Hypotheses, associations

Rob Davies

2023-11-21

## 1 PSYC401: Classes weeks 6-10

* My name is Dr Rob Davies, I am an expert in communication, individual differences, and methods

|  |
| --- |
| Tip |
| **Ask me anything**:   * questions during class in person or anonymously through slido; * all other questions on discussion forum |

## 2 Weeks 6-10

* Introduction to our approach

## 3 Our approach: weeks 6-10

* Key idea: data analysis should be taught in **context**
* Early classes introduce R, open science ideas, basic statistical tests
* Later classes develop concepts and skills
* Critical change: work in **context** of live research project

## 4 Key idea: data analysis in context

* Traditionally, psychologists teach a different procedure each week: *a-test-a-week*
* limited discussion of theory or measurement,
* focus on rules about doing and reporting null hypothesis significance tests
* but:

1. This risks student (researcher) focus on doing the significance test (when, what, how)
2. While the *real* challenges are located in figuring out what we want to find out, measure, and explain

## 5 Our approach: Concepts, skills, levels

* Each week, we focus on building identified **concepts and skills**
* We develop step-by-step through identified **levels**
* In conceptual work, we aim for deep and broad understanding of what you do and why **to grow independence**
* In practical work with R, we **introduce, consolidate, or extend**

## 6 Targets for weeks 6-10: Concepts

We are working together to develop concepts: our work will *develop across the weeks* but we introduce the key ideas at specific points

1. *Week 6* — Credibility, the data analysis pipeline, data or model uncertainty, reproducibility
2. *Week 7* — Hypotheses, measurement and associations
3. *Week 8* — Principles and evidence in visualization, learning in R and open knowledge
4. *Week 9* — Linear models and prediction
5. *Week 10* — Linear models and the challenges of variation

## 7 Targets for weeks 6-10: Skills

We are working together to develop skills in a series of classes

1. *Week 6* — Reproducibility tests, accessing and analyzing shared data
2. *Week 7* — Visualizing, estimating, and reporting correlations
3. *Week 8* — Producing visualizations like a pro, navigating the R knowledge ecosystem
4. *Week 9* — Fitting, testing and reporting linear models
5. *Week 10* — Working with multiple variables, evaluating evidence across multiple samples

## 8 Let’s take a break

* End of part 1

## 9 Week 7

* Week 7 – Hypotheses, measurement and associations

## 10 Better methods can’t make up for mediocre theory

* **Previously – in week 6**: I talked about improving science through open *reproducible* methods but we cannot make progress without better theory and data (Smaldino, 2019)
* We want open reproducible findings but we do not just want reproducibility
* We want to make sense of people in useful ways

## 11 Open, reproducible, methods are not enough

* **Now**: we need to think causally about predictions and about measurement
* We discuss health comprehension project to demonstrate critical self-reflection

1. For useful hypotheses, we need better theory so we can build clear testable predictions from explicit assumptions
2. And with better models, we need better measurement because if we cannot reliably measure something then it is hard to build a theory about it

## 12 Critical thinking and you

* Students and colleagues almost never have problems *coding* analyses in R
* The challenges are *almost always* located in the critical reflection you must do in order to develop sensible analysis, and to interpret the analysis results
* So we need to start by highlighting the work of critical reflection in data analysis

## 13 Why most psychological research findings are not even wrong

* As you will know, it is often difficult to identify a claim in an article (Scheel, 2022)
* Here are some questions you can ask to decide if a claim *you read or make* is clear:

1. Is the claim stated unambiguously: can the claim support or contradict (or is it uncertain about) a prediction?
2. Can you understand how we get back from the claim to the data, given assumptions about measurement, sampling and procedure?

## 14 Why Hypothesis Testers Should Spend Less Time Testing Hypotheses

* The response to crisis has been to teach and use better methods
* This improvement reveals a core problem (Scheel et al., 2021): we often work to test hypotheses but our hypotheses are often *undeveloped*
* We train hypothesis testing but we also need to train *hypothesising*:
* how to measure, operationalize, and how to decide if hypothesis is corroborated or not

**We want to be capable of being wrong**

## 15 Statistical rituals largely eliminate critical statistical *thinking*

* Traditionally, students learn statistical tests, and learn to identify if a test statistic is significance or not
* If we do not also talk about what is actually observed, and whether or how – or why – it is or is not compatible with theory based predictions then we do *ritual not science* (Gigerenzer, 2004)
* This is a problem: the focus on anything-but-null allows us to build or accommodate vague theories that can never be wrong

## 16 We need to think about the derivation chain

|  |  |
| --- | --- |
| |  | | --- | |  |   Figure 1: The derivation chain |

## 17 Here’s a toolkit for thinking productively about your hypotheses

The derivation chain (Meehl, 1990; Scheel et al., 2021)

1. Develop your theory: the concepts, and the assumptions about causality
2. Specify how psychological concepts will be measured
3. Identify auxiliary assumptions about how we get from theoretical concepts to observable data
4. Identify theoretical predictions
5. Link theoretical predictions to specific statistical tests that may support or contradict them

## 18 Valid measures

* We often teach and learn about different kinds of validity but the key idea is simple (Borsboom et al., 2004):
* a test is valid for measuring an attribute if and only if (a) the attribute exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure
* We want to work with valid measures but *validity* requires explaining: (Q.1) Does the thing exist in the world? (Q.2) Is variation in that thing be reflected in variation in our measurement?

## 19 Summary: our critical thinking checklist

* What is our (causal) theory?
* What measures are we using, why?
* What is our specific prediction, why?
* Does the prediction relate to sign and to magnitude?
* What analysis can test this prediction, why?
* How will our results affect our beliefs, why?

## 20 Learning targets for this week:

* *Concepts*: begin with critical thinking
* *Skills*: developing hypotheses

## 21 Learning targets for this week:

* *Concepts* – associations: correlations, estimates and hypothesis tests
* *Skills* – visualizing variation and covariation
* *Skills* – writing the code
* *Skills* – estimating correlations
* *Skills* – hypothesis tests for correlations
* *Skills* – interpreting and reporting correlations

## 22 Let’s take a break

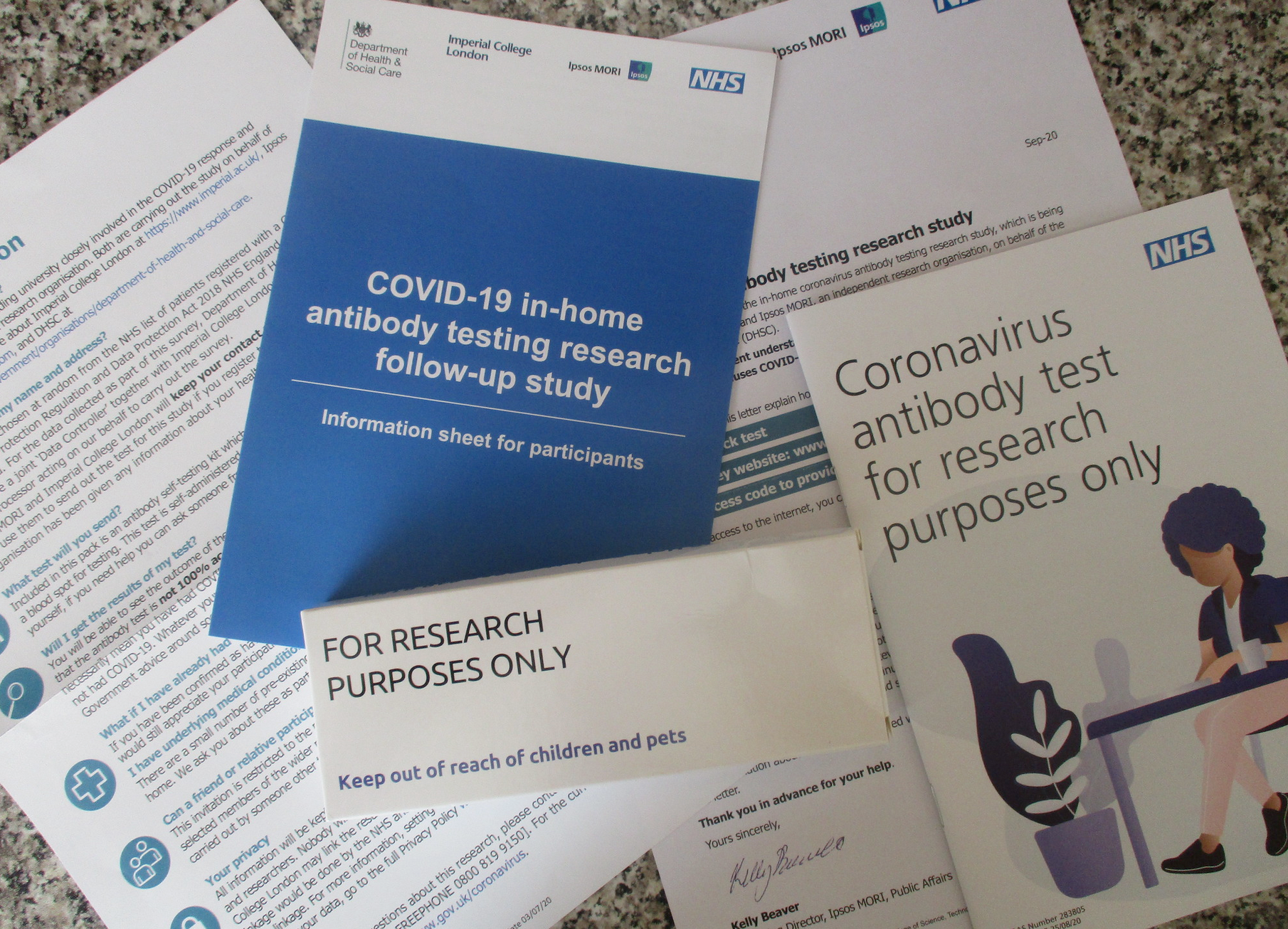
* End of part 2

## 23 Case study: the health comprehension project

* Because the important questions concern how psychologists ask and answer research questions
* We will work in the context of a live research project: *What makes it easy or difficult to understand written health information?*

|  |
| --- |
| flickr: Sasin Tipchair ‘Senior woman in wheelchair talking to a nurse in a hospital’ |

## 24 Why this? We don’t really know what makes it easy or difficult to understand advice about health



## 25 Health comprehension project: impacts

* We are working to help improve communication
* With partners at Vienna Business University, Kantar Public, and the London School of Economics
* Our work has implications for: business communication; understanding reading development; marketing communication

## 26 Health comprehension project: questions and analyses

* Our research questions were:

1. What person attributes predict success in understanding?
2. Can people accurately evaluate whether they correctly understand written health information?

* These kinds of research questions can be answered using methods like **correlation, linear models**

## 27 Health comprehension project – relevance: methods you will use in your professional work

* We got funding to collect data online using online Qualtrics questionnaire surveys
* We tested people on a range of dimensions using standardized ability and our own knowledge tests
* Many of you will go on to work with online surveys, and with data from standardized ability measures

## 28 You can get involved

* You can – if you choose – get involved
* Developing the project repository and associated documentation (forthcoming on PEP) and be a named co-author of the preprint we produce

|  |
| --- |
| Extract from Qualtrics survey |

## 29 Health comprehension project: why it is a case study

* The health project has strengths and limitations
* We show how to identify and critically evaluate this project so you can do the same for your work

## 30 Cognitive process theory of comprehension success

* When skilled adult readers read and try to understand written text (Kintsch, 1994)
* They must recognize and access the meanings of words
* Then use knowledge and reasoning to build an interpretation of what is in the text
* Based on connecting the information in the text with what they already know

## 31 Individual differences theory of comprehension success

* Successfully understanding text depends on (1.) language experience and (2.) reasoning ability (Freed et al., 2017)

|  |  |
| --- | --- |
| |  | | --- | |  |   Figure 2: Factors influencing comprehension success |

## 32 Where the data come from: our measures

* We measure reading comprehension: asking people to read text and then answer multiple choice questions
* We measure background knowledge: vocabulary knowledge (Shipley); health literacy (HLVA)
* We ask people to rate their own understanding of each text

## 33 Example critical evaluation questions

* Are multiple choice questions good ways to probe understanding? – What alternatives are there?
* Are tests like the Shipley good measures of language knowledge? – What do we miss?
* Can a person accurately evaluate their own understanding? – Can we rely on subjective judgments?

## 34 Relevance to you

* Even very good students sometimes do not question the validity of measures:
* Not asking questions like this has a real impact on the value of the interpretation of results
* Here, we are looking ahead to the critical thinking you will need to do for your dissertations

## 35 Let’s take a break

* End of part 3

## 36 Talking about the relationships between variables

* Psychologists and people who work in related fields often want to know about *associations*
* Is variation in observed values on one dimension (e.g., comprehension) related to variation in another dimension (e.g., vocabulary)?
* Do values on both dimensions *vary together*?

## 37 The language in this area can vary: we will be consistent but you need to be aware of the different terms

* **Outcome** response criterion dependent variable
* **Predictor** covariate independent variable factor
* **Linear model** regression analysis regression model multiple regression

## 38 Let’s look at the data we will use

* The person in row 1 has ETHNICITY White, is AGE 34 years, scored 33 on Shipley vocabulary, scored 7 on HLVA health literacy and, on average, self-rated their understanding of health information as 7.96 (*so* 8/9, mean.self) while scoring 0.49 accuracy in tests of understanding (49% mean.acc)

# A tibble: 4 × 6  
 mean.acc mean.self HLVA SHIPLEY AGE ETHNICITY  
 <dbl> <dbl> <dbl> <dbl> <dbl> <fct>   
1 0.49 7.96 7 33 34 White   
2 0.85 7.28 7 33 25 White   
3 0.82 7.36 8 40 43 White   
4 0.94 7.88 11 33 46 White

## 39 Destination correlation: where the correlation number comes from

**Covariance**

* If we want to estimate the correlation between two sets of numbers: and
* We want to know if variation in (given by )
* Varies together with variation in (given by )

## 40 Destination correlation: where the correlation number comes from

**Covariance divided by standard deviations**

* Because the two sets of numbers can be on different scales: e.g., SHIPLEY out of 40; mean.acc (proportion, out of 1)
* And because *covariance* values depend on the scales
* To correlations easier to compare, we need to remove scale by dividing by the variables standard deviations

## 41 Let’s think about an example correlation

* Research question: Can people accurately evaluate whether they correctly understand written health information?
* Measurement: Someone with higher scores on tested accuracy of understanding will also present higher scores on their ratings of their own understanding
* Statistical prediction: We predict that mean.acc and mean.self scores will be associated
* Test: If the prediction is correct, mean.acc and mean.self scores will be correlated

## 42 Distributions: How do scores vary?

|  |
| --- |
| Histograms showing the distribution of mean accuracy and mean self-rated accuracy scores in the ‘clearly.one.subjects’ dataset: means calculated for each participant over all their responses |

## 43 A histogram is a useful way to show the distribution of values

* We have a sample of accuracy scores:
* Mean accuracy scores vary between 0.0 and 1.0
* We draw the plot by grouping together similar values in bins
* Heights of bars represent numbers of cases with similar values in same bin

|  |
| --- |
| Distribution of mean accuracy |

## 44 When we talk about variance we are talking about how values vary in relation to the mean for the sample

* The average of these *mean accuracy* scores is marked with a red line where 0.8
* The accuracy score for the person in row 1 is located at , marked in blue

|  |
| --- |
| Distribution of mean accuracy |

## 45 We are talking about how values vary in relation to the mean for the sample

* In comparison, the mean accuracy score for the person in row 4 is located at , marked in blue

|  |
| --- |
| Distribution of mean accuracy |

## 46 The basic question when we examine covariance: do values vary *together*?

* If the person at row 1 has a mean.accuracy score of .49, lower than the average
* And the person at row 4 has a mean.accuracy score of .94, higher than the average
* What will their mean.self scores be: will *they* be higher or lower than the average mean.self score?

## 47 We can use scatterplots to examine associations

|  |
| --- |
| Scatterplots showing whether values on mean accuracy (mean.acc) vary together with values on mean self-rated accuracy (mean.self) for the participants in this sample |

## 48 A scatterplot is a useful way to examine if the values of two or more variables vary together

* Mean accuracy scores vary between 0.0 and 1.0
* The height of each point shows the observed value of accuracy on the y-axis
* Self-rated accuracy scores vary between 1 and 9
* The horizontal position of each point shows the observed value of self-rated accuracy on the x-axis

|  |
| --- |
| Scatterplot showing how values on mean accuracy and mean self-rated accuracy vary together |

## 49 A scatterplot is a useful way to examine if the values of two or more variables vary together

* We have a sample of 170 people
* For each person, we have a value for the mean accuracy and a *paired* value for the mean self-rated accuracy
* Each point shows the paired data values for a person
* In red: someone scored 3.48 on mean self-rated accuracy, 0.57 on mean accuracy

|  |
| --- |
| Scatterplot showing how values on mean accuracy and mean self-rated accuracy vary together |

## 50 Let’s take a break

* End of part 4

## 51 The R code for a correlation test, bit by bit

cor.test(clearly.one.subjects$mean.acc,   
 clearly.one.subjects$mean.self,  
 method = "pearson")

1. We specify the cor.test function, and name one variable clearly.one.subjects$mean.acc
2. Then we name the second variable clearly.one.subjects$mean.self
3. Last we specify the correlation method = "pearson" because we have a choice

## 52 Identifying the key information in the results from one correlation test

* We look at the value of the correlation (here, cor) and the p-value
* We can see that the correlation statistic is positive cor = .4863771 which we round to
* And p-value = 2.026e-11 indicating that the correlation is *significant*

cor.test(clearly.one.subjects$mean.acc,   
 clearly.one.subjects$mean.self,   
 method = "pearson")

Pearson's product-moment correlation  
  
data: clearly.one.subjects$mean.acc and clearly.one.subjects$mean.self  
t = 7.1936, df = 167, p-value = 2.026e-11  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.3619961 0.5937425  
sample estimates:  
 cor   
0.4863771

## 53 Reporting a correlation

* Usually, we report a correlation like this:

Mean accuracy and mean self-rated accuracy were significantly correlated (). Higher mean accuracy scores are associated with higher mean self-rated accuracy scores.

## 54 Interpreting correlations with the help of visualization

* The correlation statistic is positive in sign and moderate in size, about
* We can see that higher mean accuracy (mean.acc) scores are associated with higher mean self-rated accuracy (mean.self) scores

|  |
| --- |
| Scatterplot showing how values on mean accuracy and mean self-rated accuracy vary together |

## 55 What will different kinds of correlations look like?

We can simulate data to demonstrate: **(left)** the correlation is positive, ; **(right)** the correlation is negative,

|  |
| --- |
| Scatterplots showing how simulated data values on mean accuracy and mean self-rated accuracy could vary together given positive or negative correlations |

## 56 We can also imagine – again with simulated data – what correlations of increasing size might look like

|  |
| --- |
| Scatterplots showing how simulated data values on mean accuracy and mean self-rated accuracy could vary together given positive correlations of increasing size |

## 57 Summary

* We are often interested in whether or how variation in the values of two variables are associated
* We can visualize the distribution of values in any one variable using histograms
* We visualize the association of values in two variables using scatterplots
* We conduct correlation tests to examine the sign (positive or negative) and the strength of the association
* But we always need to think about our research questions, about where our data come from and about whether our measures are any good

## 58 Look ahead to week 8: grow in independence: supported

* Every problem you ever have: someone has had it before, solved it, and written a blog (or tweet or toot) about it

|  |
| --- |
| Rammstein concert crowd surfing; flickr, CC, Anirudh Koul |

## 59 Look ahead to week 8: the revolution in knowledge and you

* R is *free open statistical software*: everything you use is contributed, discussed and taught by a community of R users online, in open forums
* Learning to navigate this knowledge is an introduction to the future of knowledge sharing

|  |
| --- |
| flickr: Jeremy Brooks ‘Free speech fear free’ |

## 60 End of week 7

## 61 References

Borsboom, D., Mellenbergh, G. J., & Heerden, J. van. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>

Freed, E. M., Hamilton, S. T., & Long, D. L. (2017). Comprehension in proficient readers: The nature of individual variation. *Journal of Memory and Language*, *97*, 135–153. <https://doi.org/10.1016/j.jml.2017.07.008>

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606. <https://doi.org/10.1016/j.socec.2004.09.033>

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294–303. <https://doi.org/10.1037/0003-066x.49.4.294>

Meehl, P. E. (1990). *Why summaries of research on psychological theories are often uninterpretable 1, 2*. *i*, 195–244.

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development*, *31*(1), e2295. <https://doi.org/10.1002/icd.2295>

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, *16*(4), 744–755. <https://doi.org/10.1177/1745691620966795>

Smaldino, P. (2019). Better methods can’t make up for mediocre theory. *Nature*, *575*(7781), 9.