École polytechnique fédérale de Lausanne

Regression Methods Project Report

Coin-Data Regression Study

Authors:
Tara Fjellman
Rayan Harfouche

Professor:
Anthony Davison

Contents

Abstract

- context of the original paper (with their claims)
- our addition/contribution/comment about it

1 Introduction

Before diving into the analysis, we provide a brief overview of the datasets main features and the models we consider. Our exploratory analysis is available as a Jupyter notebook, and provides more detail.

The dataset is composed of throws from 48 people using 44 coins in total. Given the study did not impose strict guidelines for coins to be used, the design is heavily unbalanced. Eighteen coins have only been thrown by a single person, while some of them have been thrown by more than 20 people. Also, more than half the people have only flipped 5 or less different coins, while someone threw 11 different ones. As for the the person-coin pairs, most have fewer or equal to 1000 throws, while some have around 10000 ones. This severe unbalance must be kept in mind during the study, given it can pose challenges during model interpretation.

After plotting the same-side rates across people, coin and person-coin combinations, we deemed it relevant to investigate models considering both person and coin as covariates, as well as individual person-coin pairs. Following the advice given on the project statement, we also branched our analysis in Binomial-response GLM and WLS approaches.

Motivated by impacts of muscle-memory on the flipping, we additionally investigated aspects such as time-varying same-side rates and memory between successive throws.

2 Analysis

2.1 Model Comparison

In this section, we introduce and compare different models for the same side success rate.
Some GLMs with binomial responses and some WLS ones based on the ... approximation.
Should explain no a priori response transformation ...
For each, the considered formulas in terms of the covariates are:

- `1`, corresponding to a constant model.
- `1+C(person)`, corresponding to a model with the person as a covariate.
- `1+C(person)+C(coin)`, corresponding to a model with the person and the coin as covariates.
- `1+C(person)+C(coin)+C(person):C(coin)`, corresponding to a model with the person, the coin, and the interaction between the person and the coin as covariates.

* model 4 could seem redundant due to nesting-main effect, but we .... *
Should explain why eliminated some covariates ...

2.1.1 Tools For Selection

- when to use AIC vs LRT ?
- citing a few things about LRT not miting overfitting
-

2.1.2 WLS Approach

In this section, we consider the normal approximation for the binomial variable R with denominator m. Having that the success probability is fairly close to 0.5 we get $p(1-p)$ approximately equal to $1/4$. Hence we end up with: $R/m \sim N(p, 1/(4m))$.

We then make a linear fit using weighted least squares, where weight associated to each entry is proportional to the inverse of the variance, that only depending on m.

We consider different models, starting with a constant model, then adding in the following order the

Tab 1: *Model comparison based on AIC values.*

Model Formula (RHS)	AIC
1	159.76
1 + C(person)	12.89
1 + C(person) + agg	0.00
1 + C(person) + agg + C(coin)	29.21
1 + C(person) + agg + C(person):C(coin)	123.98

2.1.3 GLM Approach

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. Lorem ipsum dolor sit amet, consectetuer adipiscing elit.

Tab 2: *Model comparison for different models.*

Model	Deviance	AIC	Model DF
`1`	3943.48	173.97	0
`1+C(person)`	3677.51	0.00	46
`1+C(person)+C(coin)`	3611.12	17.61	88

Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien

est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. Lorem ipsum dolor sit amet,

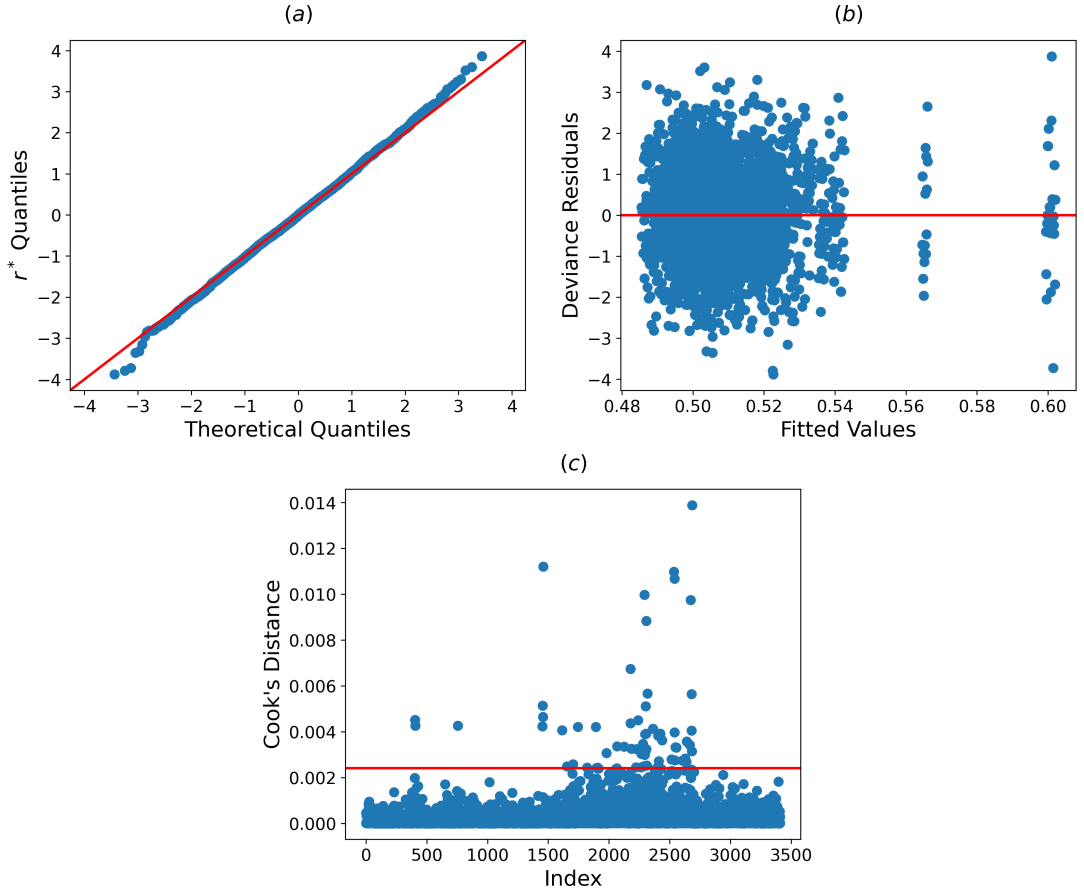

Fig 1: *Diagnostics for the selected GLM model. (a).*

consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Tab 3: *Likelihood ratio tests between models.*

Tested model	Restricted model	p-value
`1+C(person)`	`1`	0.00e+00
`1+C(person)+C(coin)`	`1+C(person)`	9.61e-03

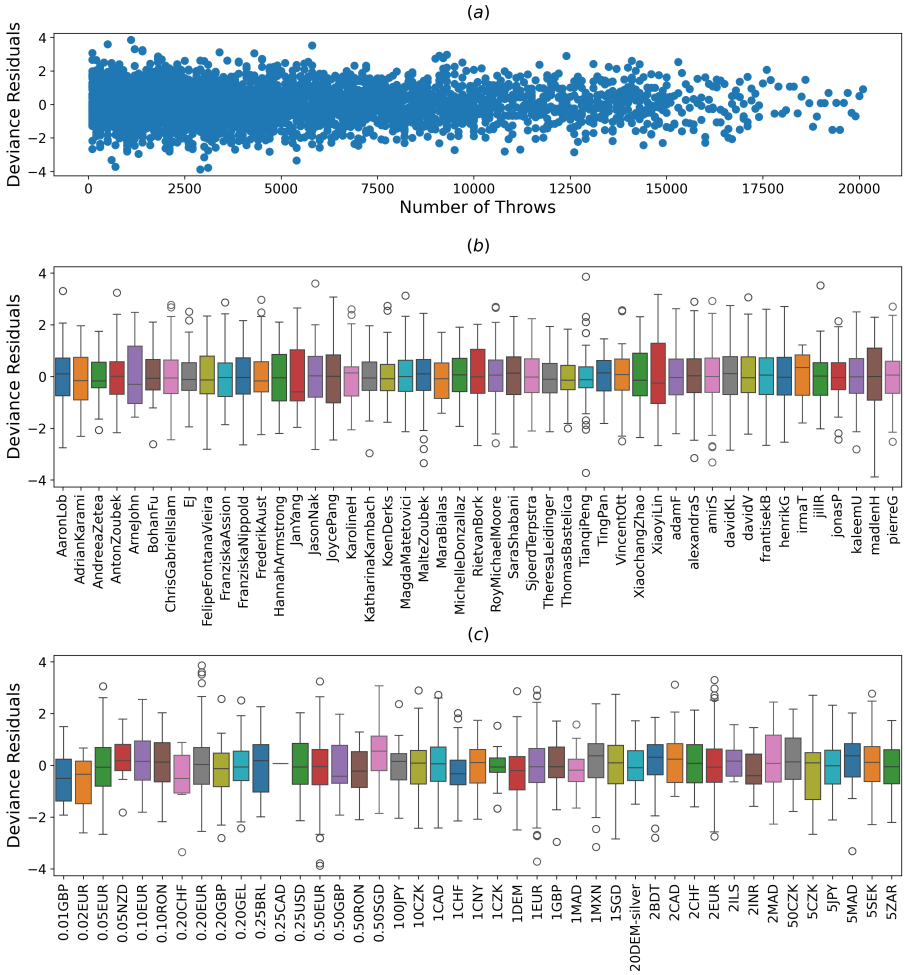

Fig 2: *Dev-resid as a function of (a) person and (b) coin.*

2.2 Unusual Observations

2.3 Zoom on Learning Effects

- bias comes from start
- amount of bias (considerable)
- wobble interpretation (consistent with physical model, citing the paper)

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

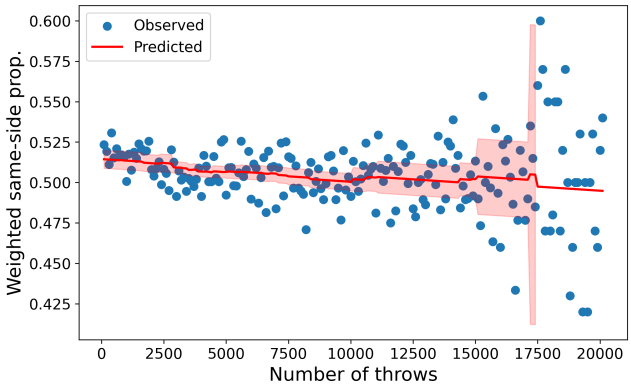

Fig 3: *Learning effects.*

2.4 Memory Effects

In this section we focus .

Tab 4: *Model comparison for models including : no memory, 1-hop memory and 2-hop memory.*

Model	Deviance	AIC	Model DF
1	264930.81	0.00	0
1+hop1_mem	264929.96	1.15	1
1+hop1_mem+hop2_mem	264929.07	2.26	2

3 Discussion

4 Conclusion

Acknowledgements

References