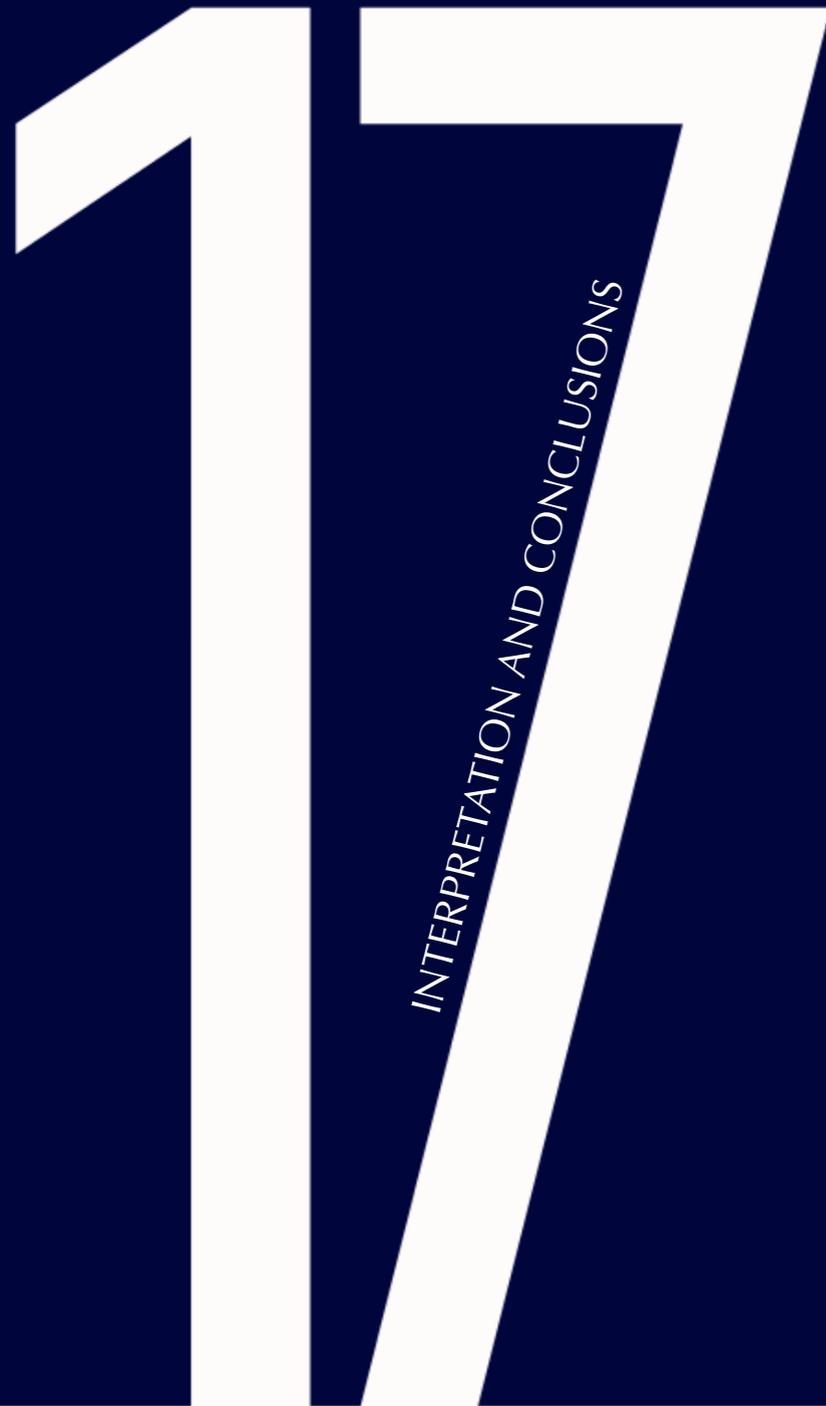


CHAPTER 17: UNCERTAINTY AND RANDOM: WHEN IS CONCLUSION JUSTIFIED?



Deduction – the use of facts to reach a conclusion – seems straightforward and beyond reproach. The reality is that uncertainty underlies every step in deductive inference. The uncertainty applies at many levels.

Introduction

Any fan of Conan Doyle's Sherlock Holmes has no doubt marveled at the stunning deductive powers of the mythical detective: Holmes's glance at a suspect's shoes evokes a proclamation that the gray smudge is of a clay found only in a particular quarry outside of Dover, that the person's wrinkled clothes indicates a recent train ride that day, and before you know it, Holmes has produced a dazzling chain of connected facts that accounts for the suspect's whereabouts for the previous 24hr. It sounds so logical and flawless.

The problem is that those deductions never recognize uncertainty. There is some chance that the gray smudge on the person's shoes is not clay from Dover, but is pigeon poop, paint, or any of 100 other materials; the wrinkling of clothes might stem from being worn two days in a row, and so on.

In applying the scientific method to reach a conclusion, we want to acknowledge the uncertainty. Ideally, we hope to reduce the major sources of uncertainty, but in any case, we should not do what Holmes does – we should not regard our conclusion as fact.

The Roots of Uncertainty

Consider the conclusion that levels of a particular protein hormone in the body (leptin) determine the body mass index (BMI): whether the person is thin, of normal weight, or obese. The initial studies of leptin were based on mice, and indeed, the biotech company Genentech spent several hundred million dollars acquiring the rights to use the leptin gene therapeutically. The data we might imagine using to reach this conclusion could include:

leptin levels and BMI in a mouse strain

leptin levels and BMI in a sample of humans

Suppose we find that BMI and leptin show a trend in both mice and humans. Where is the uncertainty in concluding that leptin is the basis of BMI? Here are some issues to consider:

1. inappropriate model - mice may not be a good model of humans
2. bad protocol – the measured leptin levels may be inaccurate, so the patterns are not real.
3. bias - the sample of people used may not be representative of most humans (e.g., perhaps they were all middle-aged, white males)
4. insufficient replication - the number of people in the sample may be small, so that any pattern may have arisen by chance. We use statistics to decide this possibility, and the matter is addressed below under ‘random.’
5. correlations - the leptin-BMI pattern may be real but leptin is not ‘causal.’ We deal with this

Thus any conclusion about leptin and BMI must acknowledge and address these and other sources of uncertainty. Initially, it may not be possible to quantify the uncertainty or even to decide that leptin is 'probably' a major determinant of BMI. As more data are obtained, the role of leptin on BMI should be increasingly resolved.

In general, uncertainty underlies the models used and data quality in many ways.

Random

Randomness is a form of uncertainty that we often attempt to quantify. When you play cards or roll dice, the so-called games of chance, you are knowingly allowing randomness to have a big influence on your short-term fate. Of course, randomness is what makes those games interesting and puts everyone on a somewhat equal basis for winning. Not all variation is due to chance – when you step on the gas pedal to make the car go faster, you are creating non-random variation in your speed. Random is specifically reserved to explain why we get different outcomes (= variation) when trying to keep everything the same, as with a coin flip. When it comes to the scientific method, we are mainly interested in whether some observed variation is due to chance or something else (e.g., is the accident rate of drivers talking on cell phones higher than that of drivers not on cell phones).

Not All Randomness is the Same

Randomness comes in different flavors. A coin flip represents one type of random – two possible outcomes with equal probability. (A die is a similar type of random but with 6 possible outcomes.) Random variation may instead fit a bell curve, as if we were considering how much your daily weight differed from its monthly average: most of the daily differences would be small, but a few might be large. Yet another type of randomness describes how many condoms are expected to fail in a batch of 1000.

Statistics: Testing Models

Most people have heard of statistics, and we mentioned it in a previous chapter. This mathematical discipline should probably be considered a top-ten phobia for most college students, but it is unfortunately useful in the scientific method. The principle behind most statistical tests is simple, however. A statistical test merely compares a particular model of randomness with some data. When a null model is rejected, it means that the data are NOT compatible with that particular brand of randomness. In essence, a statistical test is a substitute for replication, but instead of replicating the data, the test replicates the model of randomness to see often the random process fits the real data.

Wierdnesses of Random

Some properties of randomness are intuitive, but others are not. Some of the interesting properties of randomness can be explained without any use of mathematics. It can be useful to be aware of them, so you do not get ‘fooled’ by randomness. There is in fact a book with that title (‘Fooled by Randomness’) that explains how many seemingly significant events in our lives and in the stock market are due merely to chance, and the demise of many investment analysts has resulted from their failure to appreciate the prevalence of randomness in their early success.

Runs and excesses

If you flip a coin (randomly), you expect a Head half the time on average. Sampling error will cause deviation from exactly 50%, but as the number of flips gets really large, the proportion of heads will get closer and closer to $1/2$.

You can ask a different question, however. At any step in the sequence of coin flips, you will have either an excess of heads overall, an excess of tails, or have exactly 50% of each. If you have observed more heads than tails, for example, how likely is it that the number of tails will ‘catch up’ so that you then have as many or more tails than heads? From the fact that the observed proportion of heads gets closer and closer to 0.5 as more flips are done, it might seem that an excess of heads (or tails) will not last long. In fact, the opposite is true. As the number of flips increases, an excess tends to persist. From a gambler’s point of view, the fact that ‘he’ is losing does not mean that ‘he’ is ever likely to catch up, even if the game is fair and the odds of winning each hand are 50%. The longer the game goes on, there is less and less chance of ever breaking even.

A 'run' is a succession of wins with no losses (or a succession of losses with no wins). In athletics, runs can occur in a team's wins and losses or in a player's hits/baskets. There is a tendency to think that a player is 'hot' during a succession of good plays but is cold in a succession of misses. To describe a player is hot means, of course, that we don't think the string of good plays is due to chance, but instead stems from their being really good at those times. Yet when hot and cold strings have been analyzed statistically, they are usually consistent with random (like a coin flip, but one in which the odds of success differ from 50%).

Rare Encounters:

We know that the chance two unrelated people have the same birthday is approximately 1 in 365 (slightly less due to leap year and seasonal trends in birth rates). We might thus imagine that the probability of finding two people with the same birthday is small even when we consider a group of people. This intuition is wrong (again). In a group of 23 people, the chance that at least 2 of them share a birthday is approximately 1/2 -- because there are many different pairs of individuals to consider in a group of 23 (253 pairs to be exact), although not all pairs are 'independent' of the others.

There are many 'birthday problem' events in our lives. As you get older and have more experiences, there will be accidental meetings of people from your past and other coincidences that seem to improbable to arise from chance. However, when you average over the countless opportunities that you and others have for those rare events, it is not surprising that they happen occasionally.

A related phenomenon concerns the improbability of events in our lives. We often marvel at unique events and assume that something so unusual could not happen by chance. Yet our lives are a constant string of statistically improbable events. When you consider the identities of each card in a poker hand, each hand is just as improbable as every other hand. In fact, the probability of getting a royal flush is higher than the probability of getting the specific hand you were dealt; it's just that the vast majority of poker hands are worthless in terms of winning the game.

Scams

An apparently common scam in investment circles exploits randomness. It works like this. The scammer sends out monthly predictions about the stock market to 4096 potential clients. In the first month, half the clients receive a prediction that the market will go up, half receive the opposite prediction. At the end of the month, only half the predictions were correct (neglecting the possibility of no change). The scammer then sends out 'predictions' to the 2048 people who received correct predictions for the first month; once again, half of them receive predictions of an increase in the market, half receive predictions of a decrease. At the end of the second month, there are 1024 people who have received 2, consecutive correct predictions. Furthermore, if the scammer is clever, most of these prospective clients will not know the others who have been sent letters, so they will be unaware that half the letters sent out have made incorrect predictions. By continuing this methodology, after 5 months the scammer will be guaranteed of having 128 clients who have received 5 consecutive, correct predictions. If even a modest fraction of them are impressed, they may be prepared to invest heavily in the scammer's fund, with absolutely no assurance that it does any better than random.