

17 Horoscopes

Statistics courses and books—this one included—tend to resemble horoscopes. There are two senses to this resemblance. First, in order to remain plausibly correct, they must remain tremendously vague. This is because the targets of the advice, for both horoscopes and statistical advice, are diverse. But only the most general advice applies to all cases. A horoscope uses only the basic facts of birth to forecast life events, and a textbook statistical guide uses only the basic facts of measurement and design to dictate a model. It is easy to do better, once more detail is available. In the case of statistical analysis, it is typically only the scientist who can provide that detail, not the statistician.²⁴²

Second, there are strong incentives for both astrologers and statisticians to exaggerate the power and importance of their advice. No one likes an astrologer who forecasts doom, and few want a statistician who admits the answers as desired are not in the data as collected. Scientists desire results, and they will buy and attend to statisticians and statistical procedures that promise them. What we end up with is too often *horoscopic*: vague and optimistic, but still claiming critical importance.²⁴³

Statistical inference is indeed critically important. But only as much as every other part of research. Scientific discovery is not an additive process, in which sin in one part can be atoned by virtue in another. Everything interacts.²⁴⁴ So equally when science works as intended as when it does not, every part of the process deserves our attention. Statistical analysis can neither be uniquely credited with science's success, nor can it be uniquely blamed for its failures and follies.

And there are plenty of failures and follies. Science, you may have heard, is not perfect. *The Lancet* is one of the oldest and most prestigious medical journals in the world. This is what its editor-in-chief, Richard Horton, wrote in its pages in 2015:²⁴⁵

The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness.

Rethinking: Mercury rising. If I should offer you horoscopic advice, this is what I'd say. Thinking generatively—how the data could arise—solves many problems. Many statistical problems cannot be solved with statistics. All variables are measured with error. Conditioning on variables creates as many problems as it solves. There is no inference without assumption, but do not choose your assumptions for the sake of inference. Build complex models one piece at a time. Be critical. Be kind.

How do we know that much of the published scientific literature is untrue? There are two major methods.

First, it is hard to repeat many published findings, even those in the best journals.²⁴⁶ Some of this lack of repeatability arises from methodological subtleties, not because the findings are false. But many famous findings cannot be repeated, no matter who tries. There is a sense in which this should be unsurprising, given the nature of statistical testing. But the high false-discovery rate has become a great concern, partly because many placed unrealistic faith in significance testing and partly because it is hugely expensive to try to develop drugs and therapies from unrepeatable medical findings. It is even more expensive to design policy around false nutritional, psychological, economic, or ecological discoveries.²⁴⁷ But the basic reputation of science is also at stake, all material costs aside. Why pay attention to breathlessly announced new discoveries, when as many as half of them turn out to be false?

Second, the history of the sciences is equal parts wonder and blunder. The periodic table of the elements looks impressive now, but its story is unglamorous. There were more false elemental discoveries than there are current elements in the periodic table.²⁴⁸ Don't think that all these false discoveries were performed by frauds and cranks. Enrico Fermi (1901–1954) was one of the greatest physicists of the twentieth century. He discovered two heavy elements, ausonium (Ao, atomic number 93) and hesperium (Es, atomic number 94). These atomic numbers are now assigned to neptunium and plutonium, because Fermi had not actually discovered either. He mistook a mix of lighter already-discovered elements. These sorts of errors, and many other sorts of errors, were routine on the path to the current periodic table. Its story is one of error, ego, fraud, and correction. Other sciences look similar. Philosophers of science actually have a term, *the pessimistic induction*, for the observation that because most science has been wrong, most science is wrong.²⁴⁹

How can we reconcile such messy history, and widespread contemporary failure, with obvious successes like General Relativity? Science is a population-level process of variation and selective retention. It does not operate on individual hypotheses, but rather on populations of hypotheses. It comprises a mix of dynamics that may, over long periods of time, reveal the clockwork of nature.²⁵⁰ But these same dynamics generate error. So it's entirely possible for most findings at any one point in time to be false but for science in the long term to still function. This is analogous to how natural selection can adapt a biological population to its environment, even though most individual variation in any one generation is maladaptive.

What is included in these dynamics? Here's a list of some salient pieces of the dynamic of scientific discovery, in no particular order. You might make your own list here, as there's nothing special about mine.

- (1) Quality of theory and predictions: If most theories are wrong, most findings will be false positives. Karl Popper argued that all that matters for a theory to be scientific is that it be falsifiable. But for science to be effective, we must require more of theory. There was a brief quantitative version of this argument on [page 51](#). A good theory specifies precise predictions that provide precise tests, and more than one model is usually necessary.
- (2) Dynamics of funding: Who gets funded, and how does the process select for particular forms of research? If there are no sources of long-term funding, then necessary long-term research will not be done. If people who already have funding judge who gets new funding, research may become overly conservative and possibly corrupt.

- (3) Quality of measurement: Research design matters, all agree; but often this is forgotten when interpreting results. A persistent problem is designs with low signal-to-noise ratios.²⁵¹ Poor signal will not mean no findings, just unreliable ones.
- (4) Quality of data analysis: The topic of this book, but still much broader than it has indicated. Many common practices in the sciences exacerbate false discovery.²⁵² If you are not designing your analysis before you see the data, then your analysis may overfit the data in ways that regularization cannot reliably address.
- (5) Quality of peer review: Good pre-publication peer review is invaluable. But much of it is not so good. Many mistakes get through, and many brilliant papers do not. Peer review selects for hyperbole, since honestly admitting limitations of work only hurts a paper's chances. Is this nevertheless the best system we can devise? Let's hope not.
- (6) Publication: We agonize over bias in measurement and statistical analysis, but then allow it all back in during publication.²⁵³ Incentives for positive findings and newsworthy results distort the design of research and how it is summarized.²⁵⁴
- (7) Post-publication peer review: What happens to a finding after publication is just as important as what happens before. It is common for invalid analyses to be published in top-tier journals, only to be torn apart on blogs.²⁵⁵ But there is no system for linking published papers to later peer criticism, and there are few formal incentives to conduct it. Even retracted papers continue to be cited.
- (8) Replication and meta-analysis: The most important aspects of science are repetition and synthesis.²⁵⁶ No single study is definitive, but incentives to replicate and summarize are weaker than incentives to produce novel findings. Top-tier journals prioritize news. But if the literature is biased, then aggregating the literature just magnifies bias.

We tend to focus on the statistical analysis, perhaps because it is the only piece for which we have formulas and theorems. But every piece deserves attention and improvement. Sadly, many pieces are not under individual control, so social solutions are needed.

But there is an aspect of science that you do personally control: openness. Pre-plan your research together with the statistical analysis. Doing so will improve both the research design and the statistics. Document it in the form of a mock analysis that you would not be ashamed to share with a colleague. Register it publicly, perhaps in a simple repository, like Github or any other. But your webpage will do just fine, as well. Then collect the data. Then analyze the data as planned. If you must change the plan, that's fine. But document the changes and justify them. Provide all of the data and scripts necessary to repeat your analysis. Do not provide scripts and data "on request," but rather put them online so reviewers of your paper can access them without your interaction. There are of course cases in which full data cannot be released, due to privacy concerns. But the bulk of science is not of that sort.

The data and its analysis are the scientific product. The paper is just an advertisement. If you do your honest best to design, conduct, and document your research, so that others can build directly upon it, you can make a difference.