Linear models and critical perspectives on science and knowledge

Rob Davies

## 1 PSYC122: Classes in weeks 16-20

* 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 Week 19: Linear models – critical perspectives

|  |
| --- |
| Figure 1: Histograms showing the distribution of mean accuracy scores in 11 studies |

## 3 Targets for weeks 16-19: Concepts

We are working together to develop concepts:

1. *Week 16* — Hypotheses, measurement and associations
2. *Week 17* — Predicting people using linear models
3. *Week 18* — Everything is some kind of linear model
4. ***Week 19*** — **The real challenge in psychological science**

## 4 Targets for weeks 16-19: Skills

We are working together to develop skills:

1. *Week 16* — Visualizing, estimating, and reporting associations
2. *Week 17* — Using data to *predict* people
3. *Week 18* — Going deeper on linear models
4. ***Week 19*** — **Evaluating evidence across multiple studies**

## 5 Learning targets for this week

* *Concepts* – To engage with the real challenges in psychological science:

1. People vary
2. Results vary

## 6 Learning targets for this week

* *Skills* – To engage with the real challenges in data analysis skills development:

1. Growing in independence
2. Exploiting the R knowledge ecosystem

## 7 Why these targets? Key ideas

Science (including psychological science) has undergone a rolling series of crises:

* Replicability or replication crisis (Pashler & Wagenmakers, 2012)
* Statistical crisis (A. Gelman & Loken, 2014)
* Generalizability crisis (Yarkoni, 2022)

## 8 The triggers for crisis

* Failures to replicate influential claims (Nosek et al., 2022)
* Questionable research practices (John et al., 2012)
* Questionable measurement practices (Flake & Fried, 2020)
* Limited samples (Button et al., 2013; Henrich et al., 2010; Wild et al., 2022)

## 9 The credibility revolution: responses

* Pre-registration (Nosek et al., 2018, 2019) and registered reports (Nosek & Lakens, 2014)
* Replication studies (e.g. Aarts et al., 2015)
* Identification of open science principles (Munafò et al., 2017)

## 10 The credibility revolution: replication is recognised as crucial to building a science of psychology

|  |
| --- |
| Nosek et al. (2022): replication outcomes for three systematic replication studies |

## 11 The real challenge: people *vary*

* The real challenges we face as psychologists: **people vary**
* We examine the impact of this variation
* And we explore if or how we can *reproduce* or *generalize* findings in psychological science

|  |
| --- |
| flickr, Cat Walker ‘crowd’ |

## 12 The real challenge: results *vary*

|  |
| --- |
| Figure 2: Scatterplots showing the association between mean accuracy and self-rated accuracy of understanding, of health information, in 11 studies |

## 13 Variation and replication

Psychological and social processes show much more variability than the usual phenomena in the physical sciences (a. Gelman, 2015)

* The patterns or effects that interest us may (maybe *will*) vary between places, people, and times
* *Because* treatment effects can be expected to vary *then* we may not see replication of effects
* *So* we investigate *how* effects vary and we *open* our workflow

## 14 The professional data analysis workflow: from raw data to results

|  |  |
| --- | --- |
| |  | | --- | |  |   Figure 3: The data analysis pipeline or workflow |

## 15 The data analysis workflow

* Get some data
* Process or tidy the data
* Explore, visualize, and analyze the data
* Present or report your findings

|  |
| --- |
| Tip |
| * Identify the elements and the order in *your* work as the parts of a pipeline or the stages in a *workflow* |

## 16 Analysis multiverse

Different researchers: different *choices* (Silberzahn & Uhlmann, 2015)

|  |
| --- |
| Silberzahn and Uhlmann (2015): Twenty-nine research teams reached a wide variety of conclusions using different methods on the same data set to answer the same question. |

## 17 Kinds of reproducibility

Gilmore et al. (2017; following Goodman et al., 2016) present three kinds of reproducibility:

* Methods reproducibility
* Results reproducibility
* Inferential reproducibility

## 18 Inferential reproducibility

If researchers repeat a study (results reproducibility) or re-analyze original data (methods reproducibility) then they *should* come to similar conclusions as original authors

But …

* reproducibility attempt could reveal problems, uncertainty over choices
* different researchers could apply different prior expectations over the probability of possible effects

## 19 Lessons learned from crises mean we now hope to see that researchers:

1. Share data and code
2. Publish research reports in ways that enable others to check or query analyses

## 20 Let’s take a break

* End of part 1

## 21 Health comprehension project – answers to our questions

* We have been working 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’ |

## 22 Health comprehension project: questions and analyses

* Our research questions are:

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

## 23 Theory: Models of comprehension accuracy *should* include predictors:

(1.) experience HLVA, SHIPLEY and (2.) reasoning ability (FACTOR3, reading strategy) (Freed et al., 2017)

|  |  |
| --- | --- |
| |  | | --- | |  |   Figure 4: Understanding text depends on (1.) language experience and (2.) reasoning ability (Freed et al., 2017) |

## 24 Multiple candidate predictor variables

|  |
| --- |
| Figure 5: Scatterplots showing the potential association between accuracy of comprehension and variation on each of a series of potential predictor variables. |

## 25 Critical thinking: data analysis assumptions

1. **validity**: that differences in knowledge or ability cause differences in test scores
2. **measurement**: that this is equally true across the different kinds of people we tested
3. **generalizability**: that the sample of people we recruited resembles the population

## 26 Critical thinking: uncertainty

There are three levels of **uncertainty** when we look at sample data (McElreath, 2020) – uncertainty over:

1. The nature of the expected change in outcome
2. The ways that expected changes might vary between individual participants or between groups of participants
3. The random ways that specific responses can be produced

## 27 Critical thinking: working with samples

* We test who we can – convenience sampling – and who we can test has an impact on the quality of evidence (Bornstein et al., 2013)

|  |
| --- |
| Tip |
| Practice **critical evaluation**:   * If age, ethnicity or gender are not balanced does this matter to your research question? * If samples are limited in size how does that affect our uncertainty over effects estimates? |

## 28 Let’s take a break

* End of part 2

## 29 11 health comprehension studies

|  |
| --- |
| Figure 6: Dotplots showing the gender, education and ethnicity of participants across 11 studies |

## 30 Participants vary in accuracy

|  |
| --- |
| Figure 7: Grid of histograms showing the distribution of mean accuracy scores in each of 11 studies |

## 31 Participants vary in age

|  |
| --- |
| Figure 8: Grid of histograms showing the distribution of participant ages in each of 11 studies |

## 32 Participants vary in health literacy

|  |
| --- |
| Figure 9: Grid of histograms showing the distribution of health literacy (HLVA) scores in each of 11 studies |

## 33 Participants vary in vocabulary

|  |
| --- |
| Figure 10: Grid of histograms showing the distribution of vocabulary (Shipley) scores in each of 11 studies |

## 34 Participants vary in reading strategy

|  |
| --- |
| Figure 11: Grid of histograms showing the distribution of reading strategy (FACTOR3) scores in each of 11 studies |

## 35 Associations vary

|  |
| --- |
| Figure 12: Association between mean accuracy and health literacy |

## 36 Associations vary

|  |
| --- |
| Figure 13: Association between mean accuracy and mean self-rated accuracy |

## 37 Health comprehension project: answers

|  |
| --- |
| Note |
| 1. What person attributes predict success in understanding?  * Health literacy, vocabulary, and reading strategy  1. Can people accurately evaluate whether they correctly understand written health information?  * Yes but not very well |

## 38 Do we see replication across studies?

|  |
| --- |
| Figure 14: Varying estimated association between mean accuracy and health literacy |

## 39 Do we see replication across studies?

|  |
| --- |
| Figure 15: Varying association between mean accuracy and mean self-rated accuracy |

## 40 Results reproducibility

* If a researcher finds a pattern in human behaviour or in individual differences
* We may *critically evaluate* the robustness or the generalizability of the finding

|  |
| --- |
| Important |
| **Results reproducibility** means that a new study with new data, collected following the original procedures as closely as possible, yields the *same* outcomes Gilmore et al. (2017) |

## 41 Health comprehension studies evidence

* Maybe it is wiser – given levels of *uncertainty* – to expect *some* variation in results

|  |
| --- |
| Tip |
| What is your view?   * Do we see *robust* prediction of accuracy of understanding of health information, given measures of vocabulary, health literacy, and reading strategy? |

## 42 PSYC122 response data

* What will we see in a new study: with *your data*?

|  |
| --- |
| Tip |
| Will we see the same or different patterns?   * Do the practical work to find out |

## 43 Let’s take a break

* End of part 3

## 44 Kinds of reproducibility

Gilmore et al. (2017; following Goodman et al., 2016) present three kinds of reproducibility:

* Methods reproducibility
* Results reproducibility
* Inferential reproducibility

## 45 Methods reproducibility

* Other researchers should be able to get the same results if they use the analysis methods with the same data

|  |
| --- |
| Hardwicke et al. (2021): Frequency of reproducibility outcomes by value type. |

## 46 Require data and code sharing

* Analyses by Kidwell et al. (2016) and analyses reviewed by Nosek et al. (2022): study data increasingly available

|  |
| --- |
| Nosek et al. (2022): Yearly counts of users, sharing of files (research data, materials, code), and registration of studies on OSF and AsPredicted. |

## 47 Why we use R: We can share data and code helpfully

It is great that data are shared but analyses show they are not always readily usable (Towse et al., 2021) but *should be*

* **completeness**: are all the data and the data descriptors supporting a study’s findings publicly available?
* **reusability**: how readily can the data be accessed and understood by others?

## 48 Why we use R: We can write self-*documented* code

# Here: fit a linear model  
lm(mean.acc ~ SHIPLEY + HLVA + FACTOR3 + AGE + NATIVE.LANGUAGE, ...)

|  |
| --- |
| Tip |
| * R will ignore everything after # * Add a # comment after each step to briefly explain to yourself *and others* what is going on |

## 49 The R knowledge ecosystem

R is:

* a language
* a computing environment
* a knowledge ecosystem

## 50 R is a language

We use:

* functions like lm() in the same way we use *verbs* to describe doing things
* arguments like (mean.acc ~ HLVA) in the same way we use *nouns* to identify who does what to whom

# Here: fit a linear model  
lm(mean.acc ~ SHIPLEY + HLVA + FACTOR3 + AGE + NATIVE.LANGUAGE, ...)

## 51 R is a language

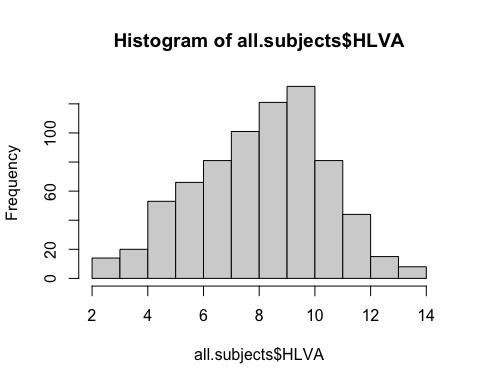
Like every language, we can often say the same thing using different words or accents

# This code does essentially the same job  
data <- read\_csv("mydata.csv")  
# As this code  
data <- read.csv("mydata.csv")

## 52 Languages have dialects

R has four different ways to draw plots: base, {lattice}, {grid}, {ggplot2}

# Base R graphics histogram of HLVA scores  
hist(all.subjects$HLVA)



## 53 The R knowledge ecosystem

Above all, R is **free**:



## 54 The R knowledge ecosystem

|  |
| --- |
| Tip |
| *Every* problem you ever have:   * someone has had it before * solved it * and written a blog (or tweet or toot) or recorded a YouTube or TikTok about it |

## 55 The R knowledge ecosystem

* 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, Cesar Salvadeo ‘Revolution’ |

## 56 How to find things out when you *know what you need*

* R has a built-in help system: typing

help(geom\_histogram)

* Gets you detailed technical information

… Examples ggplot(diamonds, aes(carat)) + geom\_histogram() …

|  |
| --- |
| Tip |
| Start with the examples |

## 57 How to find things out when you *don’t know what you need*

This code *won’t* work

all.subjects %>%  
 ggplot(aes(x = AGE, y = mean.acc)) +   
 geom\_histogram()

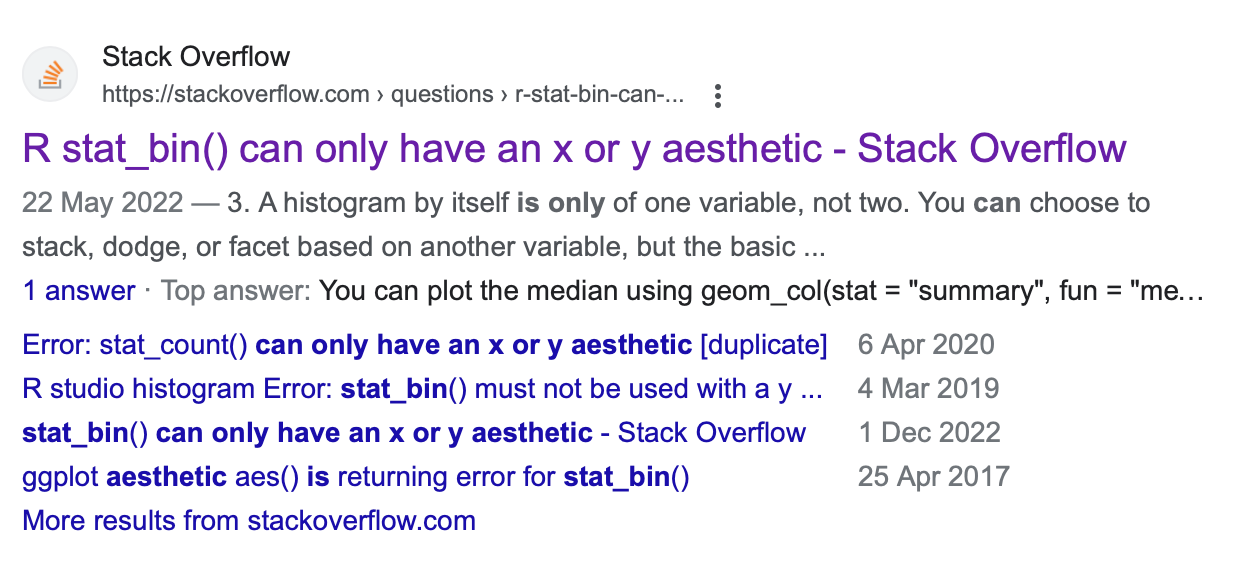
Error in `f()`:  
! stat\_bin() can only have an x or y aesthetic.  
Backtrace:

## 58 How to find things out when you *don’t know what you need*

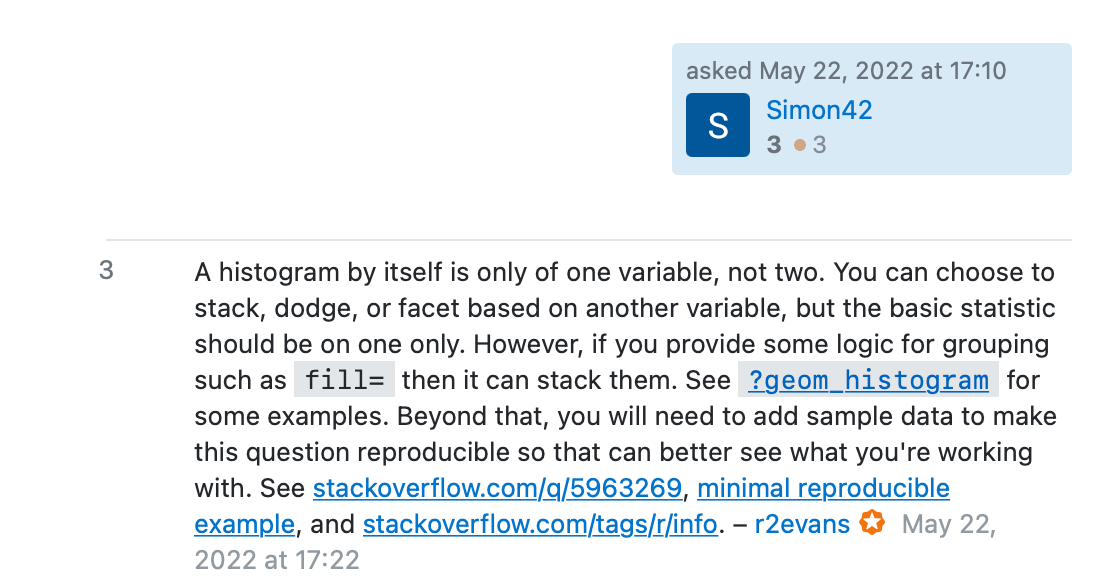
|  |
| --- |
| Tip |
| **Just google it:**   * Copy warning or error messages from R-Studio and paste them into a search engine |

## 59 Copying an error message into google gets you a list of web pages

stat\_bin() can only have an x or y aesthetic.



## 60 Stack Overflow lists question and answer discussions



## 61 How to find things out when you *don’t know what you need*

* Stack Overflow pages identify:

1. Questions asked
2. Answers, often with code solutions to problems, and helpful discussions

* With questions and answers ranked by usefulness

## 62 The wider revolution in building and sharing knowledge

We can find many excellent *free* online books like:

<https://r4ds.had.co.nz>

|  |
| --- |
|  |

## 63 A worldwide community of knowledge sharing

|  |
| --- |
| Screenshot of front page of ggplot gallery |

## 64 Summary

* In the health comprehension project: accuracy of understanding of health information can be predicted by vocabulary knowledge, health literacy and reading strategy
* People can judge their own accuracy of understanding but not well

|  |
| --- |
| Tip |
| * What do your *PSYC122* response data say? |

## 65 Summary: critical thinking

* We can expect results – associations, effects, patterns – to vary between times, places, people
* Samples will be limited, measurement under uncertainty
* Data analysis choices will vary between researchers
* *So* we share data and code *and* critically evaluate results

## 66 Summary: grow in independence

* Comment your code in your .R scripts to explain what you are doing and how
* Use online information sources to understand choices and options

|  |
| --- |
| Tip |
| * Someone has already solved your problem: you just need to find the blog/Stack Overflow discussion/TikTok where they explain the solution |

## 67 End of week 19

## 68 References

Aarts, E., Dolan, C. V., Verhage, M., & Van der Sluis, S. (2015). Multilevel analysis quantifies variation in the experimental effect while optimizing power and preventing false positives. *BMC Neuroscience*, *16*(1), 1–15. <https://doi.org/10.1186/s12868-015-0228-5>

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review*, *33*(4), 357–370. <https://doi.org/10.1016/j.dr.2013.08.003>

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. <https://doi.org/10.1177/2515245920952393>

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>

Gelman, a. (2015). The connection between varying treatment effects and the crisis of unreplicable research: A bayesian perspective. *Journal of Management*, *41*(2), 632–643. <https://doi.org/10.1177/0149206314525208>

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, *102*(6), 460–465. <https://doi.org/10.1511/2014.111.460>

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1396*, 5–18. <https://doi.org/10.1111/nyas.13325>

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341).

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3). <https://doi.org/10.1017/S0140525X0999152X>

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. <https://doi.org/10.1177/0956797611430953>

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*(5), 1–15. <https://doi.org/10.1371/journal.pbio.1002456>

McElreath, R. (2020). *Statistical rethinking*. Chapman; Hall/CRC. <https://doi.org/10.1201/9780429029608>

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. <https://doi.org/10.1038/s41562-016-0021>

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van?t Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*, 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>

Pashler, H., & Wagenmakers, E. J. (2012). Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. <https://doi.org/10.1177/1745691612465253>

Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature*, *526*(7572), 189–191. <https://doi.org/10.1038/526189a>

Towse, J. N., Ellis, D. A., & Towse, A. S. (2021). Opening Pandora’s Box: Peeking inside Psychology’s data sharing practices, and seven recommendations for change. *Behavior Research Methods*, *53*(4), 1455–1468. <https://doi.org/10.3758/s13428-020-01486-1>

Wild, H., Kyröläinen, A.-J., & Kuperman, V. (2022). How representative are student convenience samples? A study of literacy and numeracy skills in 32 countries. *PLOS ONE*, *17*(7), e0271191. <https://doi.org/10.1371/journal.pone.0271191>

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. <https://doi.org/10.1017/S0140525X20001685>