

Project planning: From design to paper and data release

Shravan Vasishth/Daniel Schad

November 08, 2019

Research/teaching interests in my lab

- Experimental psycholinguistics
- Computational modeling (sentence processing)
- Disseminating statistical methods (principally, Bayesian) in psycholinguistics (courses and tutorial articles)
- Open and transparent science

Materials from this workshop are available on github

Github source: <https://github.com/vasishth/MPILeipzig2019>

Web page: <https://vasishth.github.io/MPILeipzig2019/>

Motivation

Important problems in all areas of science

Too often, “findings” in published papers are:

- **non-replicable**

Replication attempt of Dillon et al 2013 (JML): <https://osf.io/reavs/>

Replication attempt of Levy and Keller 2013 (JML):

<https://osf.io/eyphj/>

- **non-reproducible**

<https://royalsocietypublishing.org/doi/full/10.1098/rsos.180448>

“...suboptimal data curation, unclear analysis specification and reporting errors can impede analytic reproducibility, undermining the utility of data sharing and the credibility of scientific findings.”

- **contain (serious) mistakes**

<https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02210/full>

- **impossible to use for meta-analysis**

<https://osf.io/dcbfz/>

<https://osf.io/g5ndw/>

Motivation

The underlying causes of these problems

- lack of statistical training in psych/ling
- experiments are run without checks or planning
- the result is a chaotic research process
- there is a culture of not releasing data with the paper

Actual data release statements:

“Aggregate measures are available to qualified researchers upon written request”

“The raw data supporting the conclusions of this manuscript will be made available to any qualified researcher upon request.”

- errors in analyses and coding waste researcher's time and energy

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

- The research question is whether object relatives (ORs) are easier to process than subject relatives (SRs), in Mandarin Chinese.
- In English, SRs are **easier** to process than ORs:

SR: The senator who interviewed the journalist resigned.

OR: The senator who the journalist interviewed resigned.

- Gibson and Wu 2013, and Hsiao and Gibson 2003, have argued that Chinese shows the opposite pattern (ORs easier). (Some background here: <https://tinyurl.com/y69nqnnyt>).

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

Existing data on Chinese relative clauses: Gibson and Wu 2013.

```
## load and preprocess data:  
chineseRC<-read.table("data/gibsonwu2012data.txt")  
crit<-subset(chineseRC,region=="headnoun")  
crit$region<-factor(crit$region)  
crit<-crit[,c(1,2,3,7)]
```

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

```
## the data frame:
```

```
head(crit)
```

```
##      subj item      type    rt
## 94      1   13  obj-ext 1561
## 221     1    6  subj-ext  959
## 341     1    5  obj-ext  582
## 461     1    9  obj-ext  294
## 621     1   14  subj-ext  438
## 753     1    4  subj-ext  286
```

```
## slight imbalance in design:
```

```
head(xtabs(~subj+type,crit),n=2)
```

```
##      type
## subj obj-ext subj-ext
##    1      8      7
##    2      7      8
```


Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

partially crossed, Latin square:

```
head(xtabs(~subj+item,crit),n=2)
```

```
##      item
## subj 1 2 3 4 5 6 7 8 9 10 11 13 14 15 16
##      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##      2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

```
## sum contrasts:
crit$so<-ifelse(crit$type=="obj-ext",1,-1)
library(lme4)

## Loading required package: Matrix

## "maximal" LMM (ignores convergence issues):
m<-lmer(log(rt)~so+(1+so|subj)+
        (1+so|item),
        crit,control=lmerControl(calc.derivs=FALSE))
```

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

Results:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subj	(Intercept)	5.99e-02	0.244814	
	so	3.54e-03	0.059512	-1.00
item	(Intercept)	3.32e-02	0.182104	
	so	4.74e-08	0.000218	0.85
Residual		2.65e-01	0.514322	

Number of obs: 547, groups: subj, 37; item, 15

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.0618	0.0657	92.2
so	-0.0362	0.0242	-1.5

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

Underlying model:

$$y_{kj} \sim \text{LogNormal}(\alpha + u_{0j} + w_{0k} + (\beta + u_{1j} + w_{1k}) * so_{kj}, \sigma) \quad (1)$$

where we have variance components: σ and

$$\Sigma_u = \begin{pmatrix} \sigma_{u0}^2 & \rho_u \sigma_{u0} \sigma_{u1} \\ \rho_u \sigma_{u0} \sigma_{u1} & \sigma_{u1}^2 \end{pmatrix} \quad \Sigma_w = \begin{pmatrix} \sigma_{w0}^2 & \rho_w \sigma_{w0} \sigma_{w1} \\ \rho_w \sigma_{w0} \sigma_{w1} & \sigma_{w1}^2 \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u \right), \quad \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_w \right) \quad (3)$$

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

Generate fake data to compute power:

```
## load fake data simulation code:  
source("R/gen_fake_lnorm.R")
```

Experiment design and planning

Example: power calculation

```
## extract estimates of fixed-effects parameters:  
beta<-summary(m)$coefficients[,1]  
## extract standard deviation estimate:  
sigma_e<-attr(VarCorr(m),"sc")  
## assemble variance covariance matrix for subjects:  
subj_ranefsd<-attr(VarCorr(m)$subj,"stddev")  
subj_ranefcorr<-attr(VarCorr(m)$subj,"corr")  
## choose some intermediate values for correlations:  
corr_matrix<-(diag(2) + matrix(rep(1,4),ncol=2))/2  
Sigma_u<-SIN::sdcor2cov(stddev=subj_ranefsd,  
                        corr=corr_matrix)
```

Experiment design and planning

Example: power calculation

```
## assemble variance covariance matrix for items:  
item_ranefsd<-attr(VarCorr(m)$item,"stddev")  
Sigma_w<-SIN::sdcor2cov(stddev=item_ranefsd,  
                        corr=corr_matrix)
```

Experiment design and planning

Example: power calculation

```
nsim<-100
tvals<-c()
for(i in 1:nsim){
  fakedat<-gen_fake_lnorm(nitem=16,nsbj=40,
                          alpha=beta[1],beta=beta[2],
                          Sigma_u=Sigma_u,Sigma_w=Sigma_w,
                          sigma_e=sigma_e)
  m<-lmer(log(rt)~so+(1+so|subj)+(1+so|item),
          fakedat,
          control=lmerControl(calc.derivs=FALSE))
  tvals[i]<-summary(m)$coefficients[2,3]
}
```


Experiment design and planning

Example: power calculation

```
## prospective power for nsubj=40, nitem=16:  
mean(abs(tvals)>2)  
  
## [1] 0.25
```

Experiment design and planning

Example: power calculation

Write a function for computing power:

```
compute_power<-function(nsim=100,b=-0.03625,
                        nsubj=40,nitem=16){
  tvals<-c()
  for(i in 1:nsim){
    fakedat<-gen_fake_lnorm(nitem=nitem,nsubj=nsubj,
                           alpha=beta[1],beta=b,
                           Sigma_u=Sigma_u,Sigma_w=Sigma_w,sigma_e=sigma_e)
    m<-lmer(log(rt)~so+(1+so|subj)+(1+so|item),
            fakedat,
            control=lmerControl(calc.derivs=FALSE))
    tvals[i]<-summary(m)$coefficients[2,3]
  }
  mean(abs(tvals)>2)
}
```

Experiment design and planning

Example: power calculation

- Conclusion: If we were to run a new experiment on Chinese RCs with 40 subjects and 16 items, our prospective power is approximately 30%.
- This assumes that the parameter estimates from the original data are the true values.
- Note that we are uncertain about the effect estimate (and also the variance components).

Experiment design and planning

Example: power calculation

Statistical power depends on

- sample size (subjects and items)
- standard deviations (of residual and of random effects)
- true effect size

For a full analysis, we should take these uncertainties into account. See appendix in Jäger et al., 2017 (in the folder readings).

Experiment design and planning

Example: power calculation

How to take uncertainty about the effect into account? An example:

If we have ± 1 sum coding, with OR coded $+1$ and SR -1 , then:

- Mean SR processing time (log ms): $\mu_1 = \beta_0 - \beta_1$
- Mean OR processing time (log ms): $\mu_2 = \beta_0 + \beta_1$
- Since the model is on the log scale, we can go back to the ms scale by exponentiating:

Mean difference in ms in OR vs SR: $\exp(\mu_2) - \exp(\mu_1)$

Experiment design and planning

Example: power calculation

Here is the model again:

```
m<-lmer(log(rt)~so+(1+so|subj)+(1+so|item),  
        crit,control=lmerControl(calc.derivs=FALSE))
```

Experiment design and planning

Example: power calculation

```
## estimated OR-SR difference in ms:  
exp(beta[1]+beta[2])-exp(beta[1]-beta[2])  
  
## (Intercept)  
##      -31.12
```

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

what is the slope for an effect of 20 ms?

```
exp(beta[1]+(-0.024))-exp(beta[1]-(-0.024))
```

```
## (Intercept)
```

```
## -20.601
```

what is the slope for an effect of 50 ms?

```
exp(beta[1]+(-0.06))-exp(beta[1]-(-0.06))
```

```
## (Intercept)
```

```
## -51.528
```


Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

Minimum power:

```
## there will be random fluctuation
```

```
compute_power(b=-0.024)
```

```
## [1] 0.12
```

```
compute_power(b=-0.024)
```

```
## [1] 0.14
```

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

Maximum power:

```
## there will be random fluctuation
```

```
compute_power(b=-0.06)
```

```
## [1] 0.72
```

```
compute_power(b=-0.06)
```

```
## [1] 0.73
```

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

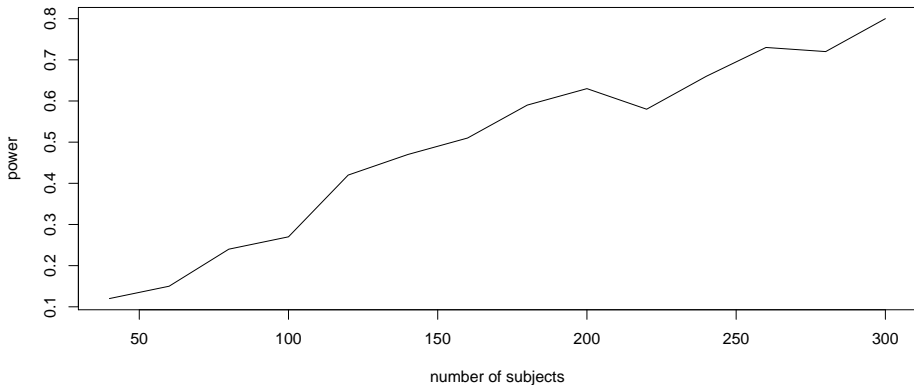
We are now in a position to figure out, for a given effect size, what sample size we need (number of subjects/items) to obtain 80% power or higher. Let's take the minimum effect size to compute a power curve.

```
n<-seq(40,300,by=20)
pow<-rep(NA,length(n))
for(i in 1:length(n)){
  #print(n[i])
  pow[i]<-compute_power(b=-0.024,
                        nsubj=n[i])
}
```

Experiment design and planning

Example: Subject and object relatives in Mandarin Chinese

power curve

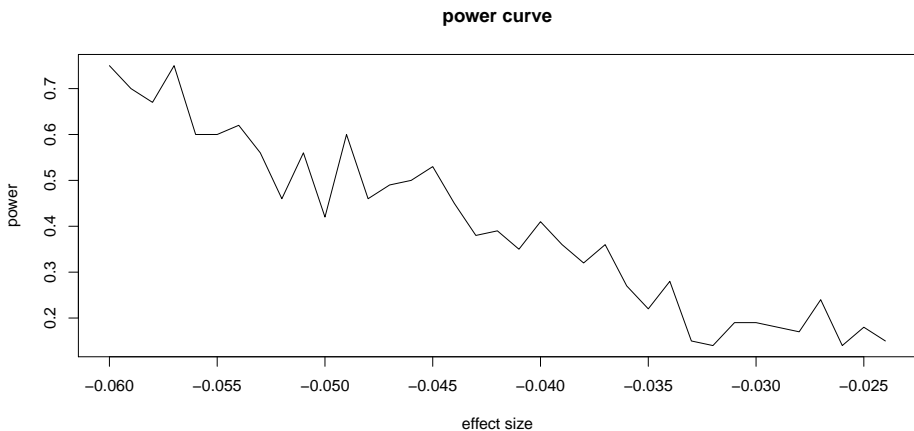


Exercise

Compute power curve as a function of different effect sizes

Exercise

Compute power curve as a function of different effect sizes



Defining the analysis plan using simulated data

Pre-registering the analysis

In 2000, ClinicalTrials.gov made it mandatory to pre-register primary outcomes.

“17 of 30 studies (57%) published prior to 2000 showed a significant benefit of intervention on the primary outcome in comparison to only 2 among the 25 (8%) trials published after 2000 ($\chi^2=12.2, df=1, p=0.0005$).

Pre-registration in clinical trials.gov was strongly associated with the trend toward null findings.”

See readings: <https://doi.org/10.1371/journal.pone.0132382>

Defining the analysis plan using simulated data

Pre-registering the analysis

What is a pre-registration for?

[See: <https://www.bayesianspectacles.org/a-breakdown-of-preregistration-is-redundant-at-best/>]

Defining the analysis plan using simulated data

Pre-registering the analysis

What a pre-registration is for (based on a twitter thread by Christina Bergmann):

- It allows the **researcher**, to clearly specify their research hypothesis in advance
- It creates documentation of the steps to be taken
- It can improve data management
- Makes a clear distinction between exploratory and confirmatory analyses
- Allows for improvements in design before the experiment is run

Defining the analysis plan using simulated data

Pre-registering the analysis

What a pre-registration is **not** for (again echoing Bergmann):

- It doesn't improve theory (the Garbage-in-garbage-out, GIGO, principle)
- It doesn't guarantee that the study was of a high quality
- It doesn't prevent exploration or p-hacking

Defining the analysis plan using simulated data

Pre-registering the analysis

Confirmatory analysis: Define the following in advance:

- the dependent variable(s) (in eyetracking, we have many choices)
- the region of interest/time window
- exclusion criteria
- the statistical model
- expected results using fake-data simulation

None of this prevents you from doing an exploratory analysis. See: - <https://osf.io/mmr7s/> - <https://psyarxiv.com/2atr/>

Defining the analysis plan using simulated data

Pre-registering the analysis

Example pre-registration: Expt 7 in

<https://osf.io/eyphj/>

One can also submit one's pre-registration at: aspredicted.org

Defining the analysis plan using simulated data

Pre-registering the analysis

Switch to slides on pre-registration.

Check that your experiment software actually collects the data you need and back-up your data

- Many people do not check whether their experimental software is actually collecting data.
- The software repeatedly crashes during live experimental runs
- They only find out after the experiment was done.
- *Solution:* Before running the experiment for real, run the study once with one subject per group (in a Latin square design), and analyze the data fully.

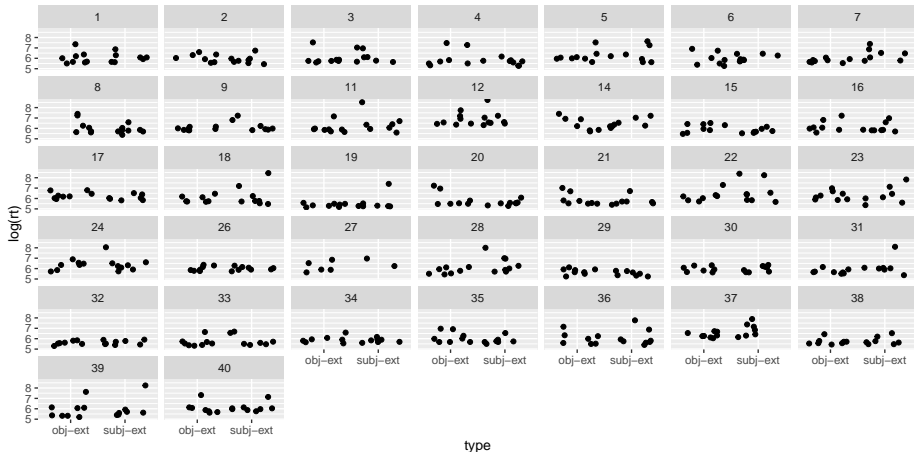
Once data are collected, visualize and summarize the data before doing any analysis

Example: Chinese RCs data

```
library(ggplot2)
p <- ggplot(crit, aes(x=type,
                      y=log(rt))) +
  geom_point(position="jitter") +
  facet_wrap(~ subj, nrow=6)
```

Once data are collected, visualize the data before doing any analysis

Example: Chinese RCs data



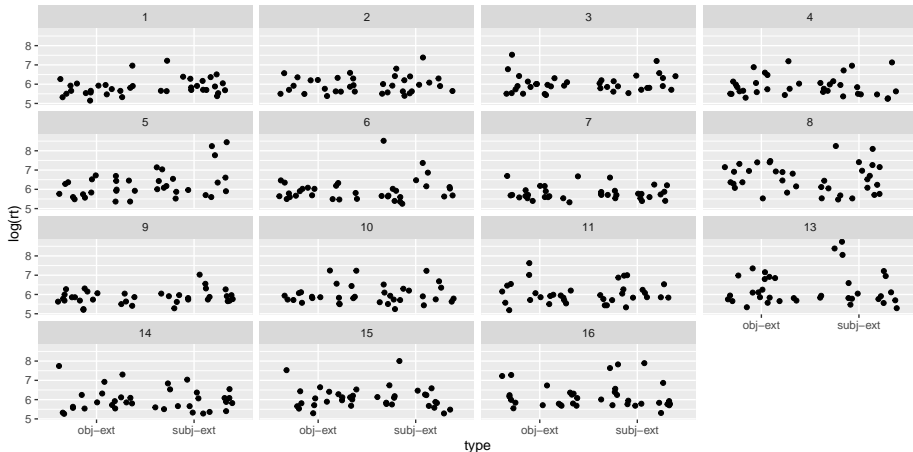
Once data are collected, visualize the data before doing any analysis

Example: Chinese RCs data

```
p2 <- ggplot(crit, aes(x=type,  
                      y=log(rt))) + geom_point(position="jitter")  
  facet_wrap( ~ item, nrow=4)
```

Once data are collected, visualize the data before doing any analysis

Example: Chinese RCs data



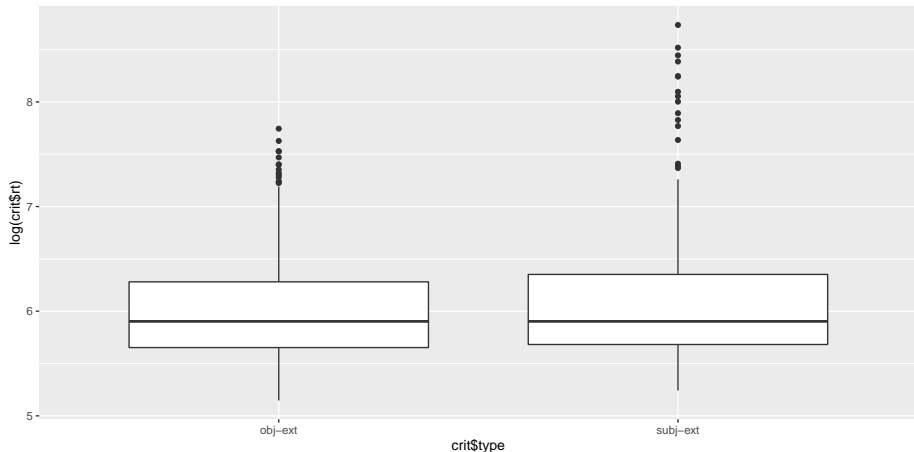
Once data are collected, visualize the data before doing any analysis

Example: Chinese RCs data

```
p3 <- qplot(crit$type, log(crit$rt), geom="boxplot")
```

Once data are collected, visualize the data before doing any analysis

Example: Chinese RCs data



Integrating the data analysis into the manuscript

Workflow checklist

Before collecting any data:

- conduct power analysis using simulated data
- pre-register an analysis plan and time-stamp it (on osf.io or aspredicted.org)
- run the entire experiment at least once per group and analyze data to check for software errors

After collecting data:

- create a data repository (using [github](https://github.com)/[bitbucket](https://bitbucket.org)/[osf](https://osf.io))
- data analysis:
 - visualize the data first
 - present the pre-registered analysis first
 - then do a separate exploratory analysis
- write code using R Markdown
- release all code and data during review and after publication
- ideally: the paper itself should be the documentation (Rmd or Rnw).

Integrating the data analysis into the manuscript

some high-level coding suggestions

- write functions for common tasks and figure types
- never hard-code variables, this will make code non-robust to changes (e.g., don't write `num_rows<-500`)
- take the time to document code
- refactor code (simplify, write modular code using customized functions)

Integrating the data analysis into the manuscript

Using R Markdown for writing papers

From RMarkdown directory, open:

- `apatemplatepapaja.Rmd`
- `apatemplate.Rmd`

Integrating the data analysis into the manuscript

Creating an R package or vignette

Read this short book: <http://r-pkgs.had.co.nz/>

Creating a package:

```
usethis::create_package("VasishtEtAl2019")
```

Package structure

- DESCRIPTION
- R/
- man/
- docs/
- vignettes/
- data/
- NAMESPACE

Data repositories: osf and github

- osf.io
- github.com