Heads and Reds

Andrew Oswald

18 November 2008

Universities of Warwick and Zurich

* For helpful discussions, I thank Bert Van Landeghem, Steve Stillman, and Michael Wolf.
Evidence is vital
Evidence is vital

The difficulty is this.
Humans have a tendency to see patterns in data when there are none.
Humans have a tendency to see patterns in data when there are none.
I want

• To illustrate why
• To make suggestions about how I attempt to do persuasive empirical work
Imagine I run a laboratory.
Imagine I run a laboratory.

My name is Andrew.
I have a collaborator
I have a collaborator

Amanda
We have a theory

We call it Time of the Day Effects.
We think Isaac Newton was not quite correct on gravity.
I am working on coin-tossing -- heads and tails.
I am working on coin-tossing -- heads and tails.
I do so 6 times in the morning, and 6 in the afternoon – for 7 days
In the other experiment, Amanda is spinning a roulette wheel.
She also does it 6 times in the morning, and 6 in the afternoon – for 7 days.
Our total observations are therefore 168.
We agree to collaborate and to send a paper to the Journal of Scientific Discoveries.
Our paper becomes famous
Our paper becomes famous

“Gravity Works Differently on Wednesday Afternoons.”
Stockholm beckons, we think?
But the key question
How likely are Andrew and Amanda to be able to write a paper (with random data) with a time-of-the-day effect that is statistically significant?
The answer is:
The answer is:

Extremely likely.
Say we are looking for a run of six heads or tails ($p < 0.02$) in a morning or afternoon.
The probability of throwing a coin 6 times in a row and getting a head each time is one half to the power 6.
The probability of throwing a dice 6 times in a row and getting a head each time is one half to the power 6. Write this as \((0.5)^6 = 1/64\).
The probability that EITHER Amanda or I find a result is
The probability that EITHER Amanda or I find a result is

\[ 1 - \text{probability there will neither a Heads-or-Tails Run nor a Red-or-Black Run}. \]
The probability that there will be neither is \((\frac{31}{32})^{28} = 0.41\).
So 59% of the time we will be able to write a paper proving, in a way that greatly exceeds the ninety-five confidence level, some version of “Heads come up on Wednesday afternoons”
In seminars

Amanda and I always say, correctly, to critics: “but our result is significant at the 0.02 level.”
Yet our paper is wrong.
The pattern is an illusion caused by too much searching.
In practice

How should we proceed in a world where the truth is blurry?
3 Suggestions
My personal checks
My personal checks

• Test for your key pattern in sub-samples.
My personal checks

- Test for your key pattern in sub-samples.
- Ask whether you made up your theory ex post.
My personal checks

• Test for your key pattern in sub-samples.
• Ask whether you made up your theory ex post.
• Ask: did you pre-search?
Replication is the best check.
The underlying point

Humans’ minds work so flexibly that they can see convincing patterns when there aren’t any.
Thank you, and good luck with your empirical work.

Evidence matters.
Heads and Reds

Andrew Oswald
www.andrewoswald.com

Universities of Warwick and Zurich

* For helpful discussions, I thank Bert Van Landeghem, Steve Stillman, and Michael Wolf.