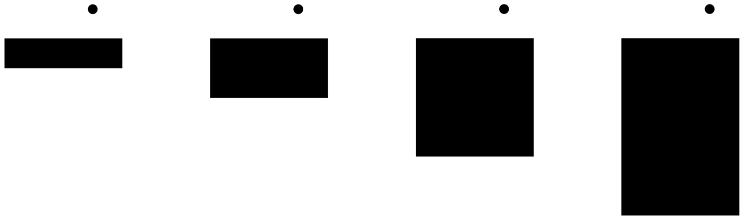
Hooray, significance! So what?

Learning more about your data using Bayesian data analysis

Thomas Kluth

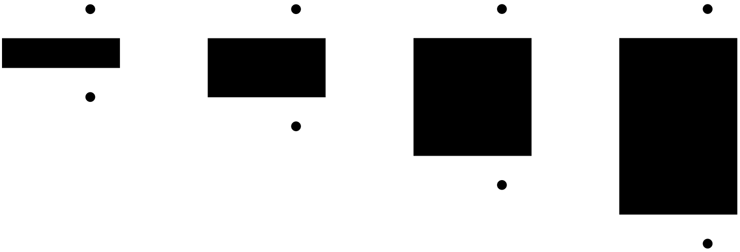
Language & Cognition Group, CITEC, Universität Bielefeld

July 6, 2017, data science meetup Münster



The circle is above the object.

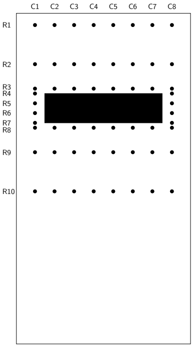
- difference in acceptability?



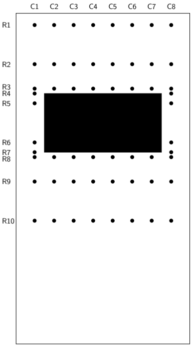
thin rectangle | thick rectangle | square | tall rectangle

The circle is above the object.

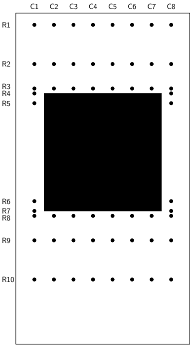
- difference in acceptability? $\longrightarrow$ empirical study
- 4 rectangles $\times$ 2 prepositions



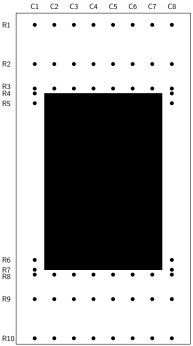
thin rectangle



thick rectangle



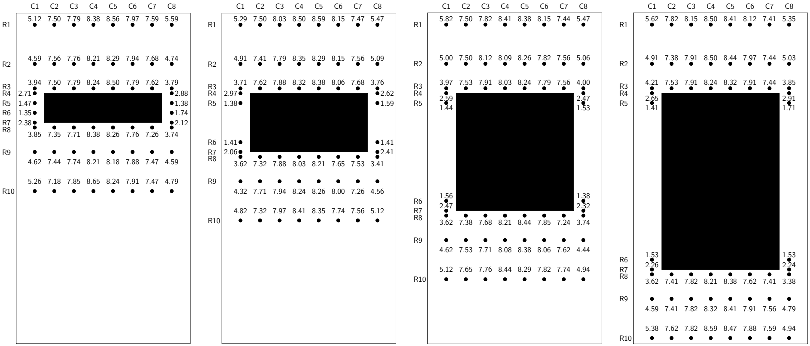
square



tall rectangle

The circle is above the object.

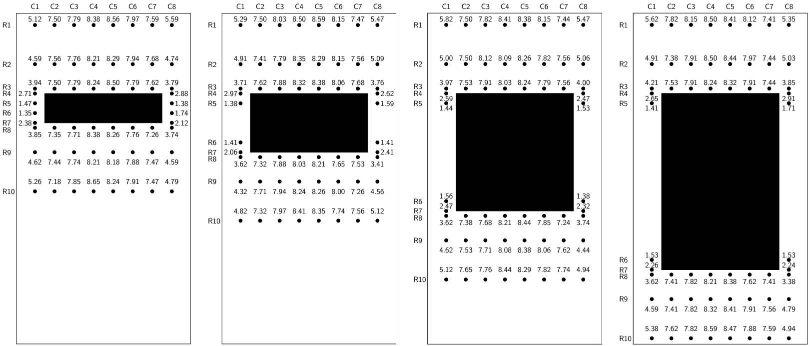
- difference in acceptability? $\longrightarrow$ empirical study
- 4 rectangles $\times$ 2 prepositions $\times$ 28 locations



thin rectangle | thick rectangle | square | tall rectangle

The circle is above the object.

- difference in acceptability? ⟶ empirical study
- 4 rectangles × 2 prepositions × 28 locations × 34 subjects ⟶ 7616 ratings (1–9 rating scale)



thin rectangle | thick rectangle | square | tall rectangle

The circle is above the object.

- difference in acceptability? $\longrightarrow$ empirical study
- 4 rectangles $\times$ 2 prepositions $\times$ 28 locations $\times$ 34 subjects $\longrightarrow$ 7616 ratings (1–9 rating scale)
- prediction[1]: lower ratings for taller rectangles

[1] Kluth, Burigo, Schultheis, and Knoeferle (submitted); Regier (1996); Regier and Carlson (2001)

Null Hypothesis Significance Testing // NHST

- prediction: difference in ratings across rectangles
- NHST $\longrightarrow$ null hypothesis: **no** difference

Null Hypothesis Significance Testing // NHST

- prediction: difference in ratings across rectangles
- NHST $\longrightarrow$ null hypothesis: **no** difference
- 5% significant difference: $<$ 5% probability that data from different rectangles come from the same ("no-difference") distribution

Null Hypothesis Significance Testing // NHST

- prediction: difference in ratings across rectangles
- NHST $\longrightarrow$ null hypothesis: **no** difference
- 5% significant difference: $<$ 5% probability that data from different rectangles come from the same ("no-difference") distribution
    - ***false*** statement: $\geq$ 95% probability that prediction is true

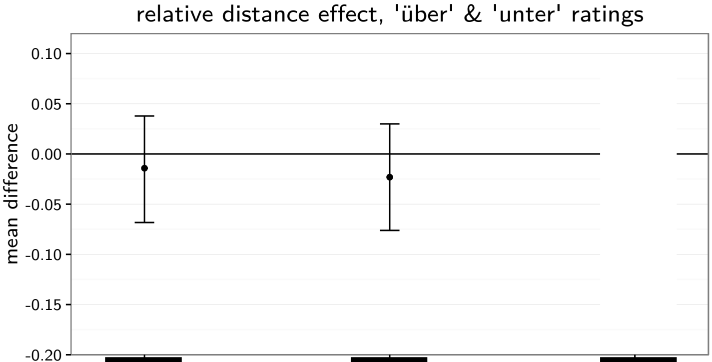
Null Hypothesis Significance Testing // NHST

- prediction: difference in ratings across rectangles
- NHST $\longrightarrow$ null hypothesis: **no** difference
- 5% significant difference: $<$ 5% probability that data from different rectangles come from the same ("no-difference") distribution
  - ***false*** statement: ~~≥ 95% probability that prediction is true~~

Null Hypothesis Significance Testing // NHST

- prediction: difference in ratings across rectangles
- NHST $\longrightarrow$ null hypothesis: **no** difference
- 5% significant difference: $<$ 5% probability that data from different rectangles come from the same ("no-difference") distribution
  - ***false*** statement: ~~$\geq$ 95% probability that prediction is true~~
- paired t-test

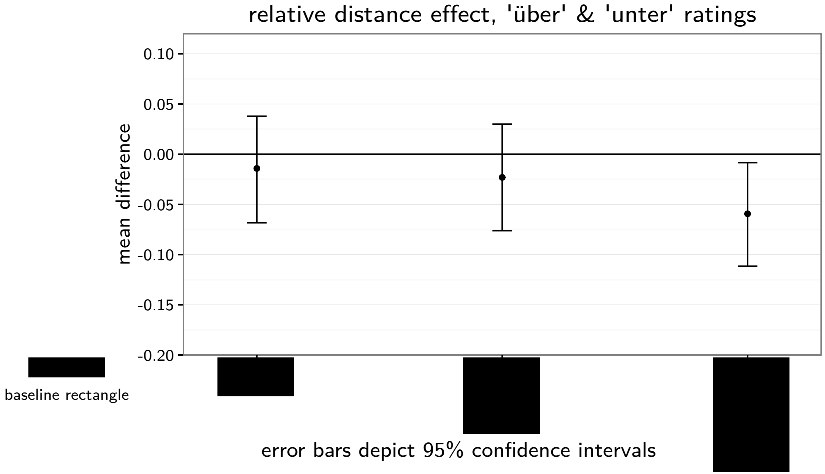
paired t-test // prediction confirmed?



baseline rectangle

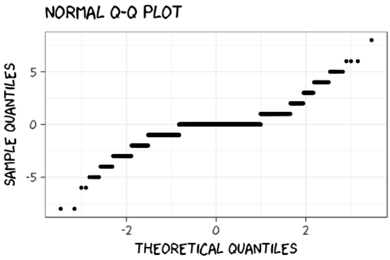
error bars depict 95% confidence intervals

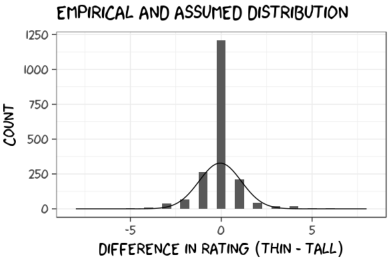
paired t-test // prediction confirmed!



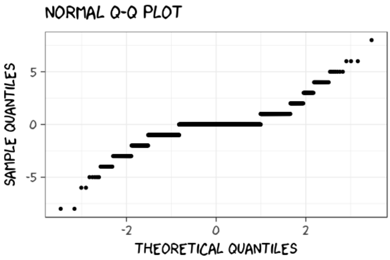
error bars depict 95% confidence intervals

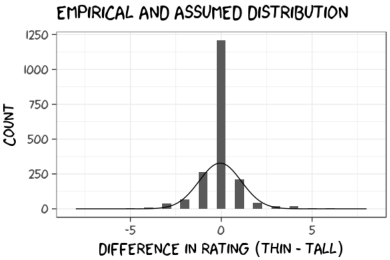
Look at your data! // normality assumption violated!

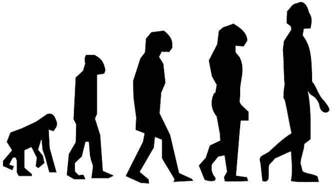




Look at your data! // normality assumption violated!







Bayesian regression model // using `brms`

predict *rating* by *rectangle* (thin, thick, square, tall)

Bayesian regression model // using brms

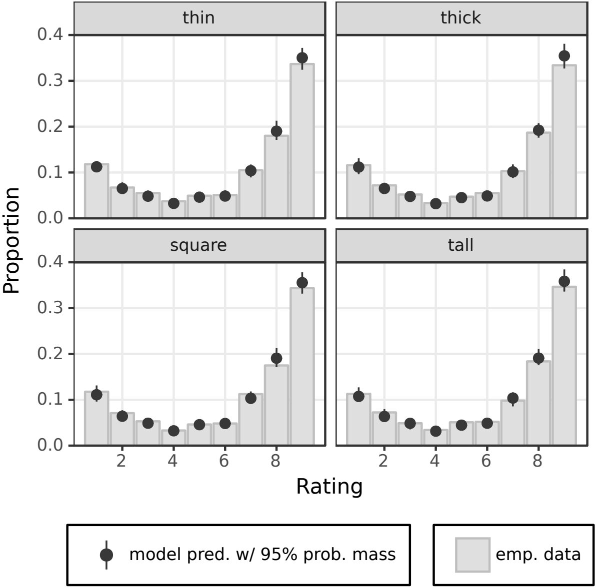
predict *rating* by *rectangle* (thin, thick, square, tall)

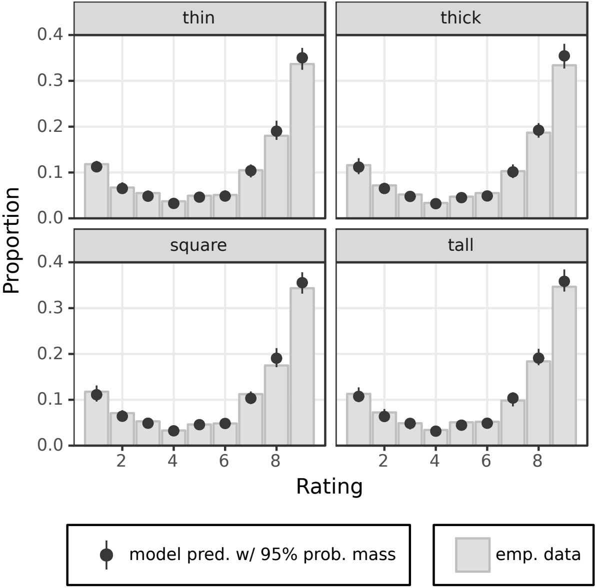
R code

```
regressionModel = brm(
        rating ~ rectangle + (1 | subject),
        family = cumulative(), # ordinal regression
        data = ratingDataFrame)
```

necessary R packages: `brms`, `rstan`

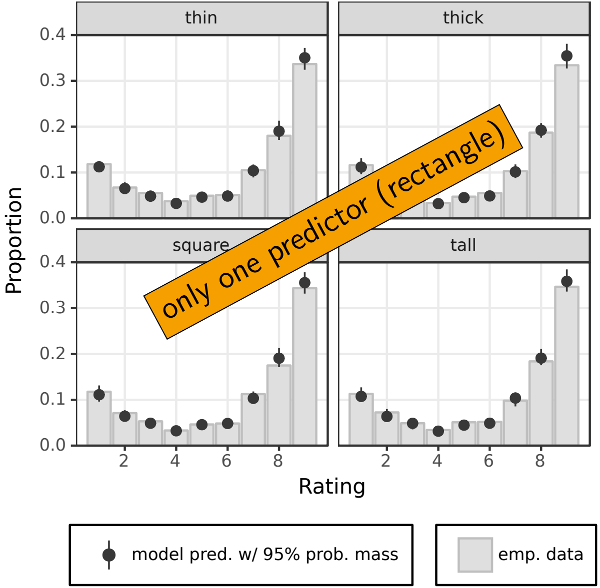
(Bürkner, in press; Stan Development Team, 2016)


thin
thick
square
tall
0.0
0.1
0.2
0.3
0.4
2
4
6
8
Proportion
Rating
model pred. w/ 95% prob. mass
emp. data



no regression slope is credibly different from zero

$\longrightarrow$ no credible difference in ratings $\longrightarrow$ prediction not confirmed



no regression slope is credibly different from zero

$\longrightarrow$ no credible difference in ratings $\longrightarrow$ prediction not confirmed

Learn from your data! // beyond binary answers . . .

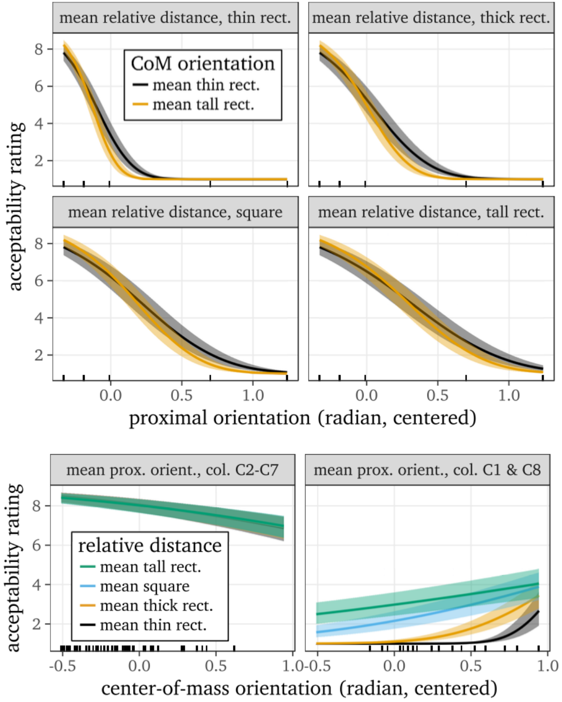
- factors hypothesized to affect rating:
  - center-of-mass orientation
  - proximal orientation
  - relative distance

Learn from your data! // beyond binary answers . . .

- factors hypothesized to affect rating:
  - center-of-mass orientation
  - proximal orientation
  - relative distance

R code

```
regressionModel = brm(
        rating ~ CoMOrient * proxOrient * relDist +
                        (1 | subject),
        family = cumulative(), # ordinal regression
        data = ratingDataFrame)
```



mean relative distance, thin rect.
mean relative distance, thick rect.
mean relative distance, square
mean relative distance, tall rect.
CoM orientation
mean thin rect.
mean tall rect.
acceptability rating
proximal orientation (radian, centered)
mean prox. orient., col. C2-C7
mean prox. orient., col. C1 & C8
relative distance
mean tall rect.
mean square
mean thick rect.
mean thin rect.
center-of-mass orientation (radian, centered)

Why Bayes? // selected reasons

regression more powerful than standard tests, but why *Bayesian* regression?

Why Bayes? // selected reasons

regression more powerful than standard tests, but why *Bayesian* regression?

- Bayesian parameter estimation allows to interpret the **whole probability distribution of regression parameters**

Why Bayes? // selected reasons

regression more powerful than standard tests, but why *Bayesian* regression?

- Bayesian parameter estimation allows to interpret the **whole probability distribution of regression parameters**
- Bayesian statistics allows to **intuitively discuss** the results
  - valid[2] statement: $\geq$ *95% probability that prediction is true*

[2] depends on the results

Why Bayes? // selected reasons

regression more powerful than standard tests, but why *Bayesian* regression?

- Bayesian parameter estimation allows to interpret the **whole probability distribution of regression parameters**
- Bayesian statistics allows to **intuitively discuss** the results
  - valid[2] statement: $\geq$ *95% probability that prediction is true*
- including **prior information** from previous studies directly into the regression analysis is part of the Bayesian framework

[2] depends on the results

Thank you!

list of useful resources follows on the next slides

References

Kluth, T., Burigo, M., Schultheis, H., & Knoeferle, P. (submitted). Does direction matter? Linguistic asymmetries reflected in visual attention. *Cognition*.

Bürkner, P.-C. (in press). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*.

Etz, A., Gronau, Q. F., Dablander, F., Edelsbrunner, P. A., & Baribault, B. (in press). How to become a Bayesian in eight easy steps: An annotated reading list. *Psychonomic Bulletin & Review*.

Kluth, T. (2017). *A C++ implementation of cognitive models of spatial language understanding as well as pertinent empirical data and analyses.* to-be-published on pub.uni-bielefeld.de.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, Mass.: MIT Press.

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*(2), 273–298. doi: 10.1037//0096-3445.130.2.273

Stan Development Team. (2016). *RStan: the R interface to Stan.* (R package version 2.14.1)

Useful resources // selected R packages

- rstan (R interface for STAN), https://cran.r-project.org/package=rstan, STAN is a programming language itself, for more information including interfaces to other languages see mc-stan.org
- brms (Bayesian regression modeling using STAN), well documented, very responsive package author, https://cran.r-project.org/package=brms, Bürkner (in press)
- rstanarm, similar to brms; faster but less flexible, https://cran.r-project.org/package=rstanarm
- bayesplot, provides great visualizations, compatible with brms and rstanarm, https://cran.r-project.org/package=bayesplot
- BEST (Bayesian estimation supersedes the t-test), https://cran.r-project.org/package=BEST
- Bayesian First Aid, http://sumsar.net/blog/2014/01/bayesian-first-aid/

Useful resources // selected tutorials & literature

- tutorial: `https://mvuorre.github.io/post/2017/how-to-compare-two-groups-with-robust-bayesian-estimation-using-r-stan-and-brms/`, based on Kruschke (2013)
- full data set including working and commented R scripts: Kluth (2017); accompanying the article Kluth et al. (submitted); not available yet, but stay tuned, should be online in a few months (check `https://www.techfak.de/~tkluth`)
- books: Kruschke (2015); McElreath (2016)
- annotated reading list: Etz, Gronau, Dablander, Edelsbrunner, and Baribault (in press)