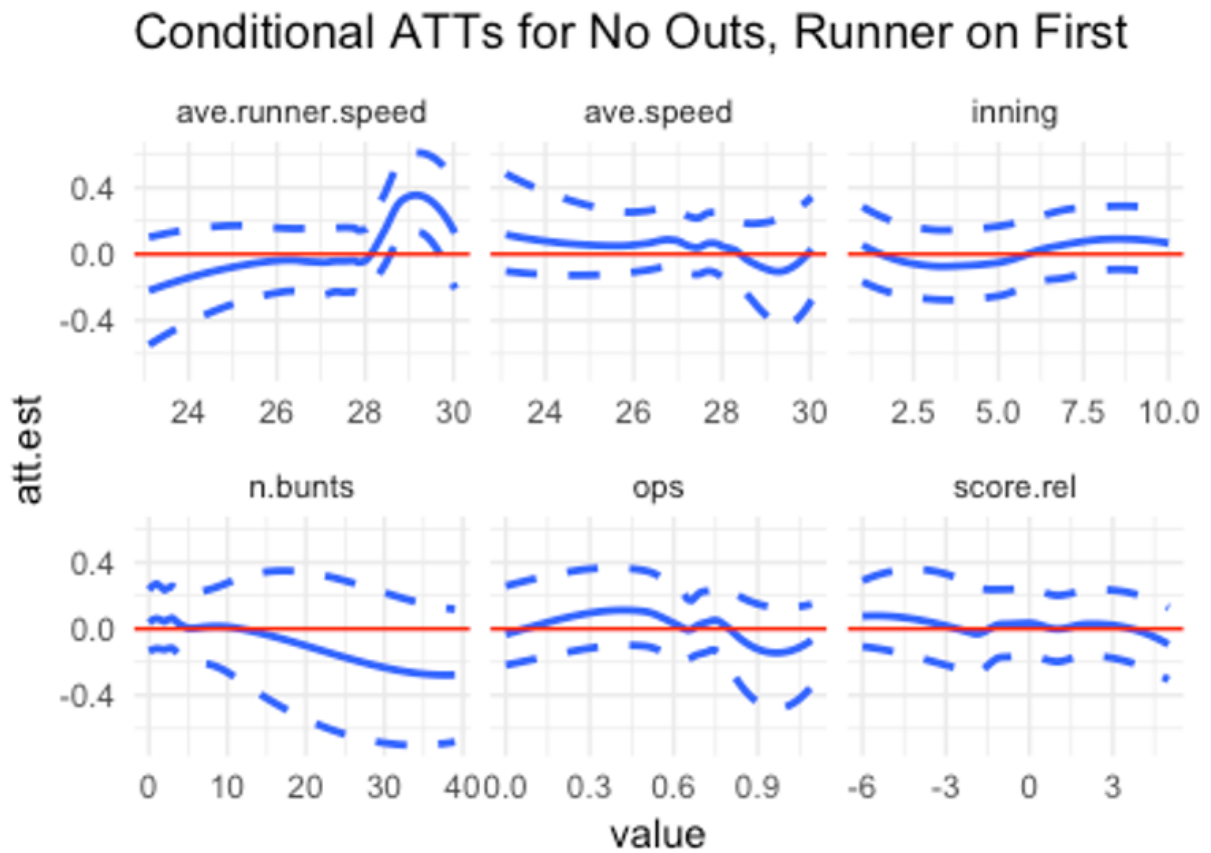
Applications

Estimating how the bunting effect varies

Authors: Katherine Evans and Michael Lopez

Title: Treatment Effect Heterogeneity in MLB Bunting Strategies

Where: JSM 2019



Player effects using location data

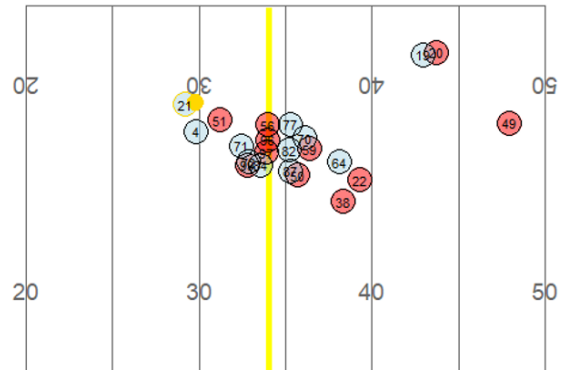
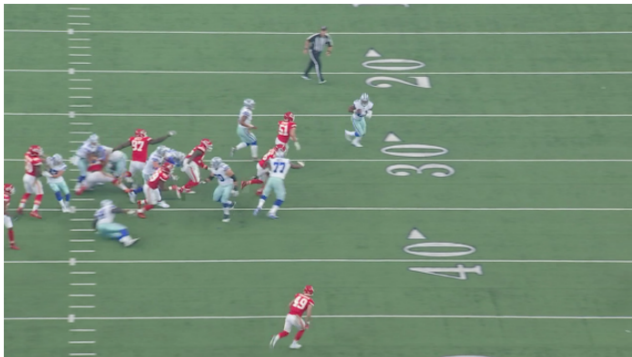
Author: Michael Lopez

Title: Building Blocks for Estimating Causal Effects of Athlete Behavior in Football and Hockey Using Player Tracking Data

Where: JSM 2019

Using player location data, we can better estimate player effects by comparing players in similar situations.

See also http://www.lukebornn.com/papers/fernandez_ssac_2018.pdf by Fernandez and Bornn for soccer.

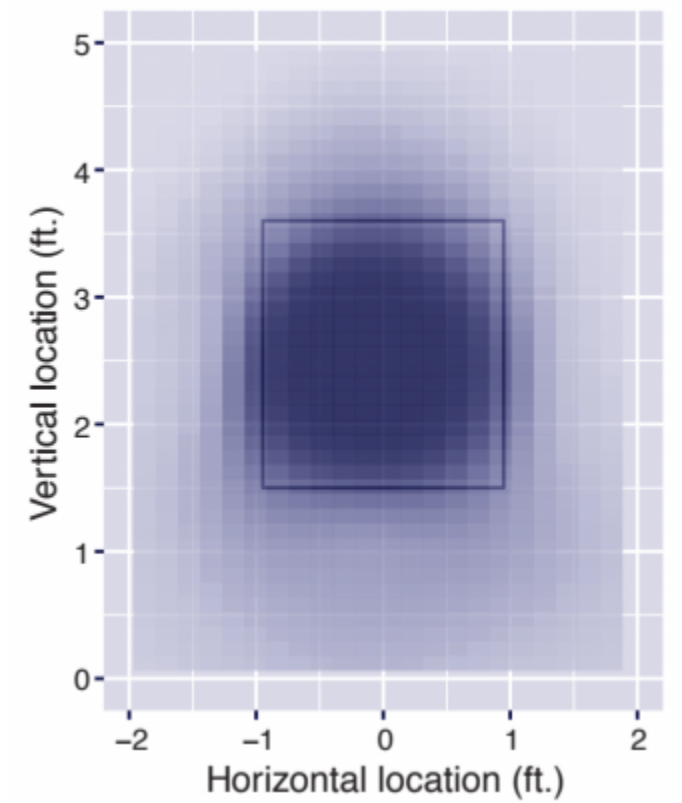


Catcher framing

Authors: Marchi, Albert, Baumer

Title: Analyzing Baseball Data with R

Where: Chapter 7



Player rewards

Authors: Romain Gauriot and Lionel Page

Title: Fooled by performance randomness: over-rewarding luck

Where: Review of Economics and Statistics, 2019

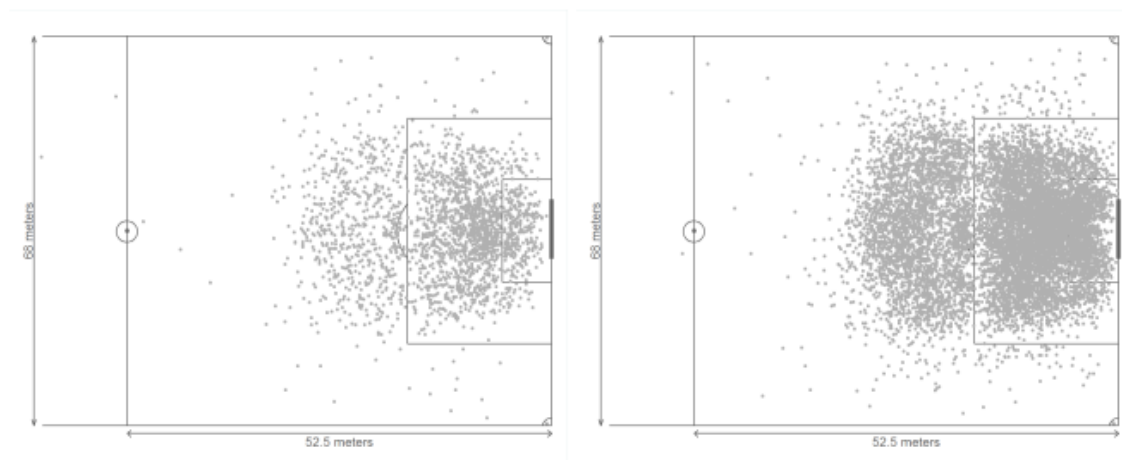


Figure 1: Graphical representation of the starting point of shots ending on the posts. On the left panel the post *in* ($N = 2,387$) and on the right the post *out* ($N = 10,679$) .

References

- Austin, P.C., 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3), pp.399-424.
- Chesnaye, N. C., Stel, V. S., Tripepi, G., Dekker, F. W., Fu, E. L., Zoccali, C., & Jager, K. J. (2021). An introduction to inverse probability of treatment weighting in observational research. *Clinical kidney journal*, 15(1), 14–20. <https://doi.org/10.1093/ckj/sfab158>
- Cummiskey, K., Villanti, L., and Crofford, I. (2024). Bunting and the ghost runner: a causal inference approach. <https://arxiv.org/abs/2404.06587>
- Hernán MA, Robins JM (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.
- Hernán, M. A., Hsu, J., & Healy, B. (2019). A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE*, 32(1), 42–49. <https://doi.org/10.1080/09332480.2019.1579578>