***Learning Statistics with R - University of Adelaide***

***Part IV – Statistical Theory***

**Chapter 11 – Hypothesis Testing**

* “The process of induction is the process of assuming the simplest law that can be made to harmonize w/ our experience. This process, however, has no logical foundation but only a psychological one. It is clear there are no grounds for believing that the simplest course of events will really happen. It is a hypothesis that the sun will rise tomorrow: and this means that we do not know whether it will rise.” – Ludwig Wittgenstein
* **Estimation** was 1 of the 2 big ideas in **inferential statistics** + the other big idea is **hypothesis testing**.
* In its most abstract form, hypothesis testing really a very simple idea: researcher has some theory about the world + wants to determine whether or not the data actually support that theory.
* However, details are messy + most people find the theory of hypothesis testing to be the most frustrating part of statistics.
* Simple example study: Seek to test whether clairvoyance exists 🡺 Each participant sits at a table + is shown a card by an experimenter, which is black on 1 side + white on the other.
* Experimenter takes card away + places it on a table in an adjacent room black or white side up completely at random, w/ the randomization occurring only after experimenter has left the room w/ the participant.
* 2nd experimenter comes in + asks the participant which side of the card is facing upwards.
* Purely a 1-shot experiment: Each person sees only 1 card + gives only 1 answer + *at no stage is the participant actually in contact w/ someone who knows the right answer.*
* Dataset is very simple = asked the question of N = 100 people + some number X = 62 got the answer right, a surprisingly large number, sure, but is it large enough to claim evidence for ESP?
* This is the situation where hypothesis testing comes in useful.
* 1st distinction you need to keep clear is between **research hypotheses** and **statistical hypotheses**.
* In ESP study, overall scientific goal = to demonstrate clairvoyance exists
* Clear research goal: hoping to discover *evidence* for ESP
* In other situations, might actually be more neutral than that, so might say research goal = to determine whether or not clairvoyance *exists*.
* Basic point: a **research hypothesis** involves making a substantive, testable scientific claim
* If you’re a psychologist, your research hypotheses are fundamentally about psychological constructs
* Any of the following would count as research hypotheses:
* Listening to music reduces ability to pay attention to other things = a claim about causal relationship between 2 psychologically meaningful concepts (listening to music + paying attention to things), so it’s a perfectly reasonable research hypothesis.
* Intelligence is related to personality = a relational claim about 2 psychological constructs (intelligence + personality), but claim is weaker: *correlational*, NOT causal.
* Intelligence is speed of information processing: This hypothesis has a quite different character:
* not actually a relational claim at all but an **ontological** claim about the *fundamental character of intelligence*
* Actually worth expanding on this one
* Usually easier to think about how to construct experiments to test research hypotheses of the form “*does X affect Y?”* than to address claims like “*what is X?*”
* In practice, usually you find ways of *testing relational claims that follow from ontological ones.*
* Ex: If I believe intelligence is speed of information processing in the brain, my experiments will often involve looking for *relationships* between *measures of intelligence* + *measures of* *speed*.
* As a consequence, most everyday research questions tend to be **relational** in nature, but are almost always motivated by *deeper ontological questions* about the *state of nature*.
* Notice in practice, research hypotheses could overlap a lot.
* Ultimate goal in ESP experiment might be to test an ontological claim “ESP exists”, but I might operationally restrict myself to a *narrower* hypothesis like “Some people can ‘see’ objects in a clairvoyant fashion”.
* That said, there are some things that really don’t count as proper research hypotheses in any meaningful sense:
* Love is a battlefield. too vague to be testable.
* While it’s okay for a research hypothesis to have a degree of vagueness to it, it has to be possible to **operationalize** theoretical ideas
* Difficult to see how this can be converted into any concrete research design.
* If that’s true, this isn’t a scientific research hypothesis, it’s a pop song.
* Doesn’t mean it’s not interesting: a lot of deep questions humans have fall into this category
* Maybe 1 day science will be able to construct testable theories of love, or to test to see if God exists, + so on; but right now we can’t
* The first rule of tautology club is the first rule of tautology club: Not a substantive claim of any kind
* True *by definition*: No conceivable state of nature could possibly be inconsistent w/ this claim
* As such, say this = an **unfalsifiable hypothesis** + as such it is outside the domain of science
* *Whatever else you do in science, claims must have the possibility of being wrong.*
* More people in my experiment will say “yes” than “no”: Fails as a research hypothesis b/c it’s a claim *about the data set*, not *about the psychology* (unless your actual research question is whether people have some kind of “yes” bias).
* This hypothesis is starting to sound more like a **statistical hypothesis** than research hypothesis
* **Research Hypotheses** can be somewhat messy at times + ultimately they *are* scientific claims.
* **Statistical hypotheses** are *neither* of these 2 things 🡺 MUST be mathematically precise + MUST correspond to specific claims about the *characteristics of the data generating mechanism* (i.e., the “population”).
* Even so, the intent is that statistical hypotheses bear a *clear relationship* to the substantive research hypotheses you care about
* Ex: ESP study 🡪 research hypothesis = some people are able to see through walls/whatever.
* What I want to do is to *map* this onto a statement about *how* data were generated
* Quantity I’m interested in w/in the experiment is P(correct), the true-*but-unknown* probability w/ which participants in my experiment answer the question correctly.
* Let’s use the Greek letter θ (theta) to refer to this probability.
* Here are 4 different statistical hypotheses:
* If ESP doesn’t exist + if my experiment is well designed, my participants are just guessing:
* should expect them to get it right 1/2 of the time
* so my statistical hypothesis is the true probability of choosing correctly is θ = 0.5.
* Suppose ESP does exist + participants can see the card.
* If true, people will perform better *than chance*.
* Statistical hypothesis is θ > 0.5.
* ESP does exist, but the colors are all reversed + people don’t realize it
* If that’s how it works you’d expect people’s performance to be below *chance*.
* Correspond to a statistical hypothesis that θ < 0.5.
* Suppose ESP exists, but I have no idea whether people are seeing the right or wrong color.
* Only claim I could make about the data would be the probability of making the correct answer is not equal to 50%
* Corresponds to the statistical hypothesis that θ != 0.5.
* All of these are legitimate examples of a statistical hypothesis b/c they are statements about a population parameter + are meaningfully related to my experiment.
* What this discussion hopefully makes clear is that *when attempting to construct a statistical hypothesis, test that the researcher actually has 2 quite distinct hypotheses to consider.*
* 1st: They have a **research hypothesis** (claim about psychology)
* 2nd: It corresponds to a **statistical hypothesis** (claim about the data-generating population).
* ESP example: these might be:
* Research hypothesis: “ESP exists”
* Statistical Hypothesis: θ != 0.5
* The key thing to recognize is a statistical hypothesis test is a test of the statistical hypothesis, NOT the research hypothesis.
* If a study is badly designed, the link between the research + statistical hypothesis is broken.
* Suppose the ESP study was conducted in a situation where participants can actually see the card reflected in a window
* if that happens, I’d be able to find very strong evidence that θ != 0.5, but this would tell us nothing about whether “ESP exists”.