

Statistical Rethinking - Reference paper notes

Thejasvi Beleyur

2021-02-01

Contents

1	Introduction	5
2	The harm done by tests of significance	7
2.1	Episode 1: the right-turn-on-red story	7
2.2	Episode 2: the safety effect of paving shoulders	8
2.3	Episode 3: speed limit increases	8
3	Assessing uncertainty in physical constants	9
3.1	Introduction	9
3.2	The psychology of judgment under uncertainty	9
3.3	Uncertainty in physical measurement	10

Chapter 1

Introduction

This book contains short notes on a selection of papers referred to in the end notes of *Statistical Rethinking* (McElreath, 2020). Some of these papers are very technically challenging statistically, and otherwise - and the notes represent my best attempts at trying to boil down the ‘essence’ of the papers. Of course, I’d be very happy to hear from any readers out there about potential additions, omission or errors.

- Thejasvi , Jan 2020

Chapter 2

The harm done by tests of significance

Reference : (Hauer, 2004)

- The issue is that university education is centred around Null-hypothesis significance tests
- *‘I have written about the paralyzing effect of statistical significance on road safety research a long time ago...However, the road safety research is a constant reminder of the continuing harm done by NHST.’*
- *“The harm is that of using sound data to reach unsound conclusions thereby giving sustenance to non-sensical beliefs.”*

2.1 Episode 1: the right-turn-on-red story

- At traffic junctions in some countries/cities, a car can turn into the direction of traffic (left/right) even when the traffic light is red. This episode relates to a series of studies which tried to assess whether this practice caused more accidents.
- Many studies observed slight increases in accident rates after intersections were converted from ‘normal’ to ‘right-turn-on-red’, however they all reported non-significant test results repeatedly.
- The non-significant results were interpreted to mean no difference.
- Only after many studies came together, the data could be put together - which pointed at the negative effects of turn-on-red intersections.

2.2 Episode 2: the safety effect of paving shoulders

- A ‘paved shoulder’ is a width of road that is outside the defined lanes.
- Adding 2-4 feet of paved shoulder seemed to suggest an overall decrease in various types of accidents.
- NHST showed no significant differences, and thus were reported as having no effect
- *‘Once again common sense and statistical ritual point in opposite directions.’*
- **‘The absence of statistical significance does not mean and should never be taken to mean that 0 is the most likely estimate.’**
- The lack of significance was taken to mean that there is no effect, which was then used to justify narrower roads.

2.3 Episode 3: speed limit increases

- In some states of the USA, the speed limits were increased. A study analysed the association between fatal crashes and speed limit increases in two years: 1987 and 1995.
- Based on structural equation modelling, the percentage increase of fatal crashes was estimated.
- The estimated increase in crashes was reported only for states and years where a significant effect was found. Non-significant effects were reported as having ‘0’ % changes - or *no change*.
- The authors state that increasing speed limits led to increased crashes only in some of the states, and not all the states.
- Here, Hauer examines the raw data - percentage change +/- RMSE and shows that overall there is a positive mean increase in crash rate across both 1987 and 1995.
- *‘However, when good data is passed through the NHST filter, a negative tends to emerge; black turns to white and white to black’*

Chapter 3

Assessing uncertainty in physical constants

Reference : (Henrion and Fischhoff, 1986)

The authors broadly discuss “*..the problem of underestimation of uncertainty.*”

3.1 Introduction

- The uncertainty of a measurement is important to 1) allow comparisons with other measurements 2) quantify the impact it has on derived parameters 3) check its match with theoretical predictions
- Authors point out examples where actual uncertainty has been grossly underreported by eg. R.A. Milikan’s estimate of e (the charge of a single electron), the accuracy of which turned out to be at least 3 standard deviations.
- “*A comprehensive assessment of uncertainty cannot rest solely on statistical analysis. Unavoidably, it involves a considerable element of subjective judgment.*”

3.2 The psychology of judgment under uncertainty

- Good estimates of uncertainty can be made by humans when there is training and scope for feedback to improve assessment.

- When the conditions aren't right, or training isn't available, it can lead to overconfidence (positing very low uncertainty) or underconfidence (positing very high uncertainty).
- Subjective confidence intervals are a common way to assess the uncertainty of a quantity. (eg. 'this carrot has a 98% probability of weighing between 100-200 gm')
 - When the true quantity falls outside the interval, it is a 'surprise'. A 'surprise index' summarises how often the quantity falls out of the interval.
 - If the surprise index in a 98% interval is above 2%, it indicates a general overconfidence, and when it's less than 2%, indicates an underconfidence.
- The 'interquartile index' is another way of measuring uncertainty assessments.
 - The interquartile index is the percent of judgments where the quantity lies between the 25-75th %ile. (ie. How often the true value lies in the interval given by the statment 'This carrot has a 50% chance of weighing 100-200gm')
 - A interquartile index < 50% indicates overconfidence.
- In general, nonspecialist subjects tend to be somewhat overconfident showing surprise indices between 20-40%, and interquartile indices of 30-40%

3.3 Uncertainty in physical measurement

- Measurements are always accompanied by their uncertainty, as std or confidence intervals.
- 'Error' and 'uncertainty' are used interchangeably, and the authors argue for their dissociation:
 - 'error': the difference between the actual value and the measurement - which *'is generally unknown at the time of measurement'*
 - 'uncertainty': the investigator's (subjective?) assessment of the amount of error.
-

Bibliography

- Hauer, E. (2004). The harm done by tests of significance. *Accident Analysis & Prevention*, 36(3):495–500.
- Henrion, M. and Fischhoff, B. (1986). Assessing uncertainty in physical constants. *American Journal of Physics*, 54(9):791–798.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.