

朝ごはんと成績に統計的関係があるか

– Peanuts のデータを用いて – *

Charlie Schultz [†] Chishio Furukawa [‡]

March, 1988

概要

Peanuts の登場キャラクターたちは、自分たちの経験に基づき「好きな朝ごはんを食べることが、学校の成績につながる」と考えている。本レポートでは、この統計的根拠を探索し、以下の結果を得た。(1) 朝ごはんに “pancake” を食べることは、小テストの成績と強く相関はしていなかった。(2) Snoopy が “corn flake” の代わりに “dog flake” ばかり買ったときには「朝ごはんに “dog flake” を食べ、アレルギーで勉強に集中できなかった」と Charlie と Sally は言っているが、これについては整合的な根拠が示された。

* 本レポートは、データサイエンスの授業において、R/RStudio、Git/GitHub、LaTeX/Overleaf を組み合わせて、再生可能なプログラミングとレポート文書作成の例として作成したものである。レポートの書式としては、あくまで内部資料として作成しており、外部向けの論文の形式を取っていない。また、Peanuts の設定は全て仮想である。

[†] Department of Comics. Email: cshultz@comics.edu

[‡] Department of Economics. Email: cfurukawa@economics.ac.jp

1 はじめに

1.1 背景

この統計分析は、朝ごはんと成績の関係性を検証することを目的としている。Charlie や Lucie らは、「おいしい朝ごはんを食べることが学校の成績に欠かせない」と信じている。特に、pancakes (日本語：ホットケーキ) は好物で (Schultz and Tanikawa 1990)、toast や oatmeal よりも、pancakes をもっと食べたいと考えている。学術的にも、朝ごはんを食べることが学校の成績と相関していると言われている (Ann-Marie K. Chandler 1995)。

1.2 問い

本レポートでは、James Street Elementary School と、それぞれの朝ごはんの記録のデータを結び付けて、次の2つの問いについて考える。

1. pancake を食べることは、学校の成績に相関しているだろうか。
2. 「dog flakes を食べることによって、勉強に集中できず、成績が悪くなった」という主張は、統計的に整合的だろうか。

1.3 記述統計

まず、これらの問いについて考える前に、登場キャラクターが、どのような朝ごはんを食べているかをまとめよう。

breakfast_renamed	count	frequency
cereal	292	0.33
dog flakes	6	0.01
no breakfast	41	0.05
oatmeal	206	0.23
pancakes	178	0.20
toast	172	0.19

(この表においては、キャラクターの記録した名前を整理してある。例: corn flake → cereal、など)

所感:

- もっともよく食べるものは、cereal や oatmeal などである。
- pancake を食べることは、それほど多くないが、ものすごく珍しいわけでもない。

- dog flakes を食べなくてはいけないことは数回しかなかった。

2 分析

2.1 pancakes と成績の関係

2.1.1 散布図

まず探索的な分析として、重要な変数の相関などを知るために散布図を描くことが効果的である。

図1は、学生ごとに色を変え、ある月のうちの「pancake 率」と math と English の成績の平均をプロットしたものである。

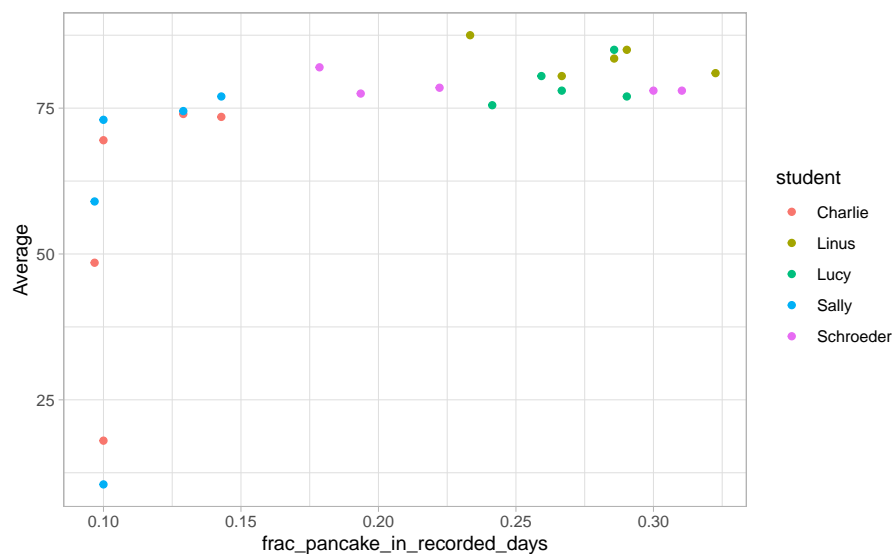


図1: pancake 率と成績の散布図

(この図1は、ggplot2 パッケージを用いて作成した「探索的図画」である。例えば、変数名やタイトルなどを調整していない。)

所感

- 全体としては、pancake と成績の間に弱い相関があるように見える。
- しかし、この相関は、学生間の関係であって、それぞれの学生に着目をすれば「pancake をよく食べた月の方が成績がよかった」というわけではないようである。

2.1.2 回帰分析

この散布図で見られた関係を、回帰分析によってより厳密に考えたい。以下の回帰式を推定したいと思います。

表1: Initial regressions

	(1)	(2)
	OLS	FE
frac(pancakes)	141.58 (20.80)	89.02 (117.94)
Constant	42.61 (4.84)	
Clustering	Y	Y
Num.Obs.	25	25
R2	0.373	0.388
R2 Adj.	0.346	0.227

Note:

Heteroskedasticity-robust standard errors clustered at students level are reported in the parenthesis.

$$\text{Test}_{it} = \beta_0 + \beta_1 \text{frac(pancakes)}_{it} + f_i + \varepsilon_{it} \quad (1)$$

- Test_{it} は、数学と英語の平均の点数
- $\text{frac(pancakes)}_{it}$ は、学生 i にとって月 t に食べた pancake の回数/ 観察が欠損していない日数
- f_i は、学生 i のダミー変数であり、固定効果を推定するためにある

ここで、関心のある推定量は β_1 であり、pancake 率が上昇することによって、成績が何ポイント改善するかを表している。

この回帰分析の推定結果が、表 1 にまとめられている。

所感:

- OLS によって、全体の相関を見ると、ある程度の相関が見られる。
- FE によって、個々人の学生にとっての相関を見ると、その相関はほとんどなくなっている。

(ここでは、統計的有意水準は示していない。これは、帰無仮説検定によってのみ結論の判断をすることを避けるためである。)

2.2 dog flakes と成績の関係

Charlie と Sally は、Snoopy が dog flakes のみしか朝ごはんを買っていなかったときがあったため、アレルギーで勉強ができなかった、と言っている。そこで、dog flakes を食べた月の試験は、実際に Brown 兄弟のみ成績の低下が見られたかどうかを検証する。

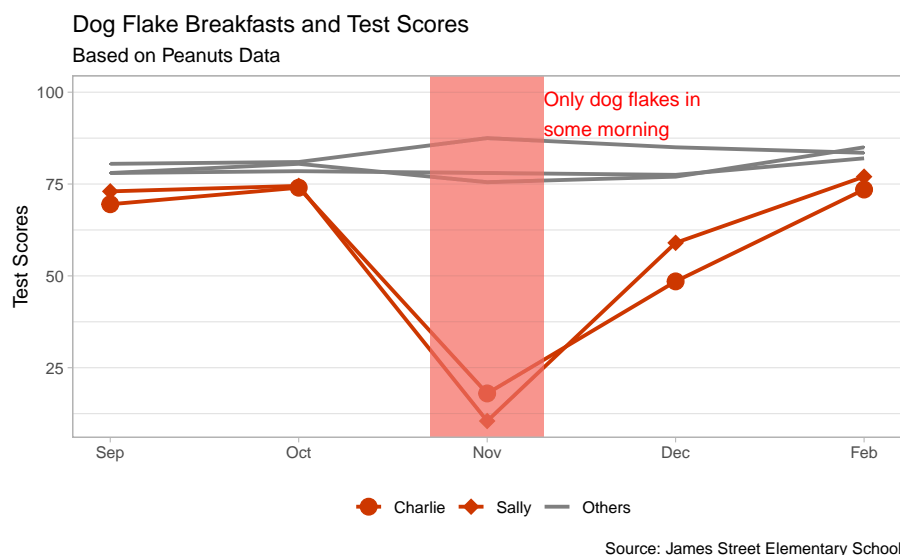


図2: dog flakes ショックと Brown 兄弟の成績

(この図2は、ggplot2 パッケージを用いて作成した「発表向け図」である。軸の名称やタイトル、またキャプションなどをつけ、また配色もマニュアルで調整している。)

所感:

- 他の学生と比べ、dog flake を食べた月と、その次の月の成績が低下していることが見られる。
- サンプル数が少ないため統計的検定をすることは難しいかもしれないが、Charlie と Sally の主張は正しい可能性が示唆された。
- pancake に関する分析は、コントロール変数などをあまり入れられていないこともあり、因果的解釈が難しい。これに対して、もし「Snoopy が dog flakes のみを買ったこと」がその他の諸要因とは関係がなかった「外生的ショック」と考えることが妥当ならば、dog flakes の項かとして捉えることができそうである。

参考文献

Ann-Marie K. Chandler, Kevin Connolly Sally M. Grantham-McGregor, Susan P. Walker (1995) "School Breakfast Improves Verbal Fluency in Undernourished Jamaican Children," *The Journal of Nutrition*, 125 (4), 894–900.

Schultz, Charlie and Shuntaro Tanikawa (1990) *A Peanuts book featuring Snoopy 1*: Kodokawa.