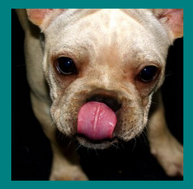
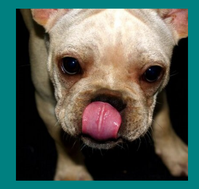


escaping batpigday

Charles T. Gray[1]

[1]Department of Mathematics and Statistics, La Trobe University, Victoria, Australia



batpigday

batpigday *noun* The coding equivalent of groundhogday.

the problem

Simulating data is a bitch.

Debugging frequently dominates the time of students in mathematical science. These students know how to solve equations, and next to nothing about code.

New tools[1] are emerging daily to enable researchers to avoid these timesink pitfalls.
These tools have lowered the programmatic barrier for researchers, but it still a learning curve.

We consider a case study in meta-analysis.

meta-analysis Statistical methodology for combining the results of several studies.

meta-analysis of medians

Conventional meta-analytic tools, such as `metafor::rma`[2], require an **effect** and a **variance** of that effect.

But what if the reported statistics are **median** and **interquartile** range? Existing estimators, such as [3], estimate a mean and standard deviation.

To test our proposed estimator for the variance of the sample median, I found myself repeating tasks and checks in the algorithms.

I tried to find a better way of debugging and writing simulations.
This lead to:

1. a packaged analysis[4], `varameta::*`, which is built on
2. the simulation package for meta-analysis data, `metasim::*`.

(*in development)

coding is the easiest part of coding

- Modular code with functions rather than script
- Reproducibility is more than `set.seed()`
- Versioning and collaboration via Git
- Packaged analyses

escaping batpigday

Generate sample sizes for k studies.

```
library(tidyverse)
library(metasim)
```

```
# simulate 2 studies where most have at most 25
sim_n(k = 2, min_n = 10, max_n = 25) %>% output_table()
```

Table 1:

study	group	n
study_1	control	18
study_2	control	15
study_1	intervention	14
study_2	intervention	15

```
# generate simulation dataframe
sim_df() %>% head(2) %>% select(-n) %>% output_table()
```

Table 2:

k	tau2_true	median_ratio	prop	rdist	parameters	id	true_effect
3	0	1	0.3	norm	list(mean = 67, sd = 0.3)	sim_1	67.0
3	0	1	0.3	exp	list(rate = 2)	sim_2	0.3

Each **row** of this dataframe represents a set of **simulation** parameters. Each simulation runs a **trial** function.

```
metatrial() %>% output_table()
```

Table 3:

conf_low	conf_high	tau_sq	k	effect	measure	true_effect	coverage	bias
19.1	95.9	238.9	3	57.5	m	50.0	TRUE	7.5
-1.3	1.2	0.3	3	-0.1	lr	0.2	TRUE	-0.2

Each **simulation** reruns the trial function a given number of times.

```
metasim() %>% pluck("results") %>%
  select(-coverage_count) %>% output_table()
```

Table 4:

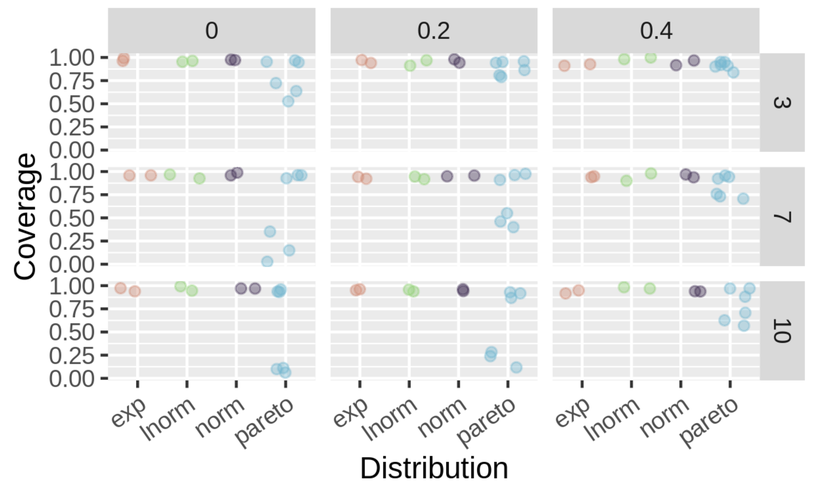
measure	tau_sq	ci_width	bias	successful_trials	coverage	id
lr	0.5	3.6	0.0	4	1	simulation1
m	388.9	94.1	3.4	4	1	simulation1

For all **simulations**, run `metasim` over each row of the dataframe.

```
sims <- metasims(trials = 100,
         trial_fn = metatrial,
         probar = FALSE) %>%
  filter(measure == "lr")

sims %>%
  select(id, k, rdist, coverage, ci_width, bias) %>%
  head() %>%
  output_table()
```

```
# plot
sims %>%
  ggplot(aes(x = rdist, y = coverage)) +
  geom_point(aes(colour = rdist), alpha = 0.4, position = "
  facet_grid(k ~ tau2_true) + theme(
    axis.text.x = element_text(angle = 35, hjust = 1),
    legend.position = "none",
    plot.caption = element_text(hjust = 0)
  ) +
    hrbrthemes::scale_colour_ipsum() +
  labs(x = "Distribution",
       y = "Coverage")
```



This poster was created with `posterdown::Thorne` [5].

References

[1] Hadley Wickham. *tidyverse: Easily Install and Load the 'Tidyverse'*. 2017. URL: https://CRAN.R-project.org/package=tidyverse.

[2] Wolfgang Viechtbauer. "Conducting meta-analyses in R with the metafor package". In: *Journal of Statistical Software* 36.3 (2010), pp. 1–48. URL: http://www.jstatsoft.org/v36/i03/.

[3] Xiang Wan et al. "Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range". In: *BMC Medical Research Methodology* 14.1 (Dec. 19, 2014), p. 135. ISSN: 1471-2288. DOI: 10.1186/1471-2288-14-135. URL: https://doi.org/10.1186/1471-2288-14-135 (visited on 02/10/2019).

[4] Ben Marwick, Carl Boettiger, and Lincoln Mullen. "Packaging Data Analytical Work Reproducibly Using R (and Friends)". In: *The American Statistician* 72.1 (Jan. 2, 2018), pp. 80–88. ISSN: 0003-1305. DOI: 10.1080/00031305.2017.1375986. URL: https://doi.org/10.1080/00031305.2017.1375986 (visited on 11/24/2018).

[5] W. Brent Thorne. *posterdown: An R Package*. 2019. URL: https://github.com/brentthorne/posterdown.