The **ordinal** package: Analyzing ordinal data

Per Bruun Brockhoff

DTU Compute
Section for Statistics
Technical University of Denmark
perbb@dtu.dk

August 20 2015

DTU Compute
Department of Applied Mathematics and Computer Science

Outline

Ordinal scales are commonly used (attribute rating or liking)

5-point scales:

1	2	3	4	5
□	□	□	□	□

7-point scales:

1	2	3	4	5	6	7
□	□	□	□	□	□	□

9-point scales:

1	2	3	4	5	6	7	8	9
□	□	□	□	□	□	□	□	□

Ordinal data — the wine data

Randall, J (1989). The analysis of sensory data by generalised linear model. Biometrical journal 7, 781-793.

AND: included in **ordinal** pacakge.

A sensory experiment:
Temperature and contact between juice and skins can be controlled when crushing grapes during wine production. These factors are thought to affect the bitterness of the wine.

Objective:
How does perceived bitterness depend on temperature and contact?

Ordinal data — the wine data

Objective:
How does perceived bitternes depend on temperature and contact?

Table: The wine data (Randall, 1989), N=72

Variables	Type	Values
bitterness	response	1, 2, 3 ,4, 5
		less — more
temperature	predictor	cold, warm
contact	predictor	no, yes
judges	random	1, . . . , 9

Temperature and contact between juice and skins can be controlled when cruching grapes during wine production.

Data for the bitterness of white wines

Table: Ratings of the bitterness of some white wines. Data are adopted from Randall (1989).

			Judge								
Temperature	Contact	Bottle	1	2	3	4	5	6	7	8	9
cold	no	1	2	1	2	3	2	3	1	2	1
cold	no	2	3	2	3	2	3	2	1	2	2
cold	yes	3	3	1	3	3	4	3	2	2	3
cold	yes	4	4	3	2	2	3	2	2	3	2
warm	no	5	4	2	5	3	3	2	2	3	3
warm	no	6	4	3	5	2	3	4	3	3	2
warm	yes	7	5	5	4	5	3	5	2	3	4
warm	yes	8	5	4	4	3	3	4	3	4	4

Bitterness ratings: 1(least), 2, 3, 4, 5(most)

Ordinal data — Income group data

McCullagh, P. (1980) Regression Models for Ordinal Data. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 42, No. 2., pp. 109-142.

AND: included in **ordinal** pacakge.

```
head(income)

##   year  pct income
## 1 1960  6.5    0-3
## 2 1960  8.2    3-5
## 3 1960 11.3    5-7
## 4 1960 23.5   7-10
## 5 1960 15.6  10-12
## 6 1960 12.7  12-15
```


Ordinal data — Soup data

Christensen, R. H. B., Cleaver, G., & Brockhoff, P. B. (2011). Statistical and Thurstonian models for the A-not A protocol with and without sureness. Food Quality and Preference, 22, 542-549.
Industrial product development experiment - Unilever.

AND: included in **ordinal** pacakge.
A-not A with sureness scale:

'Reference'			'Not Reference'		
Sure	Not Sure	Guess	Guess	Not Sure	Sure

```
head(soup)

##   RESP PROD PRODID SURENESS DAY SOUPTYPE SOUPFREQ COLD EASY GENDER
## 1    1  Ref      1        6   1   Canned >1/week  Yes    7 Female
## 2    1 Test      2        5   1   Canned >1/week  Yes    7 Female
## 3    1  Ref      1        5   1   Canned >1/week  Yes    7 Female
## 4    1 Test      3        6   1   Canned >1/week  Yes    7 Female
## 5    1  Ref      1        5   2   Canned >1/week  Yes    7 Female
## 6    1 Test      6        5   2   Canned >1/week  Yes    7 Female
##   AGEGROUP LOCATION
## 1    51-65 Region 1
## 2    51-65 Region 1
## 4    51-65 Region 1
## 5    51-65 Region 1
## 6    51-65 Region 1
```

Ordinal data — (Paired) Degree-of-difference data

25 panelists — 8 replications.

Table: Paired degree-of-difference test, data adopted from (?)

Pair	Identical	Uncertain	Different	Total
Identical pair	45	40	15	100
Different pair	36	34	30	100

Christensen, R. H. B. and P. B. Brockhoff (2011) Analysis of replicated categorical ratings data from sensory experiments. *Journal of the French Statistical Society, SFdS*, 154(3), 58-79.

Ordinal data — The 2-Alternative choice test (2-AC)

2-Alternative Choice (2-AC):

- Do you prefer A or B or do you not have a preference?
- Which is strongest, A or B, or is there no difference?

Table: 208 consumers with 4 replications

Condition	"Prefer A"	"No-preference"	"Prefer B"	Total
A	260	37	119	416
B	217	38	161	416

Christensen, R. H. B., H.-S. Lee and P. B. Brockhoff (2012) Estimation of the Thurstonian model for the 2-AC protocol. *Food Quality and Preference, 24*, 119-128.

Appropriate models for ordinal data

Ordinal data — **not continuous** data

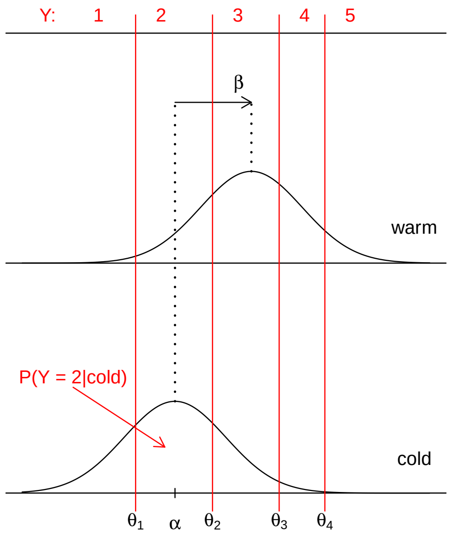
A linear regression model on the scores (1,. . . ,5)?
Breach of assumptions:

- The scores are **not** normally distributed
- A score of "4" is not twice as much as "2"
- Variance not likely to be constant

Our approach:
A cumulative link model (CLM)

- Only use information about ordering
- Intuitively: A linear model that respects the ordinal nature of the response

Understanding the cumulative link model



- *Latent* bitterness follows a linear model:

$$S_i = \alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$
$$= \alpha + \beta(\texttt{temp}_i) + \varepsilon_i$$

- We only observe a grouped version of S_i:
- $\theta_{j-1} \leq S_i < \theta_j \rightarrow Y = j$

$$P(Y_i \leq j) = F(\theta_j - \boldsymbol{x}_i^\top \boldsymbol{\beta})$$

Fitting cumulative link models with clm

```
data(wine)
fm1 <- clm(rating ~ contact + temp, data=wine, link="probit")
summary(fm1)

## formula: rating ~ contact + temp
## data:    wine
##
##  link   threshold nobs logLik AIC    niter max.grad cond.H
##  probit flexible  72   -85.76 183.52 5(0)  1.43e-13 2.2e+01
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## contactyes    0.868      0.267    3.25   0.0011 **
## tempwarm      1.499      0.292    5.14  2.8e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##     Estimate Std. Error z value
## 1|2   -0.773      0.283   -2.73
## 2|3    0.736      0.250    2.94
## 3|4    2.045      0.322    6.35
## 4|5    2.941      0.387    7.60
```


A cumulative link model for the wine data

Additive effects for temperature and contact:

$$P(Y_i \leq j) = \Phi\left(\theta_j - \beta_1(\texttt{temp}_i) - \beta_2(\texttt{contact}_i)\right)$$

- Is there an interaction between temp and contact?

Table: ANODE table for the wine data.

Source	df	deviance	p value
Total	12	39.407	< 0.001
Treatment	3	34.606	< 0.001
Temperature, T	1	26.928	< 0.001
Contact, C	1	11.043	< 0.001
Interaction, $T \times C$	1	0.1514	0.6972
Residual	9	4.8012	0.8513

An extended CLM Framework

Standard CLM:

$$F(\theta_j - \boldsymbol{x}_i^\top \boldsymbol{\beta})$$

Extended CLM:

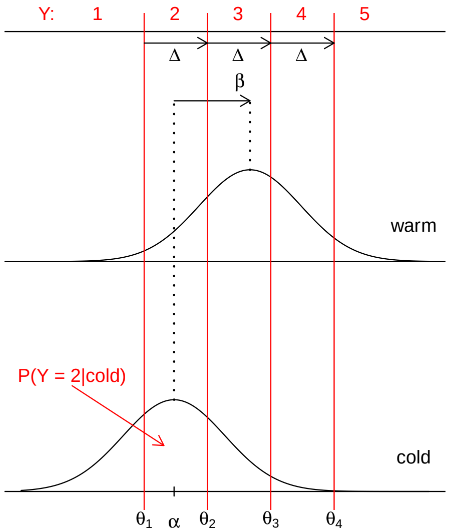
$$F\left(\frac{g(\theta_j) - \boldsymbol{w}_i^\top \boldsymbol{\beta}_j^* - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{\exp(\boldsymbol{z}_i^\top \boldsymbol{\zeta})}\right)$$

Threshold effects
Nominal effects
Scale effects

CLMM (Mixed effects):

$$F(\theta_j - \overset{\text{fixed}}{\boldsymbol{X\beta}} - \overset{\text{random}}{\boldsymbol{Zb}})$$

Understanding structured thresholds — restrictions



- The cumulative link model:

  $$P(Y_i \leq j) = F(\theta_j - \beta(\texttt{temp}_i))$$

- θ_j — ordered, but otherwise not restricted
- Require symmetry?
- Require equidistance?

Fitting models with structured thresholds

```
fm.equi <- clm(rating ~ contact + temp, data=wine,
               link="probit", threshold="equidistant")
summary(fm.equi)

## formula: rating ~ contact + temp
## data:    wine
##
##  link   threshold   nobs logLik AIC    niter max.grad cond.H
##  probit equidistant 72   -87.24 182.47 4(0)  1.40e-08 3.2e+01
##
## Coefficients:
##            Estimate Std. Error z value Pr(>|z|)
## contactyes    0.857      0.264    3.24   0.0012 **
## tempwarm      1.489      0.288    5.17  2.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##             Estimate Std. Error z value
## threshold.1   -0.587      0.233   -2.52
## spacing        1.241      0.128    9.67
```


Sensory applications of structured thresholds

Symmetric thresholds for sureness scales:

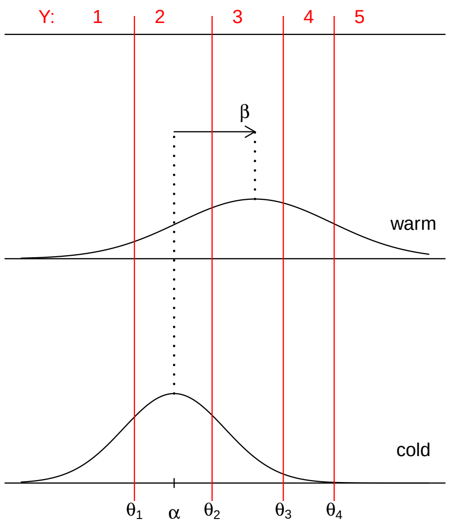
'Reference'			'Not Reference'		
Sure	Not Sure	Guess	Guess	Not Sure	Sure

Equidistant thresholds for 7- or 9-point preference scales:

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Equally spaced categories is a necessary condition for using linear models for continuous data on ordinal data.

Understanding scale effects in CLMs



Model for latent bitterness:

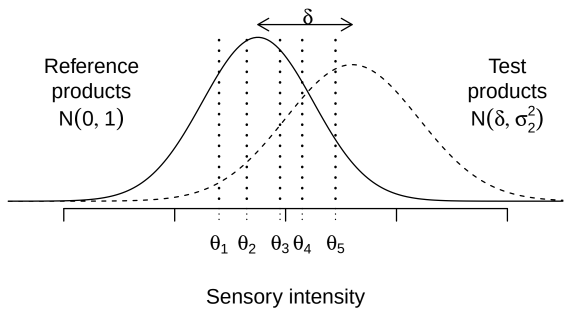
$$S_i = \alpha + \beta_1(\texttt{temp}_i) + \beta_2(\texttt{contact}_i) + \varepsilon_i,$$

$$\varepsilon_i \sim N(0, \sigma^2(\texttt{temp}_i))$$

Mccullagh, 1980, Cox, 1995, Agresti 2002

$$\gamma_{ij} = F\left(\frac{\theta_j - \beta_1(\texttt{temp}_i) - \beta_2(\texttt{contact}_i)}{\zeta_1(\texttt{temp}_i)}\right)$$

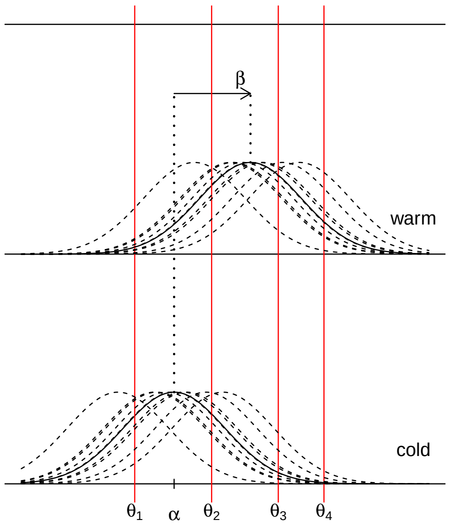
Thurstonian model for the A-not A with sureness protocol



	'Reference'			'Not Reference'		
Product	Sure	Not Sure	Guess	Guess	Not Sure	Sure
Reference	132	161	65	41	121	219
Test	96	99	50	57	156	650

Table: Discrimination of packet soup Christensen, Cleaver and Brockhoff, 2011

Including random effects in CLMs



The cumulative link model:

$$\gamma_{ij} = F(\theta_j - \beta_1(\texttt{temp}_i) - \beta_2(\texttt{contact}_i))$$

- Judges perceive wine bitterness differently
- Judges use the response scale differently

Add random effects for judges:

$$\gamma_{ij} = F(\theta_j - \beta_1(\texttt{temp}_i) - \beta_2(\texttt{contact}_i)$$
$$-b(\texttt{judge}_i)), \quad b \sim N(0, \sigma_b^2)$$

Cumulative link mixed models

$$\gamma_k = F(\boldsymbol{B}_k\boldsymbol{\psi} - \boldsymbol{Z}\boldsymbol{v} - \boldsymbol{o}_k) \qquad \boldsymbol{V} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma_\tau})$$

The log-likelihood function:

$$\ell(\boldsymbol{\psi}, \boldsymbol{\tau}; \boldsymbol{y}) = \log \int_{\mathbb{R}^r} p_{\boldsymbol{\psi}}(\boldsymbol{y}|\boldsymbol{v}) p_{\boldsymbol{\tau}}(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}$$

Integration methods:

- Laplace approximation Tierney and Kadane, 1986, Pinheiro and Bates, 1995, Joe 2008
- Gauss-Hermite quadrature (GHQ) Hedeker and Gibbons, 1994
- Adaptive Gauss-Hermite quadrature (AGQ) Liu and Pierce, 1994

A Newton-Raphson algorithm updates the conditional modes of the random effects (Laplace and AGQ)

Allowing for differences between judges

Research questions:

- Are judges rating the wines differently?
- Are there differences between bottles?

Additive random effects for judges:

$$P(Y_i \leq j) = \Phi\left(\theta_j - \beta_1(\texttt{temp}_i) - \beta_2(\texttt{contact}_i) - u(\texttt{judge}_i)\right)$$

$$u(\texttt{judge}_i) \sim N(0, \sigma_u^2)$$

Additive random effects for judges and bottles:

$$P(Y_i \leq j) = \Phi\left(\theta_j - \beta_1(\texttt{temp}_i) - \beta_2(\texttt{contact}_i) - u(\texttt{judge}_i) - b(\texttt{bottle}_i)\right)$$

$$u(\texttt{judge}_i) \sim N(0, \sigma_u^2) \quad b(\texttt{bottle}_i) \sim N(0, \sigma_b^2)$$

ANODE for mixed effects CLM

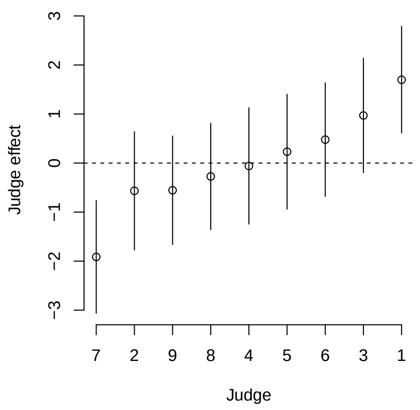
Table: ANODE table for the wine data with random effects.

Source	df	deviance	p value
Total	14	45.577	< 0.001
Var(Judge)	1	9.661	< 0.001
Var(Bottle)	1	0.001	0.998
Treatment	3	34.606	< 0.001
Temperature, T	1	25.384	< 0.001
Contact, C	1	14.238	< 0.001
Interaction, $T \times C$	1	0.1086	0.7417

Results:

- Bottles are probably not that different
- Judges do rate the wines differently

Panel inference — judge effects



CLMMs for sensory ratings data

25 panelists — 8 replications.

Table: Paired degree-of-difference test, data adopted from (?)

Pair	Identical	Uncertain	Different	Total
Identical pair	45	40	15	100
Different pair	36	34	30	100

Christensen, R. H. B. and P. B. Brockhoff (2011) Analysis of replicated categorical ratings data from sensory experiments. *Journal of the French Statistical Society, SFdS*, 154(3), 58-79.

CLMMs for sensory ratings data

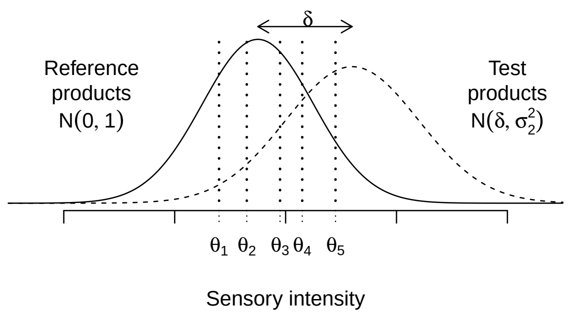
25 panelists — 8 replications.

Table: Comparison of tests of product differences.

Test	χ^2-value	df	p-value
Naive Pearson test	6.49	2	0.039
Stuart-Maxwell test (?)	3.85	2	0.149
LR test in CLMM	5.84	1	0.016

Christensen, R. H. B. and P. B. Brockhoff (2011) Analysis of replicated categorical ratings data from sensory experiments. *Journal of the French Statistical Society, SFdS*, 154(3), 58-79.

Thurstonian model for the A-not A with sureness protocol



	'Reference'			'Not Reference'		
Product	Sure	Not Sure	Guess	Guess	Not Sure	Sure
Reference	132	161	65	41	121	219
Test	96	99	50	57	156	650

Table: Discrimination of packet soup Christensen, Cleaver nd Brockhoff, 2011

Including assessor effects

Assumptions:

- Assessors do not use the response scale differently
- Assessors do not have different d's

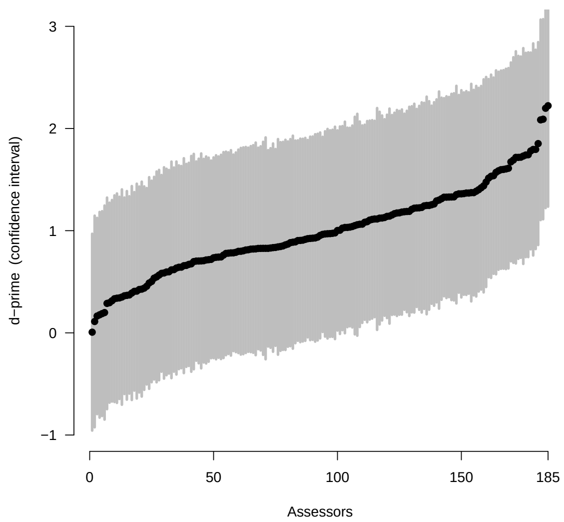
Accommodate this with mixed model extensions:

- Allow normally distributed random effects for assessors

$$P(S_i \leq \theta_j) = \Phi\left(\theta_j - \delta(\texttt{prod}_i) - u(\texttt{assessor}_i)\right) \qquad u \sim N(0, \sigma_u^2)$$

Note: This is similar to assessor effects in models for sensory profiling!

Inference for respondents — respondent-specific d's



The 2-Alternative choice test (2-AC)

2-Alternative Choice (2-AC):

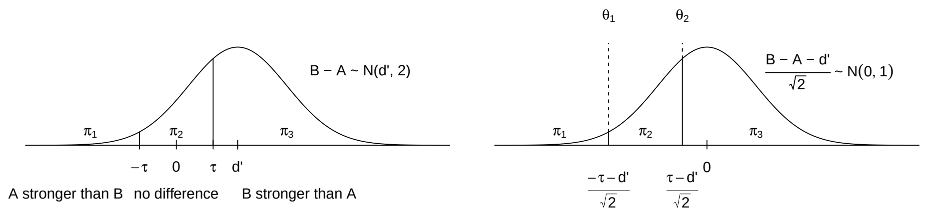
- Do you prefer A or B or do you not have a preference?
- Which is strongest, A or B, or is there no difference?

Table: 208 consumers with 4 replications

Condition	"Prefer A"	"No-preference"	"Prefer B"	Total
A	260	37	119	416
B	217	38	161	416

Christensen, R. H. B., H.-S. Lee and P. B. Brockhoff (2012) Estimation of the Thurstonian model for the 2-AC protocol. *Food Quality and Preference*, *24*, 119-128.

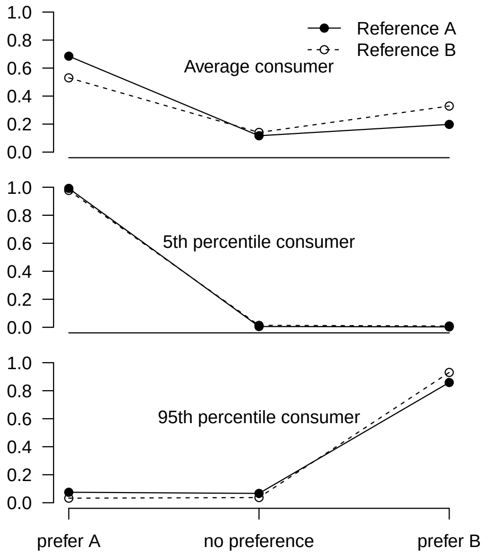
Thurstonian model for the 2-AC protocol



The Thurstonian model for the 2-AC protocol can be formulated as a cumulative link model:

$$\hat{\tau} = (\hat{\theta}_2 - \hat{\theta}_1)/\sqrt{2}$$
$$\hat{\delta} = (-\hat{\theta}_2 - \hat{\theta}_1)/\sqrt{2}$$
$$\text{se}(\hat{\tau}) = \sqrt{\{\text{var}(\theta_2) + \text{var}(\theta_1) - 2\text{cov}(\theta_2, \theta_1)\}/2}$$
$$\text{se}(\hat{\delta}) = \sqrt{\{\text{var}(\theta_2) + \text{var}(\theta_1) + 2\text{cov}(\theta_2, \theta_1)\}/2}$$

Illustrating the model



- 95% of population within $\pm 1.96\sigma_{d'} = \pm 3.3$ (d' units)
- The largest effect is consumer differences: $\chi^2_1 = 153.6$, $p < 0.001$.
- Effect of reference in duo-trio test only for consumers with an average preference

Outline

Excercises - Day 2 afternoon

- NO ordinal exercises
- IF you want: look at **ordinal** vignettes and/or manuel/help material
- Recommend instead - work with exercises from the previous three teaching blocks:
  - **sensR** part 1 exercises
  - **sensR** part 2 exercises
  - **lmerTest** exercises
  - **SensMixed** exercises