Individual differences in replicated multi-product 2-AFC data with and without supplementary difference scoring: Comparing Thurstonian mixed regression models for binary and ordinal data with linear mixed models

Christine Borgen Linander[1,*], Rune Haubo Bojesen Christensen[1], Rebecca Evans[2], Graham Cleaver[2], Per Bruun Brockhoff[1]

[1]Department of Informatics and Mathematical Modeling, Section for Statistics, Technical University of Denmark
[2]Unilever Research and Development Port Sunlight, UK
*Contact author: chjo@imm.dtu.dk

Outline

Experiment

Data:

- 12 Products - denoted A, B, C, . . . , L
- 24 Assessors
- 8 Attributes
  - 5 different attributes - denoted A, B, C, D, E
  - 3 of those evaluated again after 5 minutes - denoted C.5m, D.5m, E.5m
- 2 Sessions leading to replications

Each Product was compared to a control.

Not all combinations of Assessor and Product are present. Hence incomplete data. (Have included Assessors with at least 50% of the observations)

For each Attribute a maximum of 24 observations for each Assessor.

Experiment

Test protocol:

- 2-AFC protocol
- With an additional task to score the difference from 'not different' to 'extremely different' with 5 categories

The data from the additional task is transformed into a scale from -4 to 4 where:

- 0 corresponds to no difference
- negative values are in favor of the control
- positive values are in favor of the Product

Thus possibility to consider the response in 3 ways:

- Binary with values 0, 1
- Quantitative with values -4, 3, . . . , 3, 4
- Ordinal with values -4, 3, . . . , 3, 4

Aims for the analysis

Analytical aims:

- We want to illustrate that it is possible to fit a Thurstonian mixed model with the binary response
- We want to have the capability to take account of other sources of variation when analysing data from the 2-AFC protocol
  - To be more specific - we want to model the individual assessors
- We want to analyse the results from a set of products in one analysis rather than one product comparison at a time

Business oriented aim:

- We want to investigate if we gain valuable extra information adding the additional task to the usual 2-AFC protocol.

Models

We consider 2 types of models:

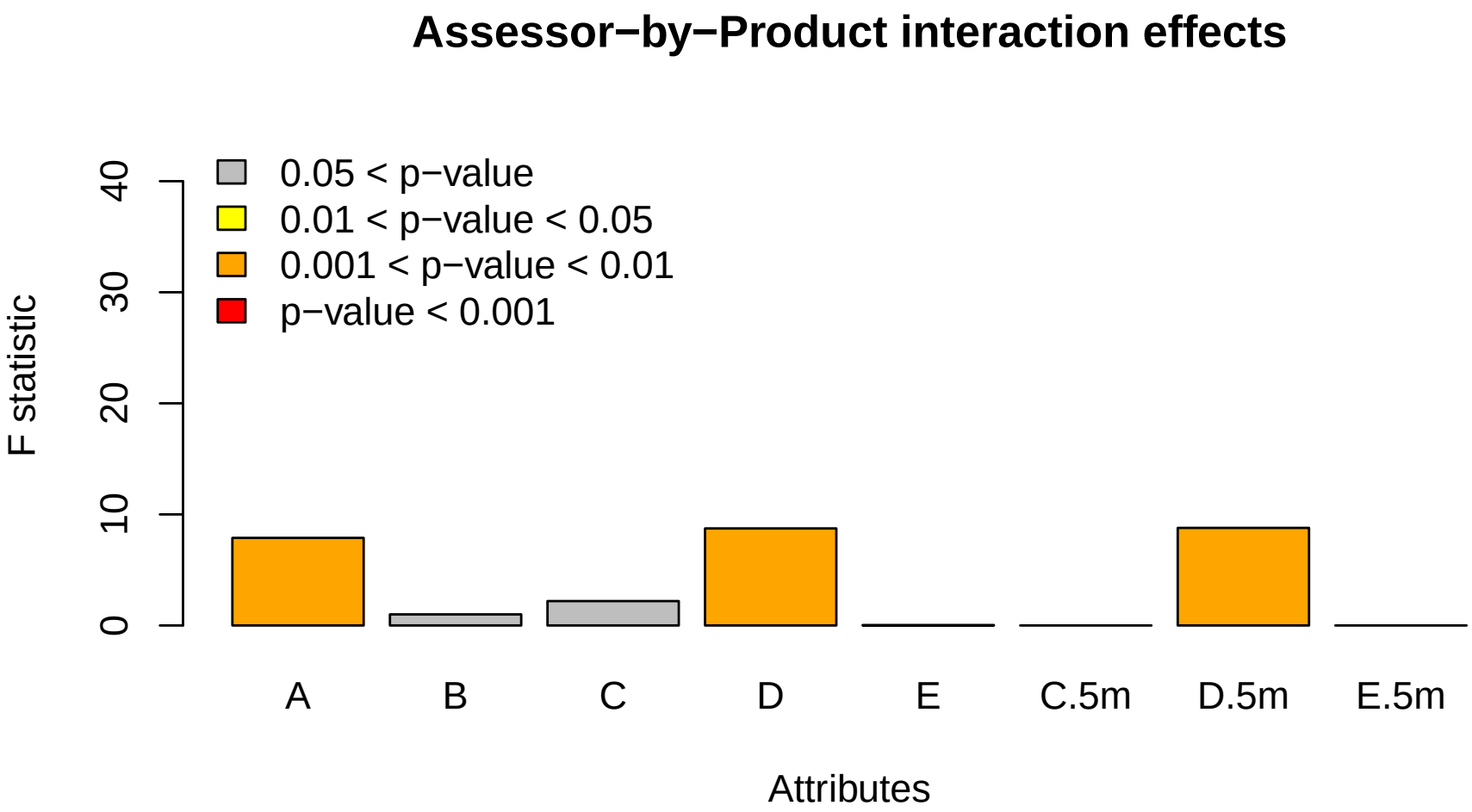
- Naive aggregation model - ignoring the replicated structure
- Mixed model - including Assessor as a random effect, modelling the replicated structure

Where the naive model will be fitted with the binary response.

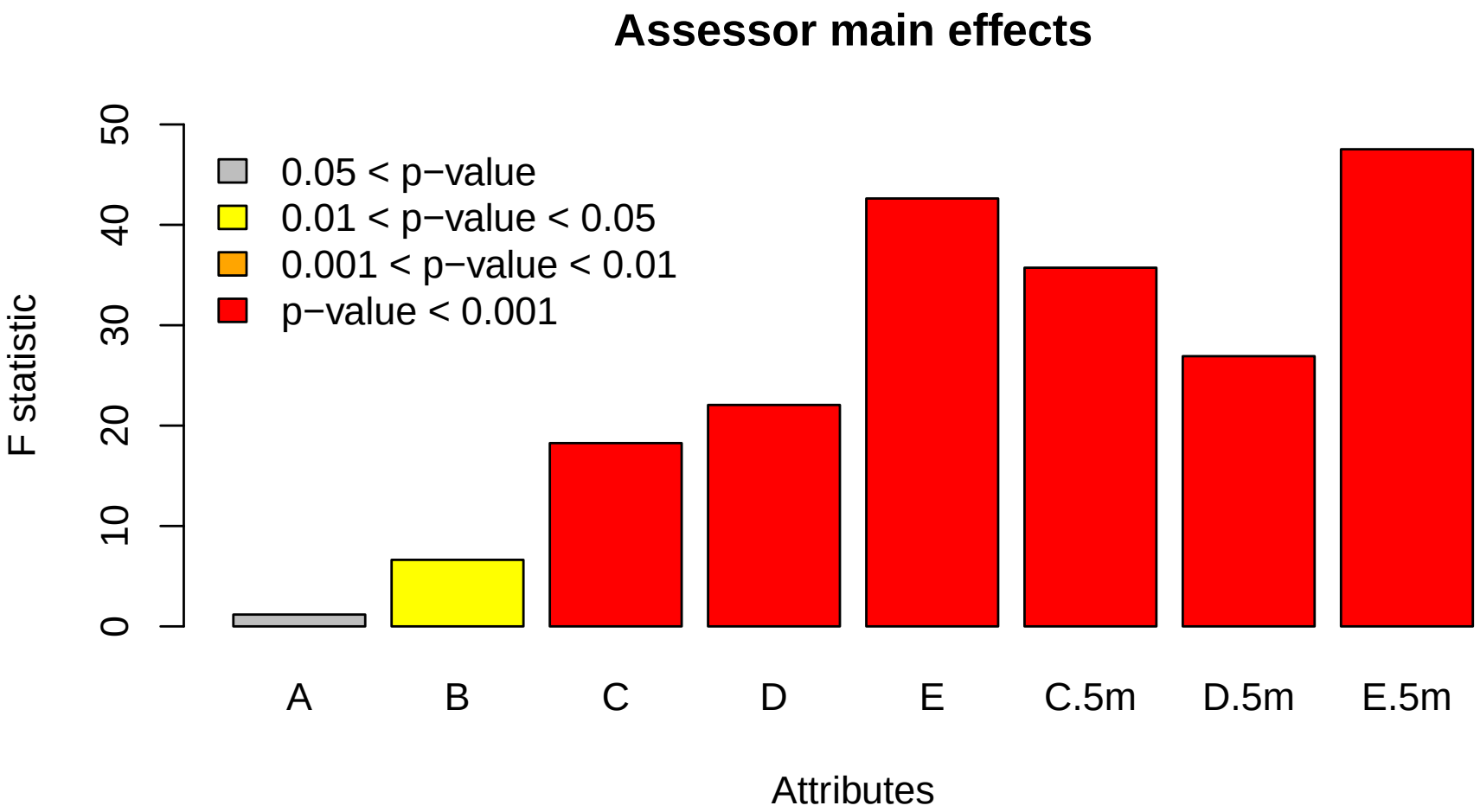
And the mixed model will be fitted with the 3 different responses.

F test - mixed model - Quantitative response

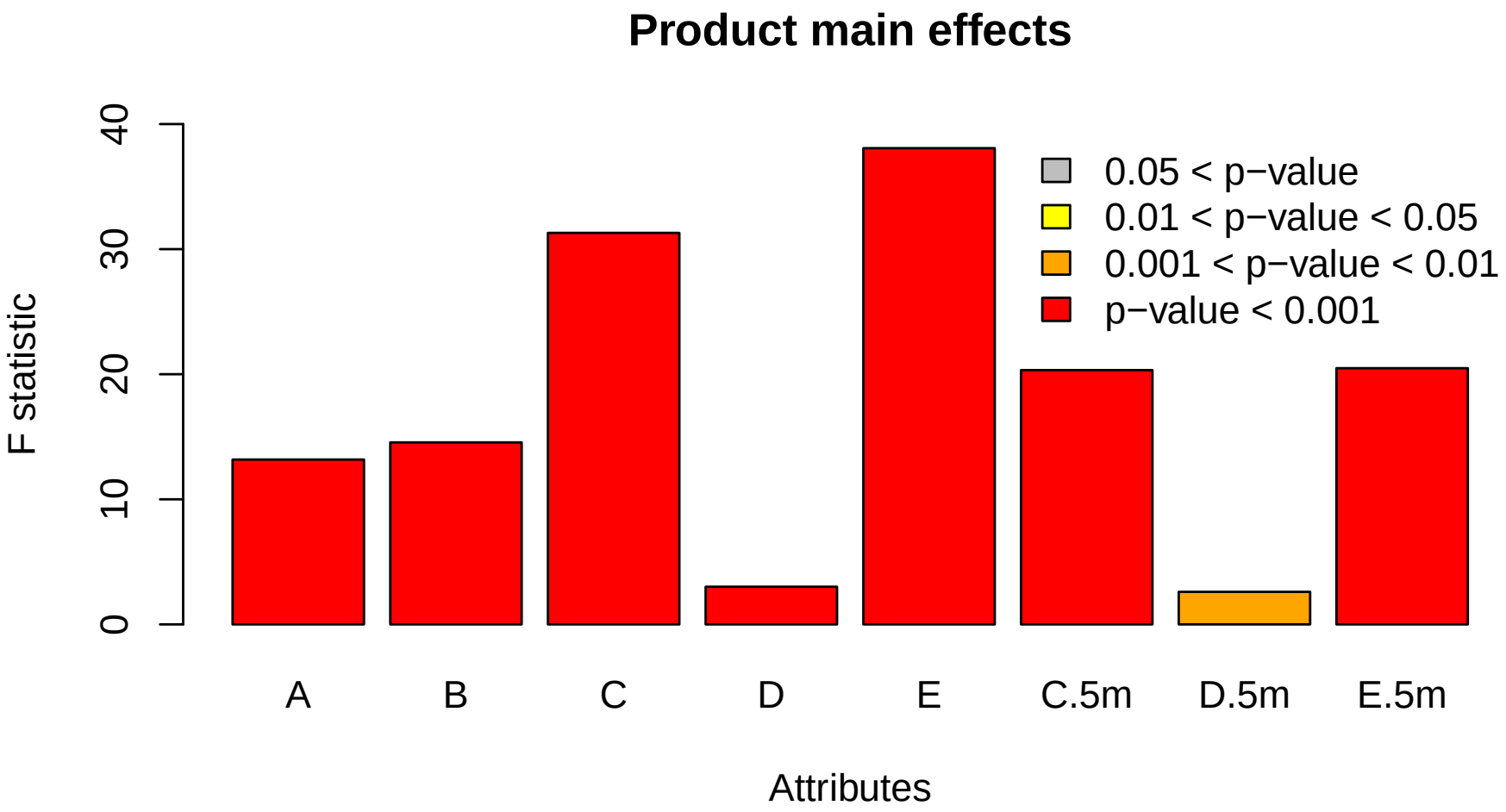
The PanelCheck way by the R-package **lmerTest**



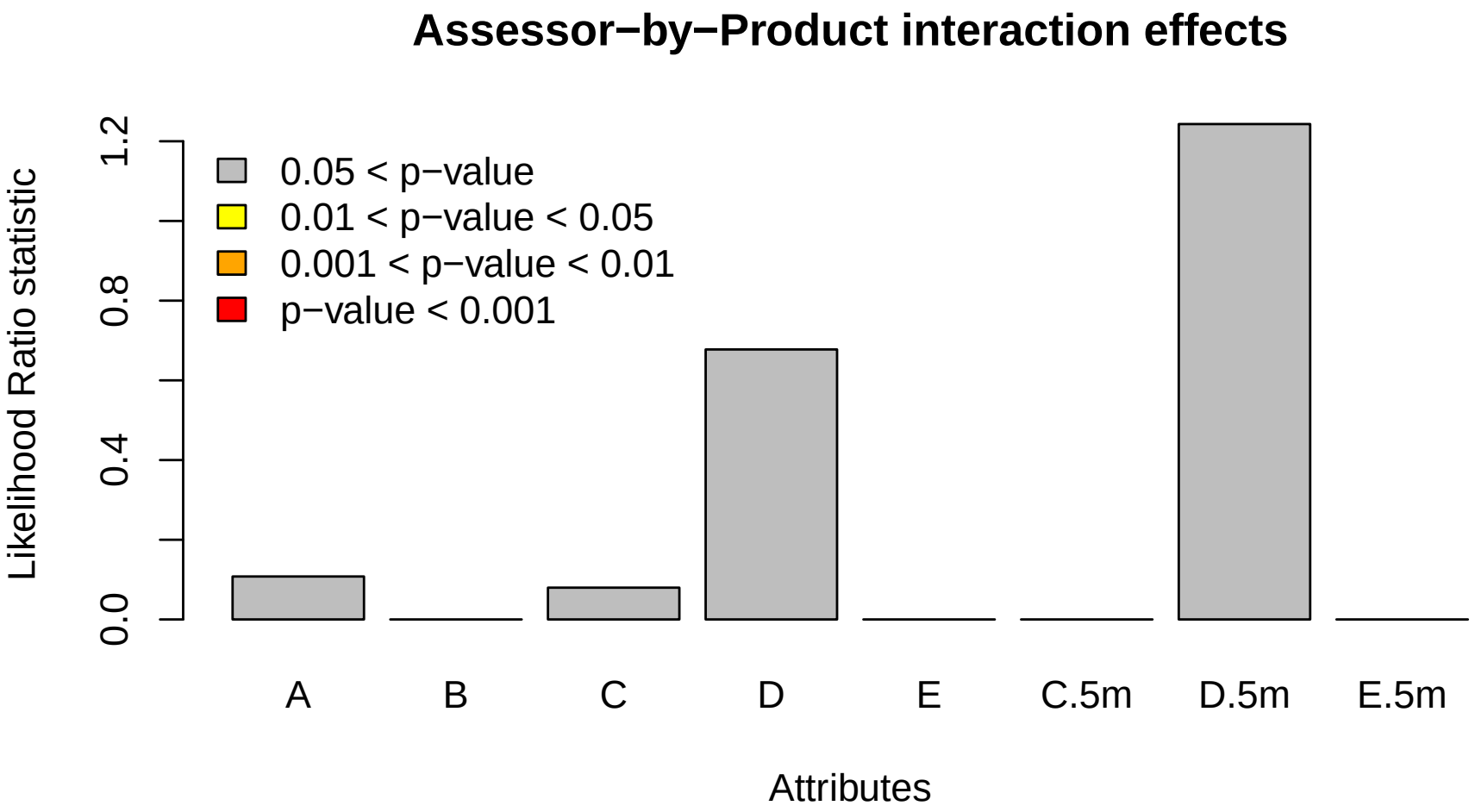
F test - mixed model - Quantitative response



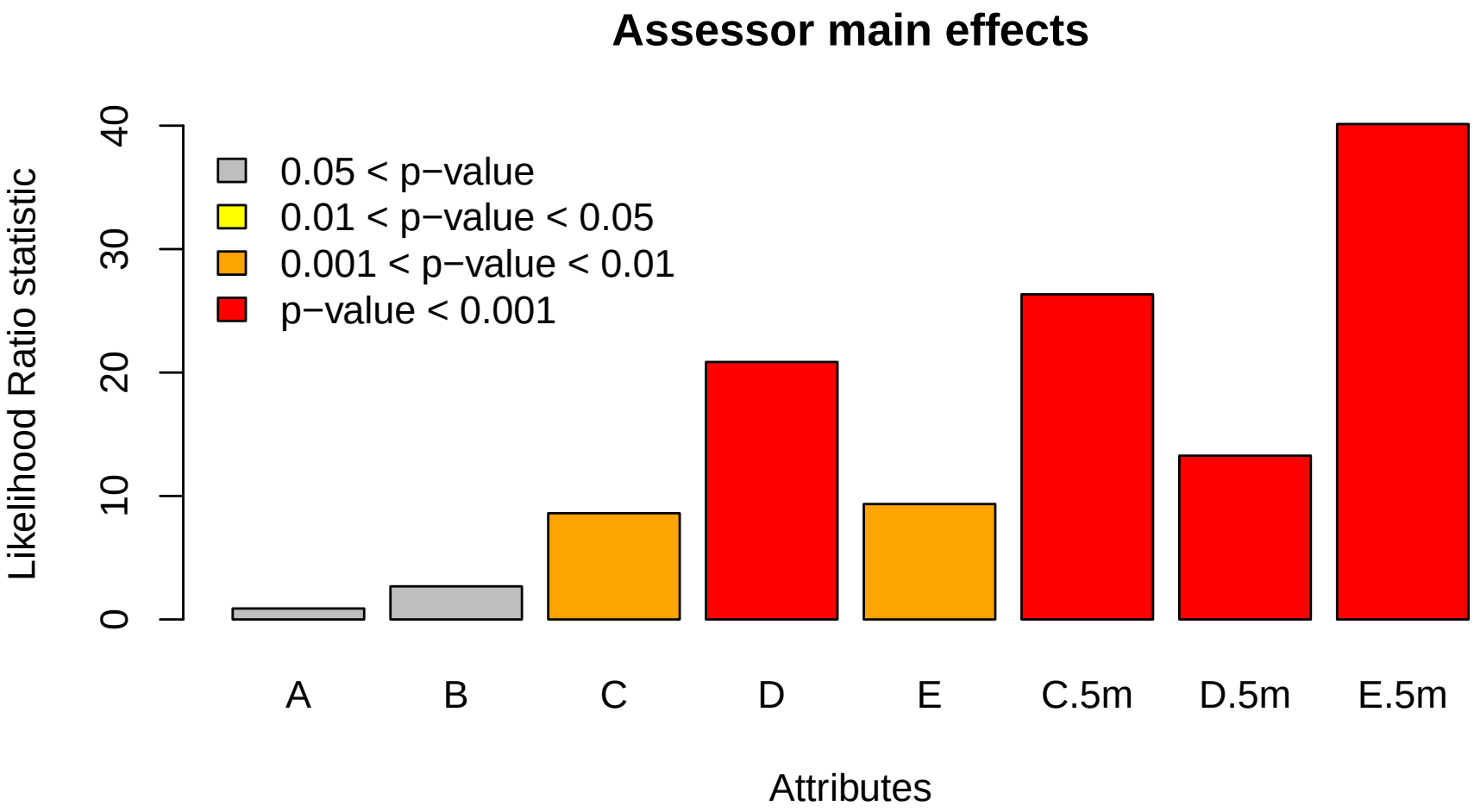
F test - mixed model - Quantitative response



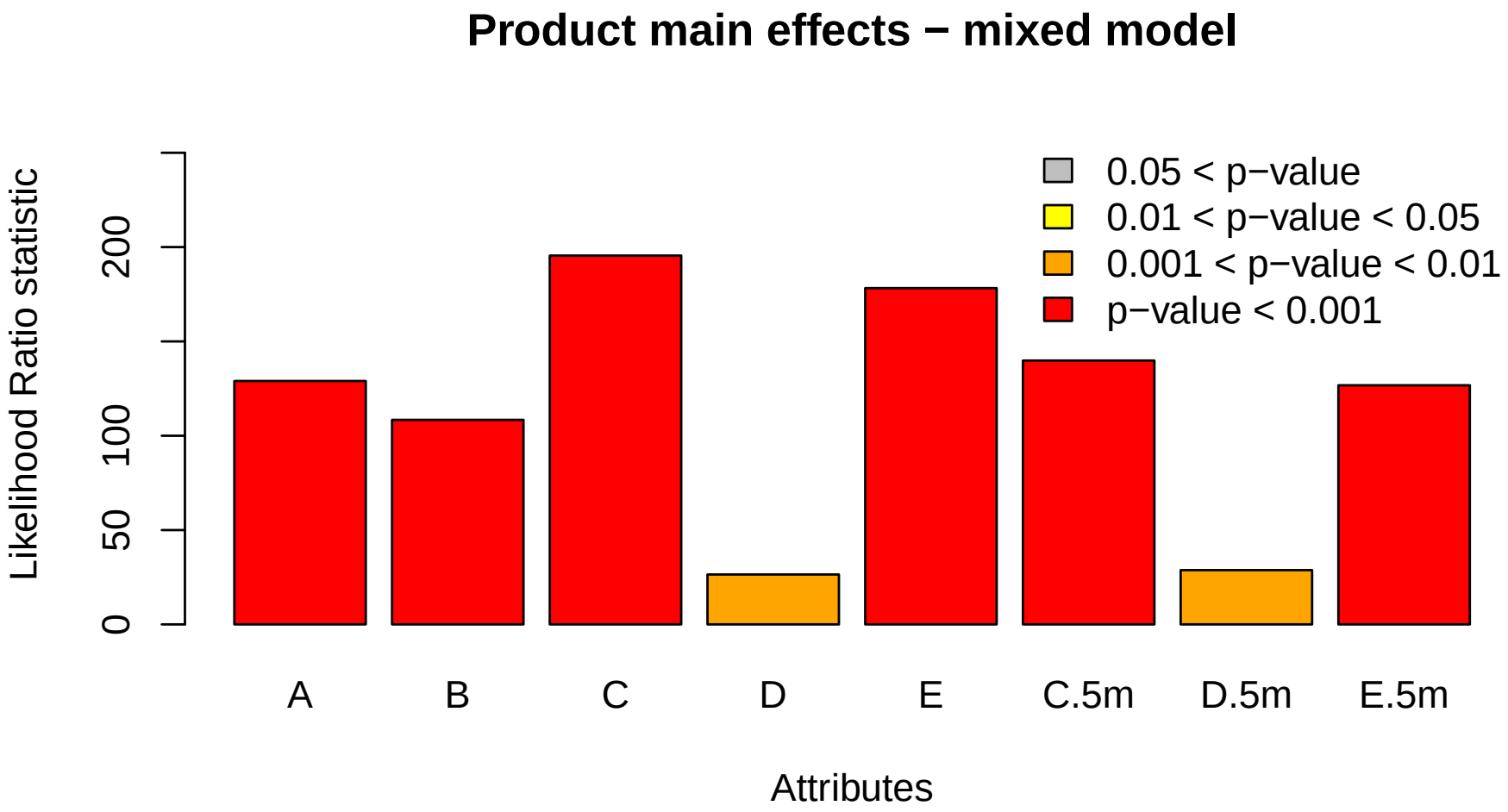
Likelihood Ratio test - mixed model - Binary response



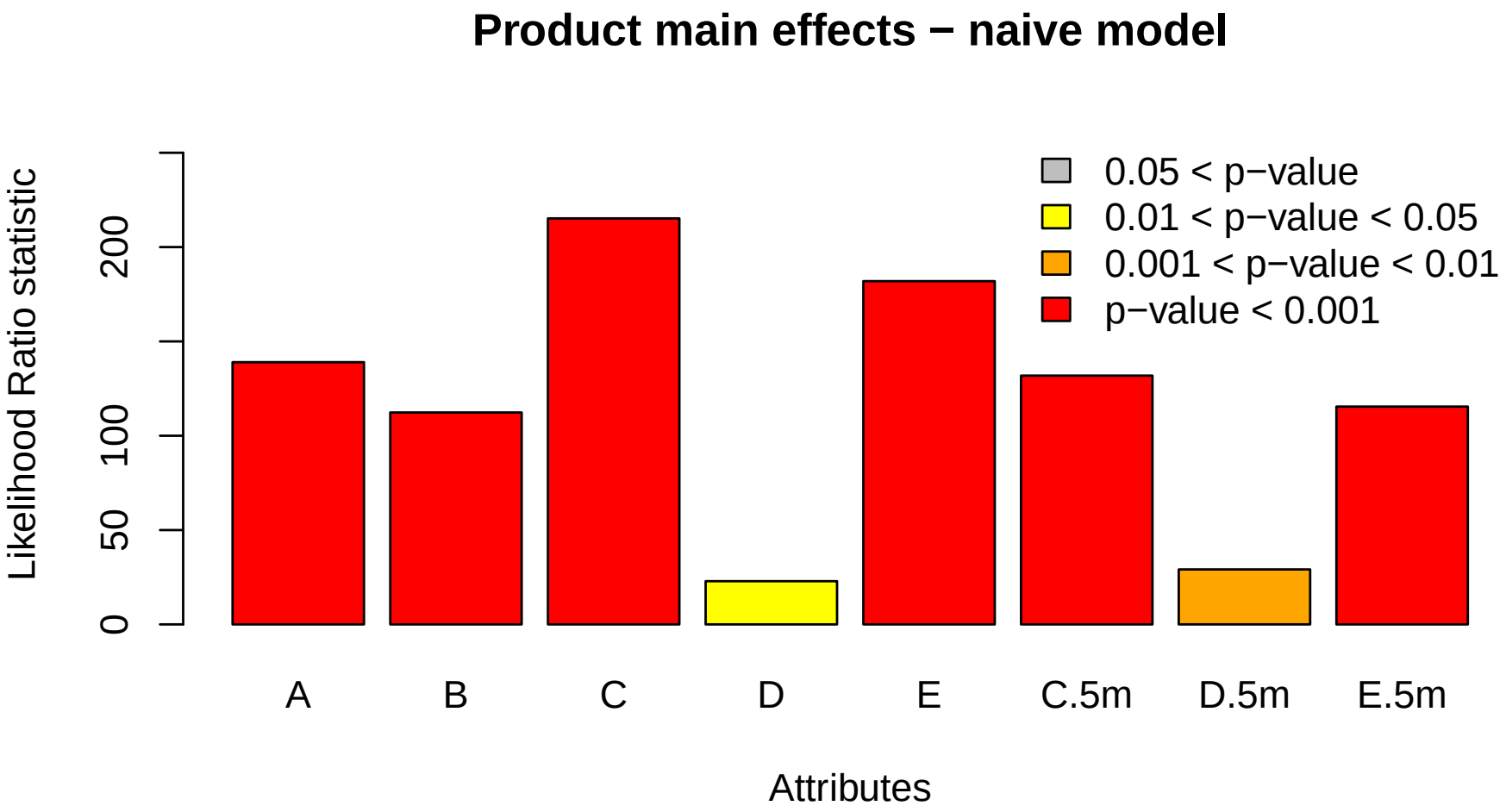
Likelihood Ratio test - mixed model - Binary response



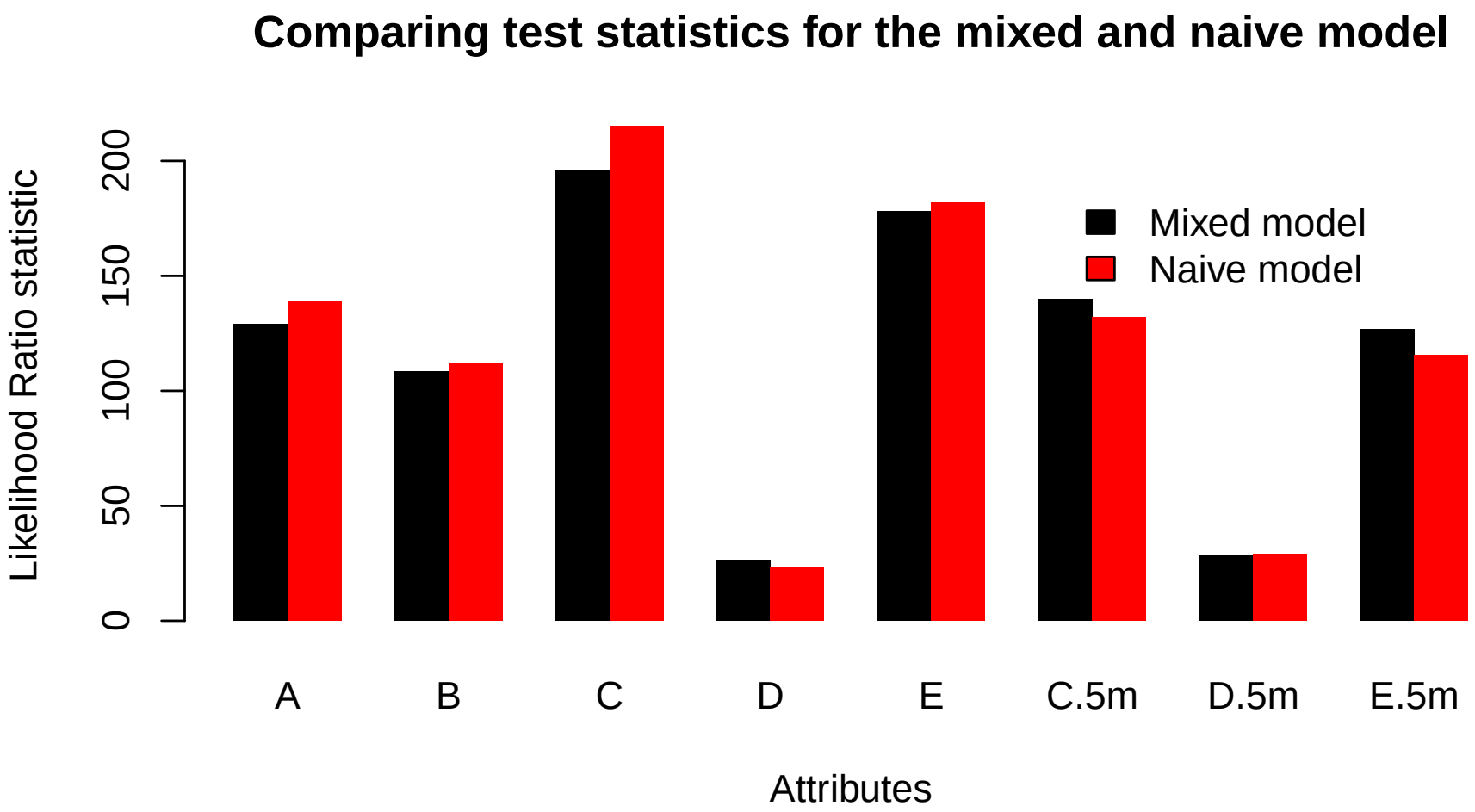
Likelihood Ratio test - mixed model - Binary response



Likelihood Ratio test - naive model - Binary response



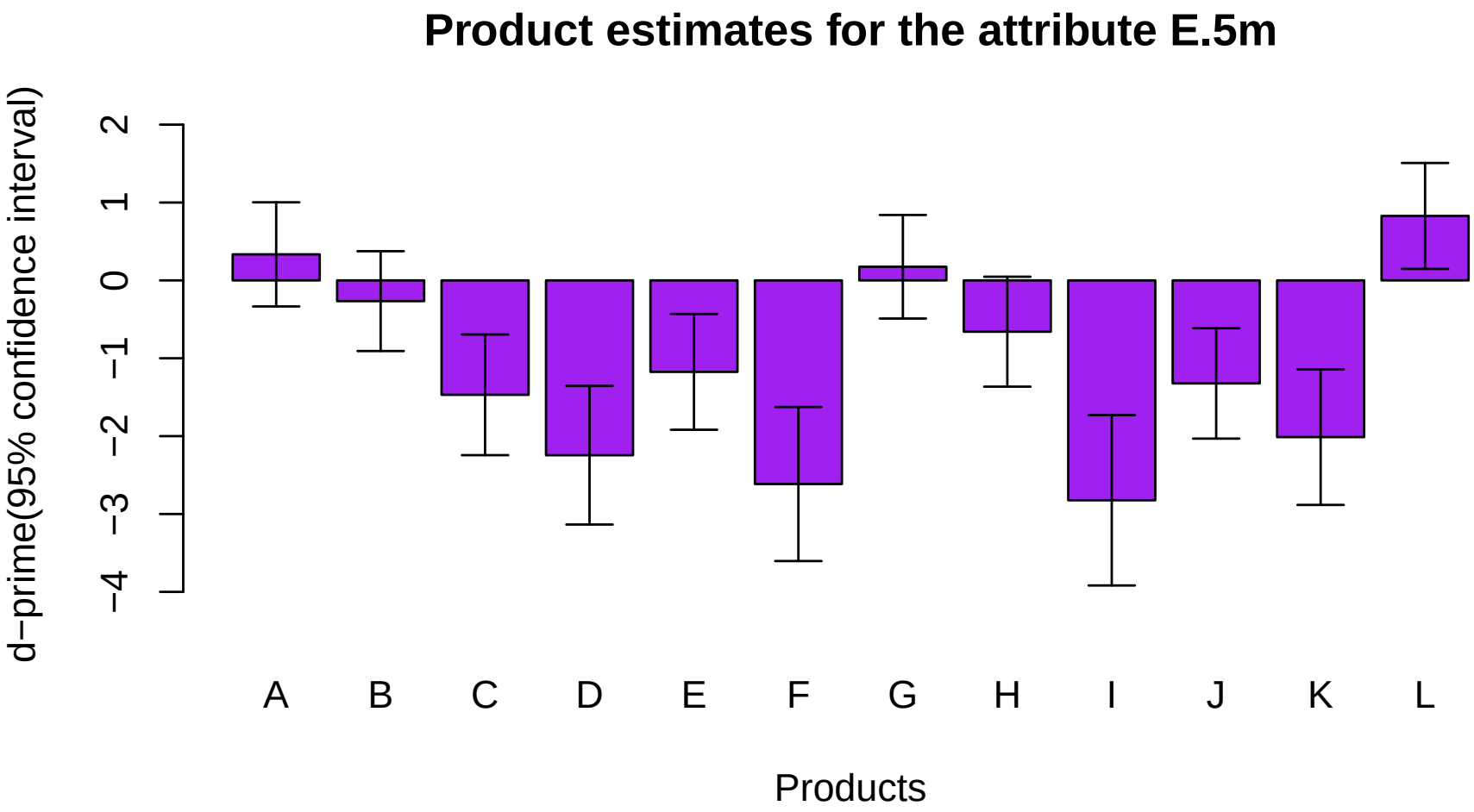
Likelihood Ratio test - comparison of Product - Binary response



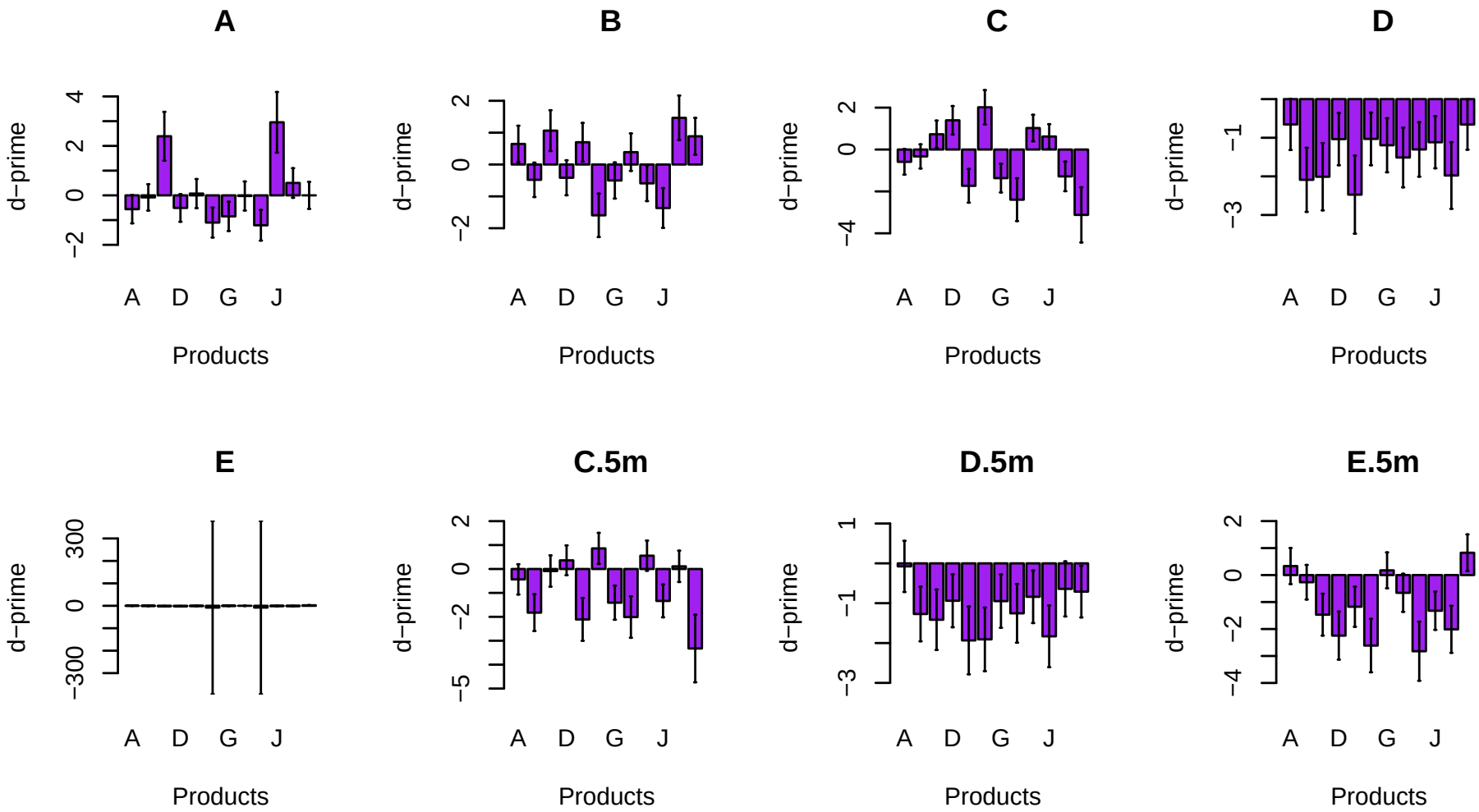
Comparing the two models for Product - Binary response

- Not a big difference in values between the methods. This fits nicely with the Assessor-by-Product interaction being non-significant
- In general you would expect the values from the mixed model to be less than the values from the naive model
- For some attributes this is not the case. These are the ones with the highest values of the test of Assessor main effect
- Looking at the plot of the Assessor main effect you would expect that this was also the case for for the D.5m attribute.
  But looking at the plot of the Assessor-by-Product interaction this is the attribute with the highest value

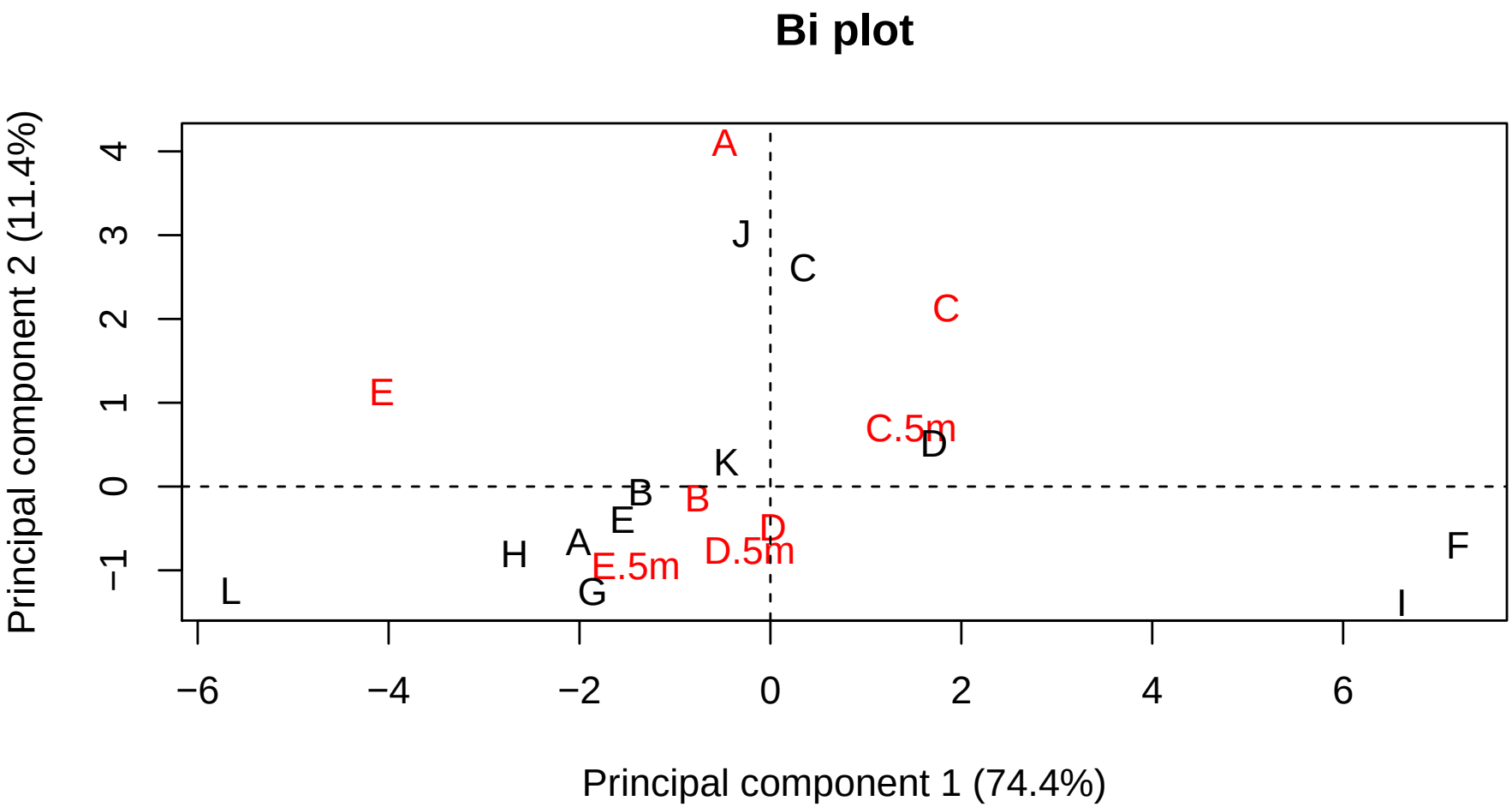
Post hoc - Product differences in terms of d-prime



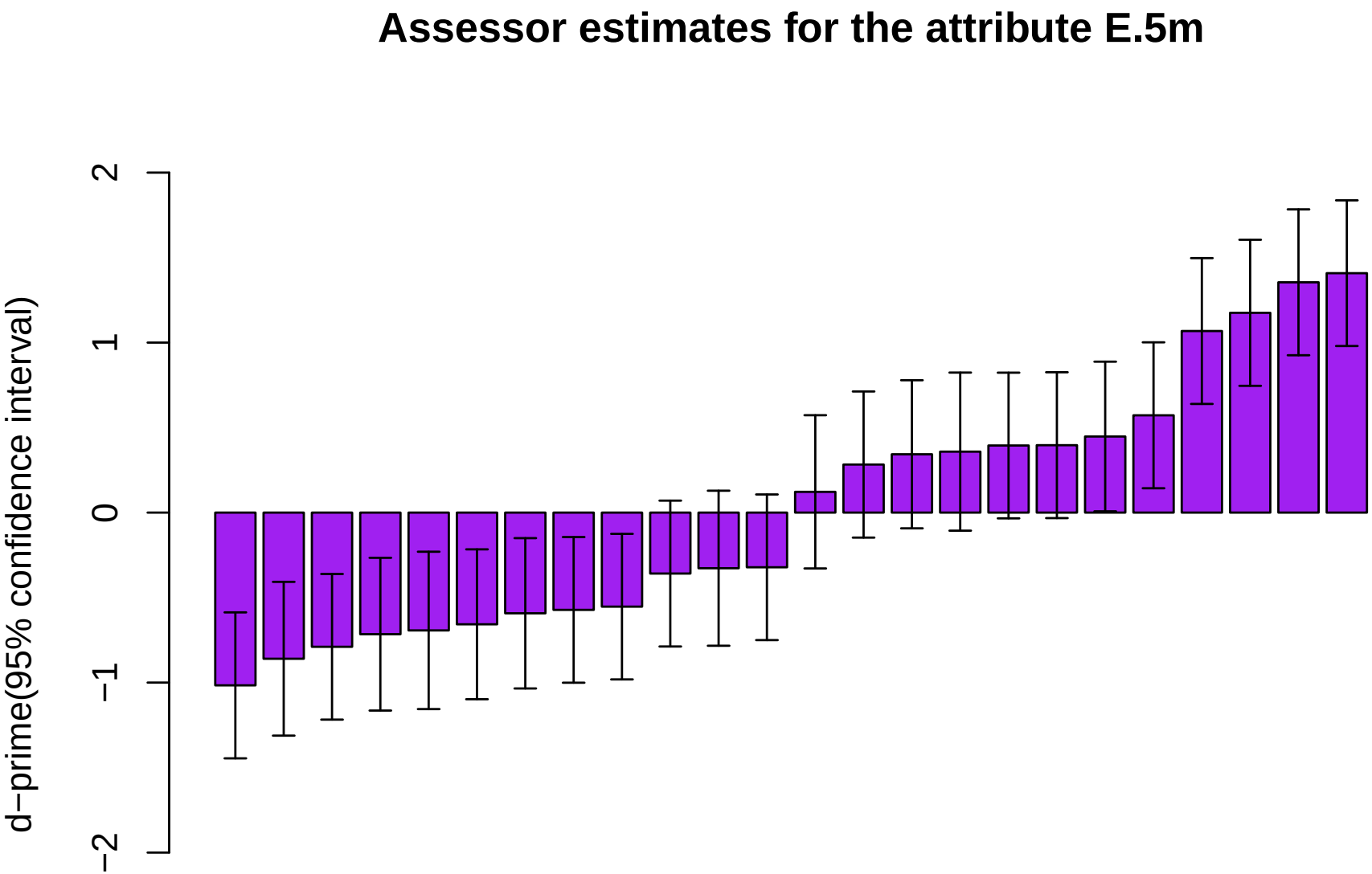
Post hoc - Product differences in terms of d-prime



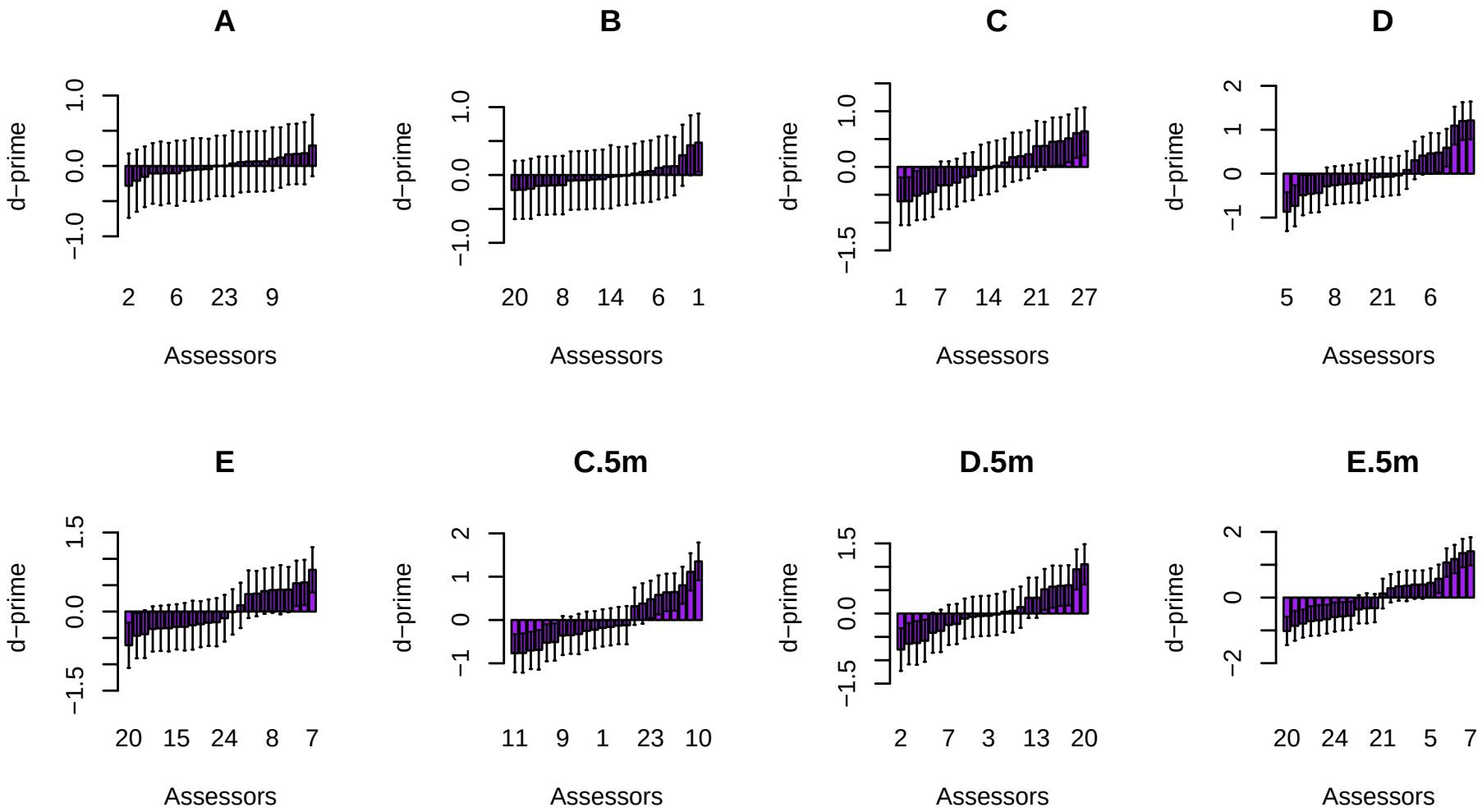
Post hoc - PCA of product d-prime values



Post hoc - Assessor performance in terms of d-prime

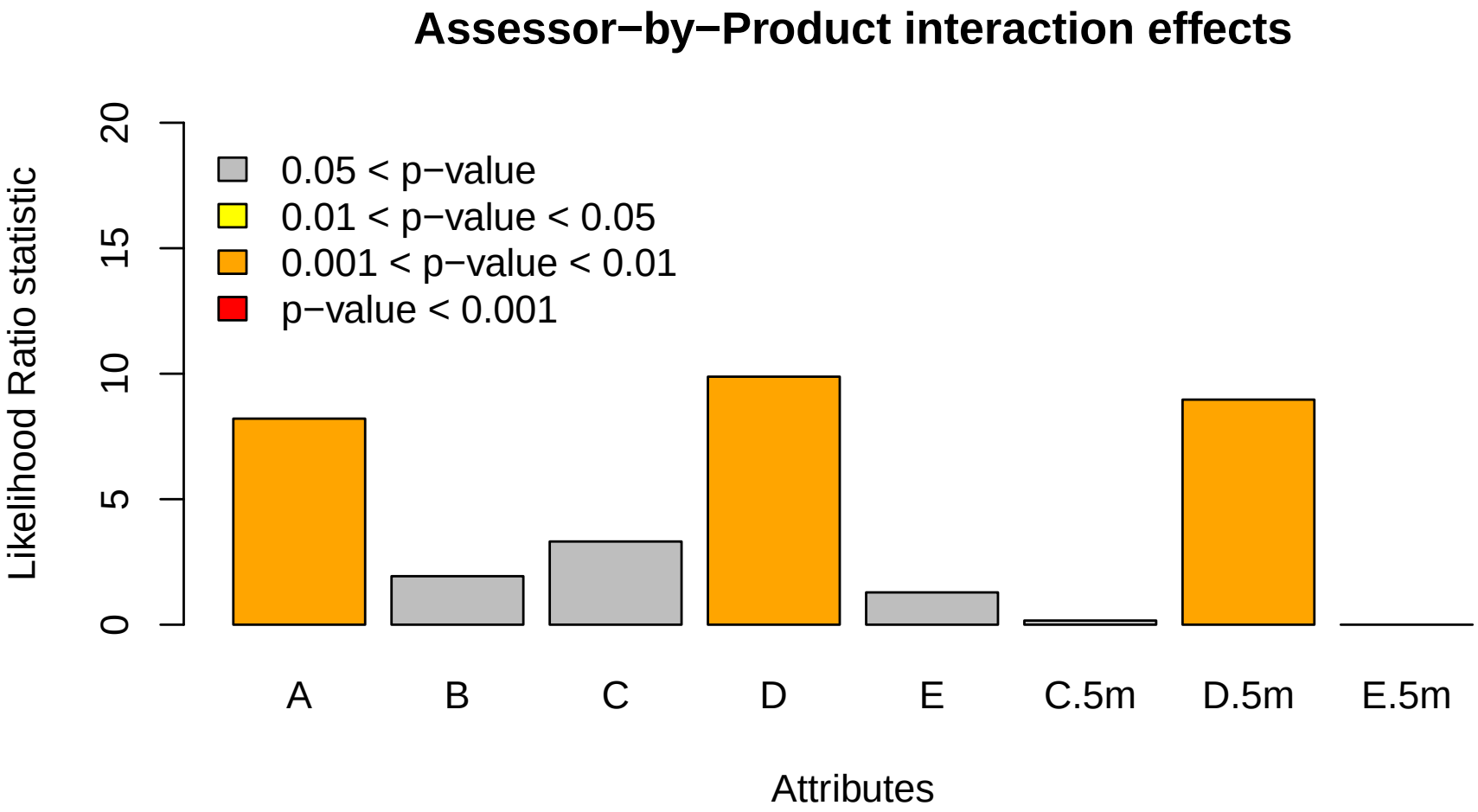


Post hoc - Assessor performance in terms of d-prime

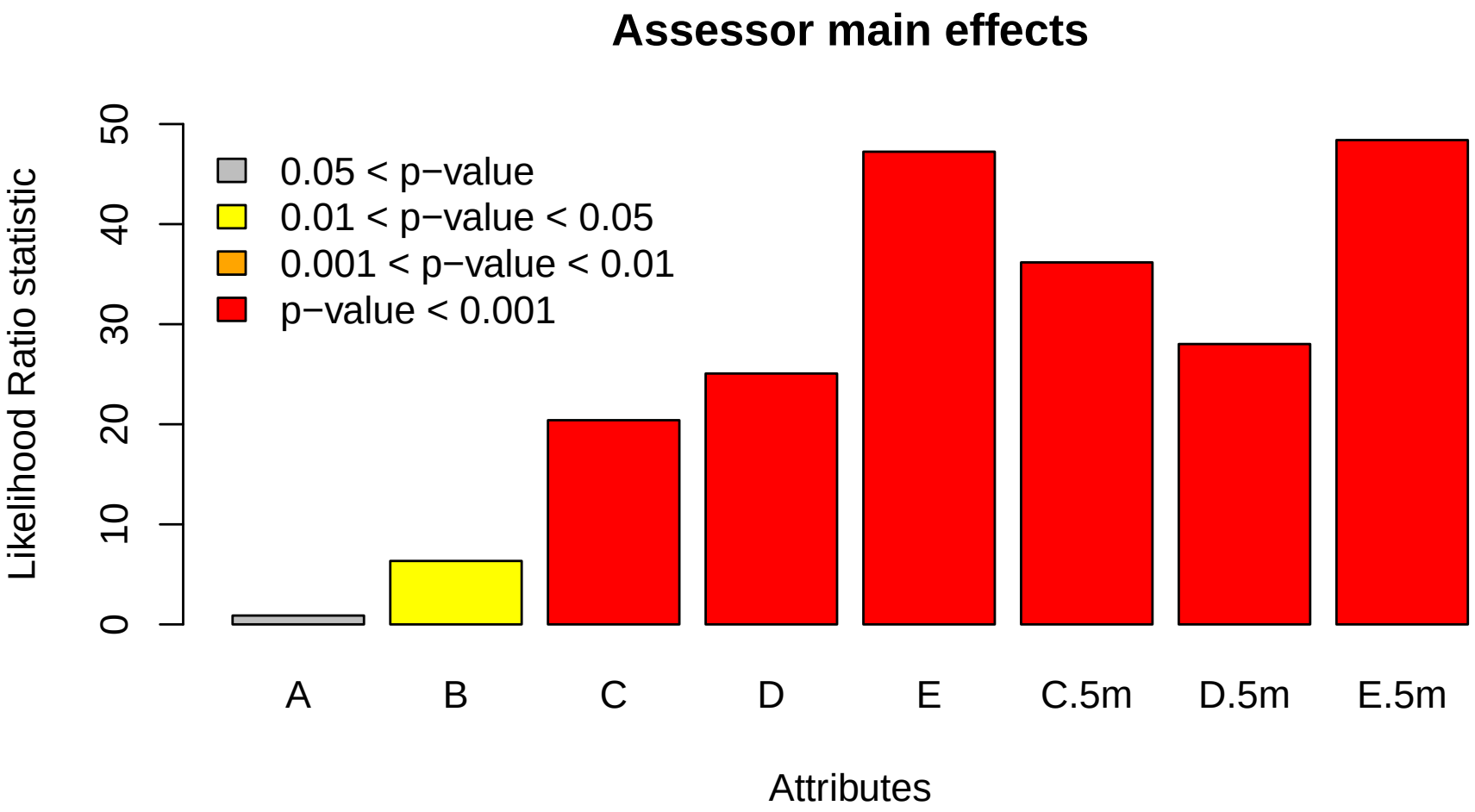


Likelihood Ratio test - mixed model - Ordinal response

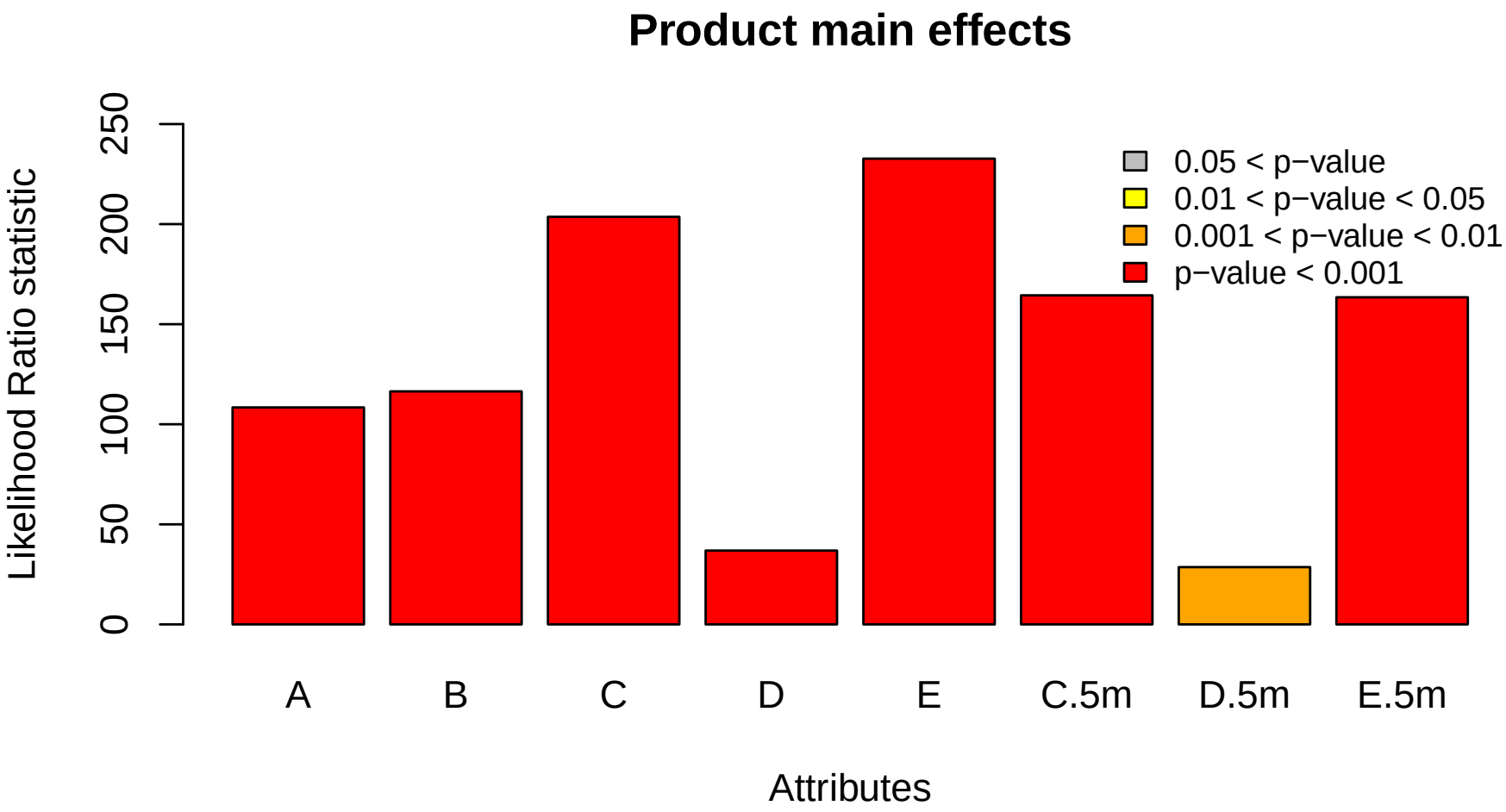
Using the R-package **ordinal**



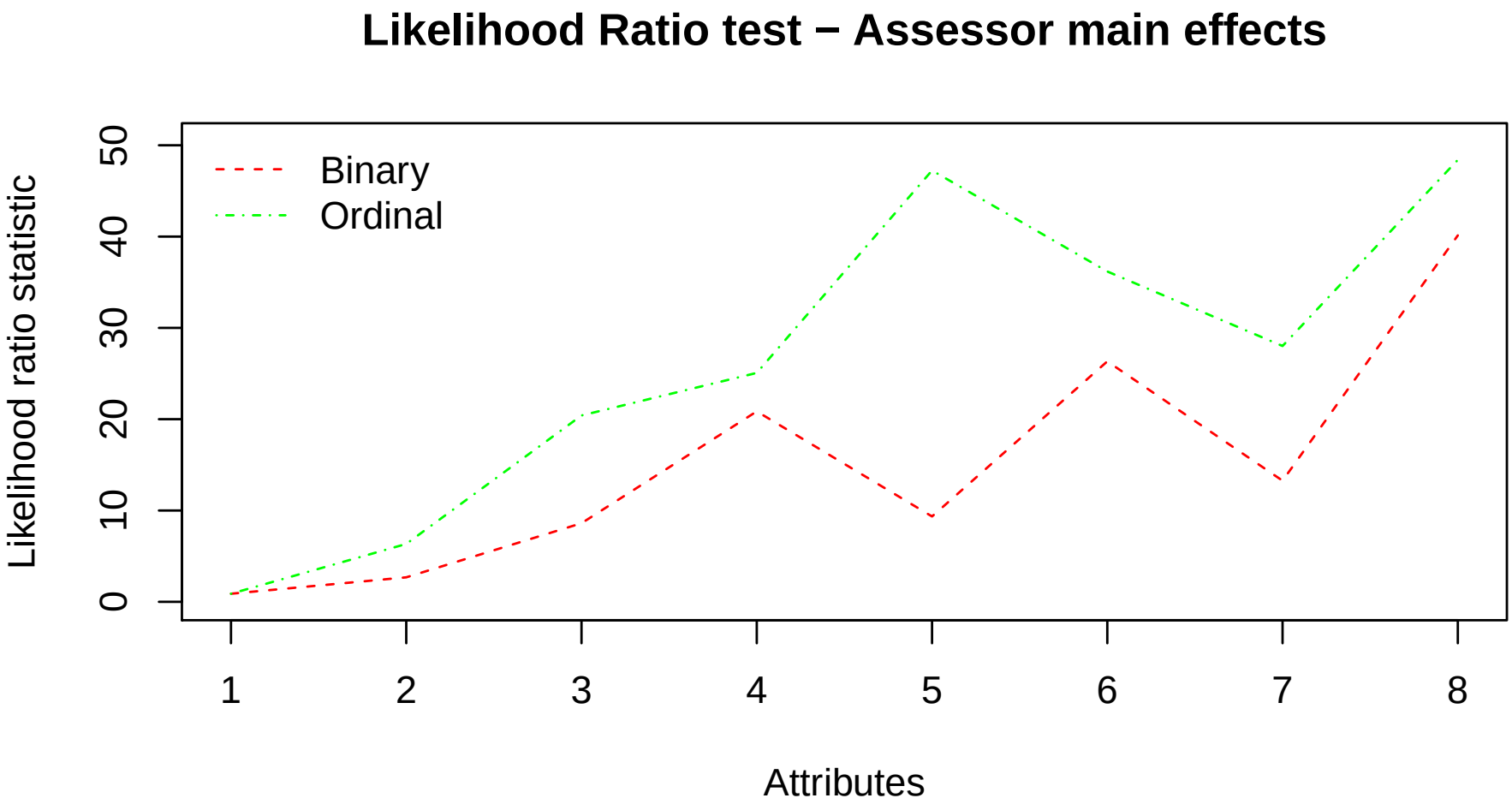
Likelihood Ratio test - mixed model - Ordinal response



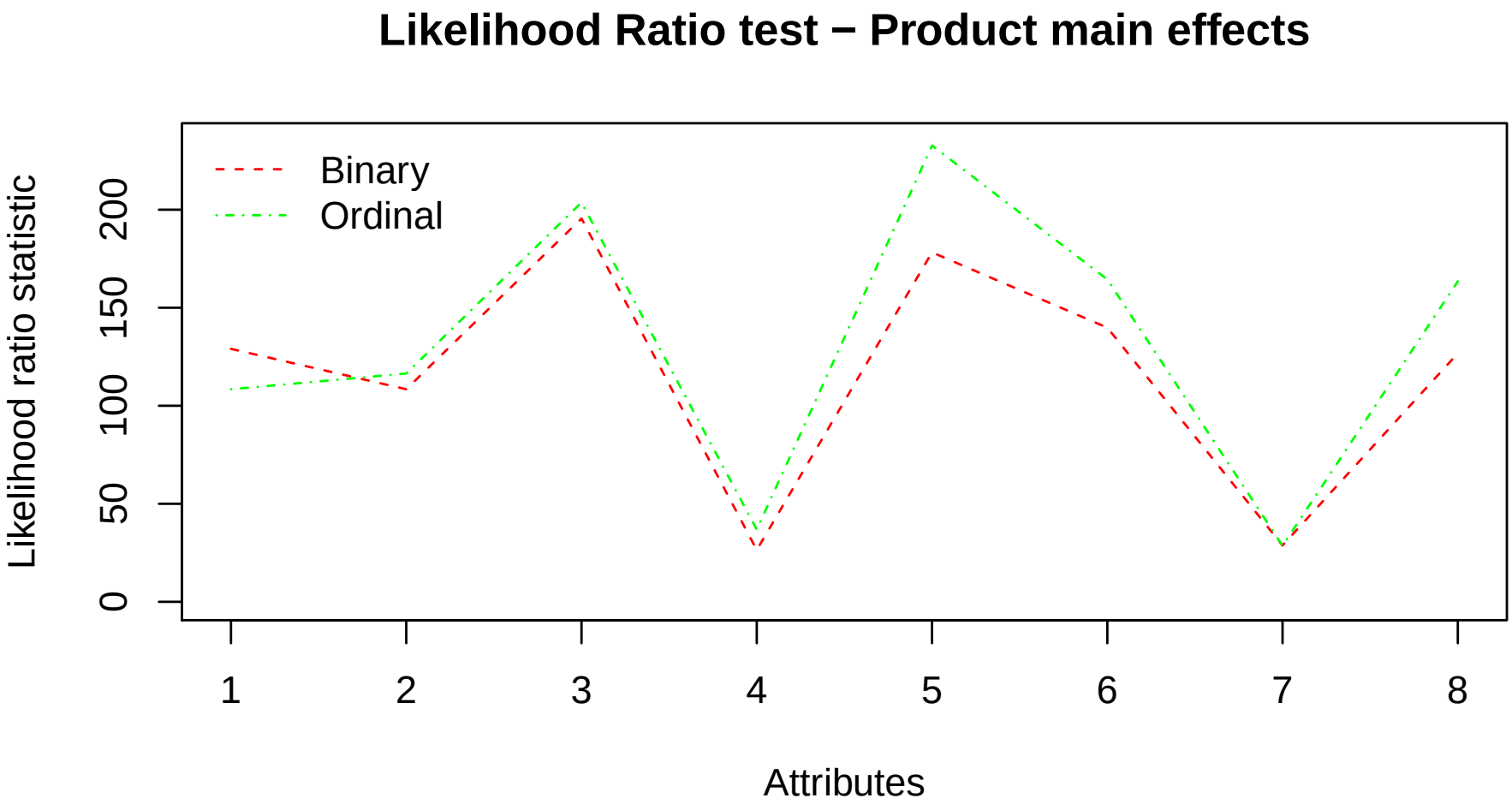
Likelihood Ratio test - mixed model - Ordinal response



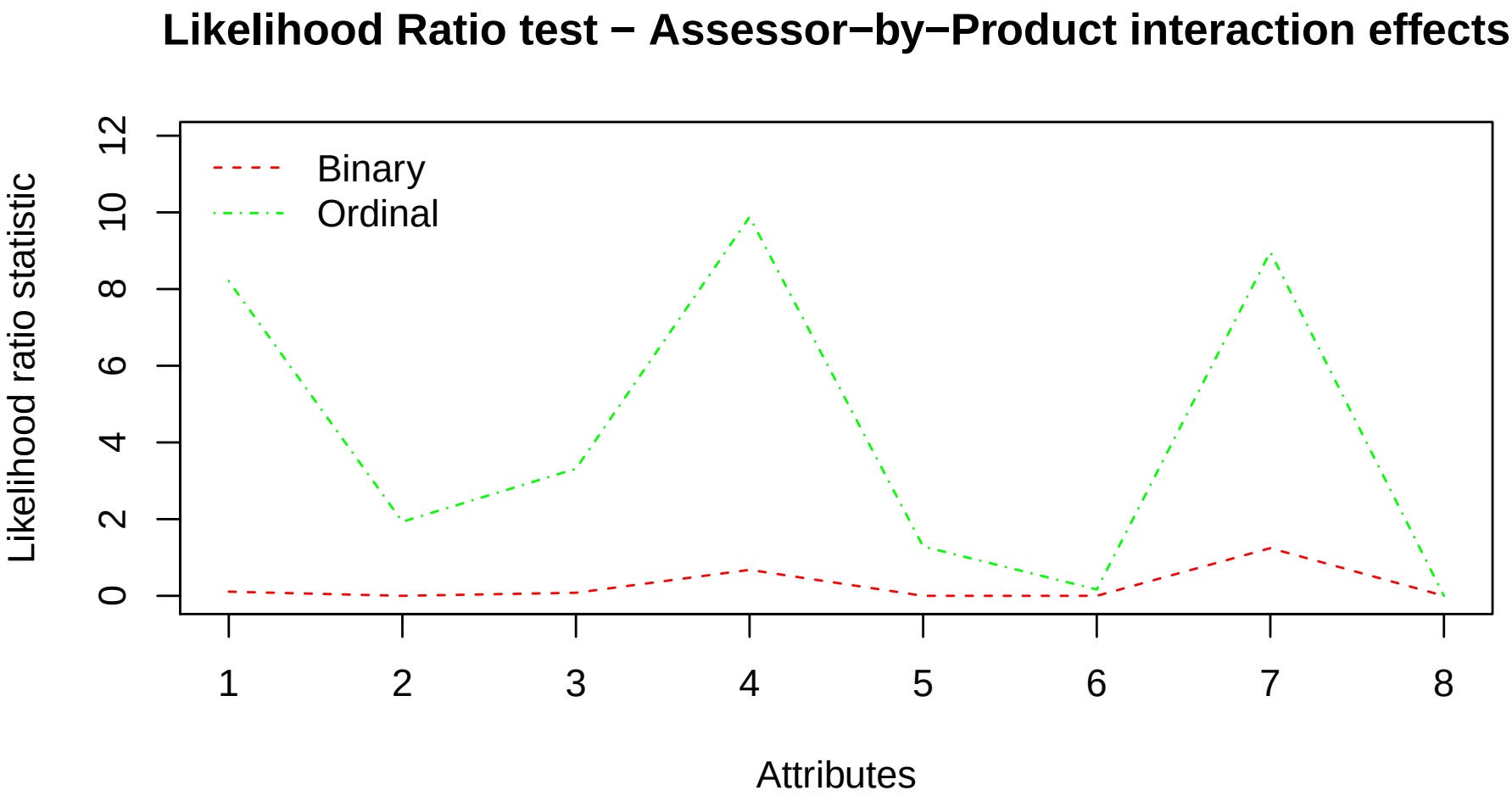
Likelihood Ratio test - comparison Assessor



Likelihood Ratio test - comparison Product



Likelihood Ratio test - comparison Assessor-by-Product



Final remarks

We can handle incomplete observations

We can fit the Thurstonian mixed model with binary data

The Thurstonian mixed model provides additional information (compared to the naive aggregation model)

- d-prime interpretations of Product estimates
- d-prime interpretations of Assessor estimates

These preliminary results indicate that the ordinal information is more sensitive than the binary information

Future work

At some point in the future:

- Investigation of how much binary data is needed to find a significant Assessor-by-Product interaction
- Come up with a proper Thurstonian interpretation of the ordinal approach
- General modelling - for instance how to handle order effects

Thank you for your attention!

References

- PanelCheck: Open source software for sensory profile data, www.panelcheck.com
- Christensen, R. H. B. (2012). Ordinal - Regression Models for Ordinal Data. R package version 2012.01-19 http://www.cran.r-project.org/package=ordinal/
- Christensen, R. H. B. and P. B. Brockhoff (2012). Analysis of replicated categorical ratings data from sensory experiments. Working paper.
- Alexandra Kuznetsova, Per Bruun Brockhoff and Rune Haubo Bojesen Christensen (2012). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package).. R package version 1.0.
- Brockhoff, P.B. and Christensen, R.H.B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. Food Quality and Preference 21(3), 330-338