Threshold models with fixed and random effects for ordered categorical data

Hans-Peter Piepho

Universität Hohenheim, Germany

Edith Kalka

Universität Kassel, Germany

Contents

1. Introduction

Analysis of ordered categorical data:

- ANOVA, LSD
- Transformation to normality
- Nonparametric methods
- Threshold model

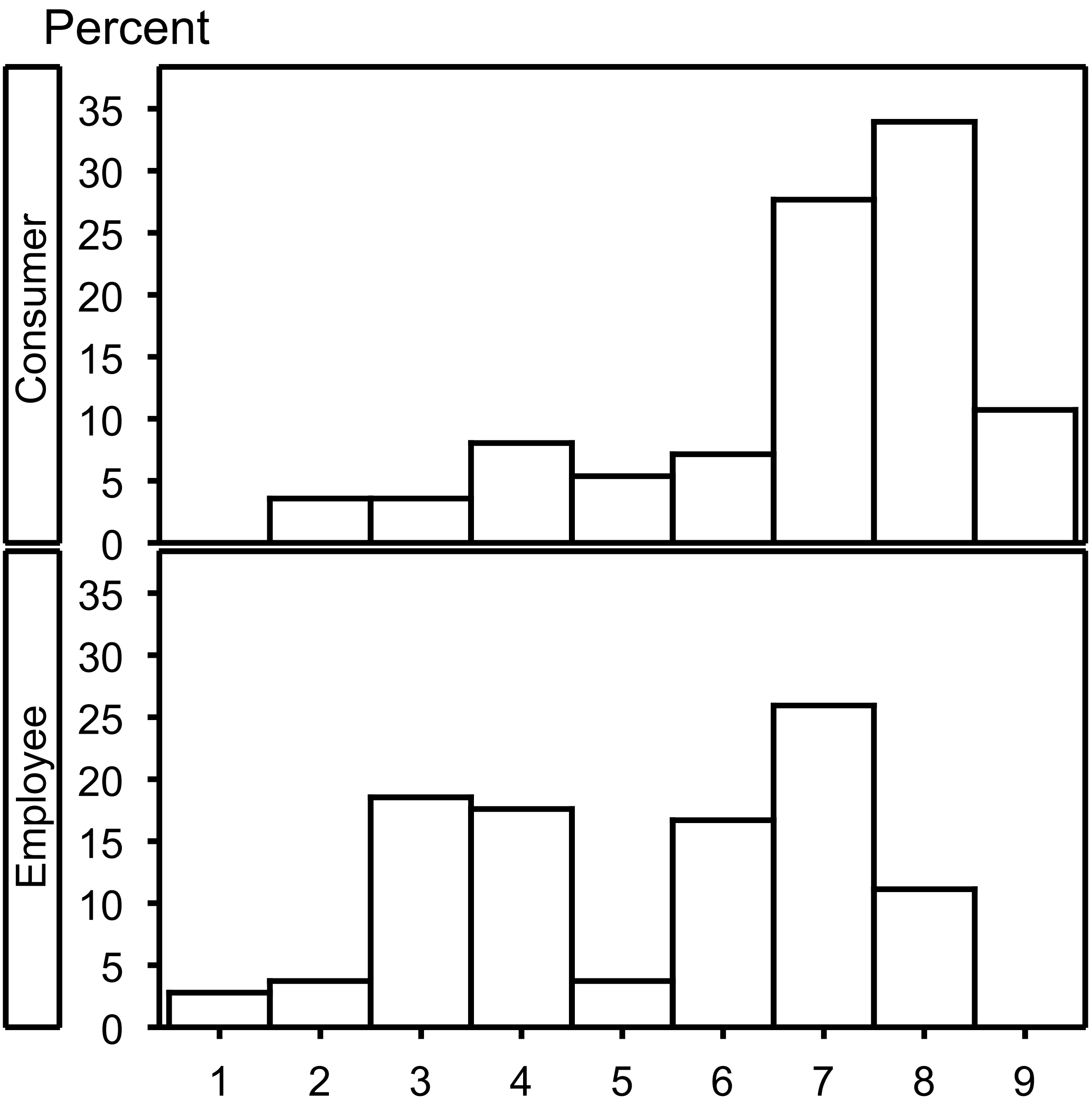
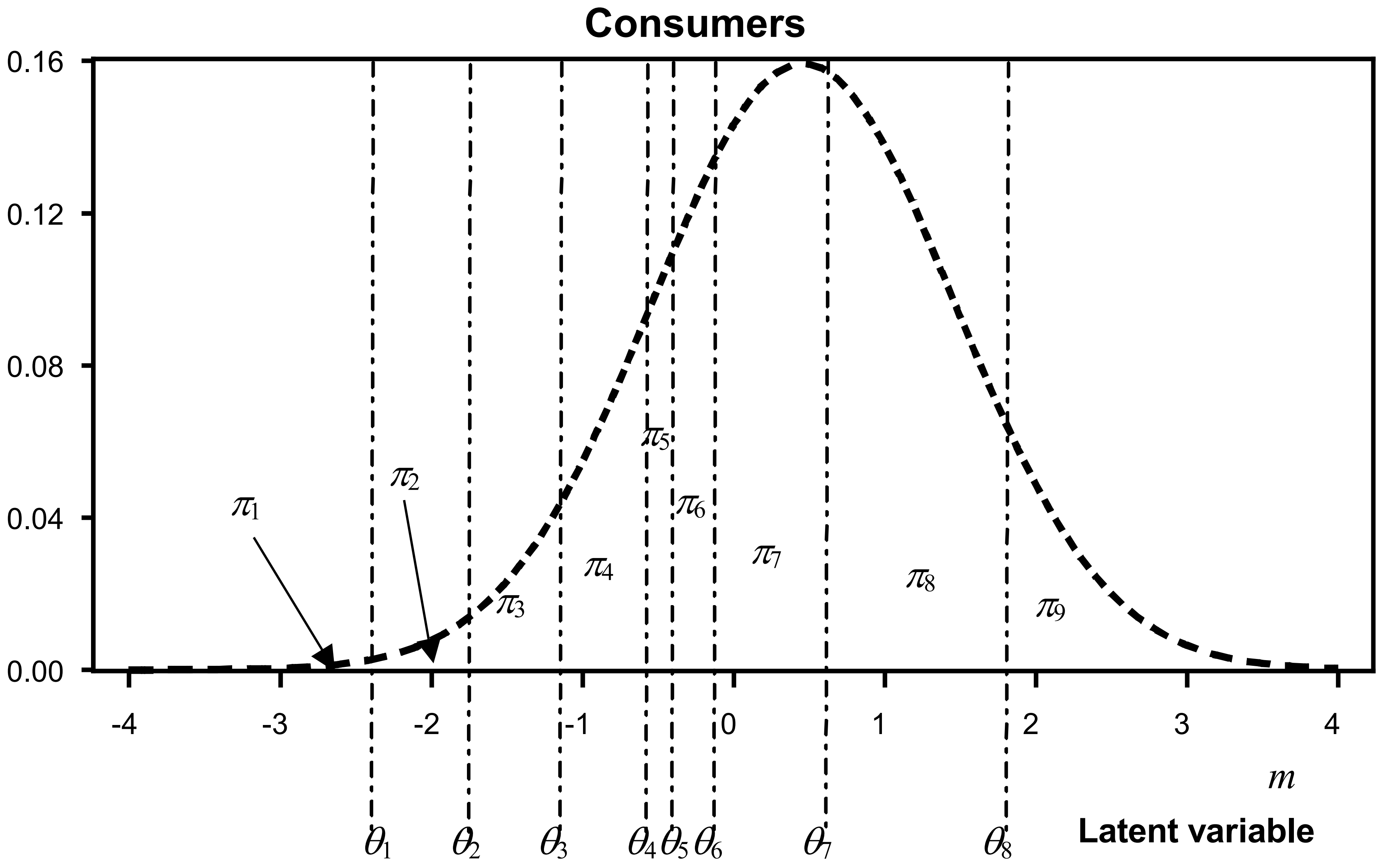
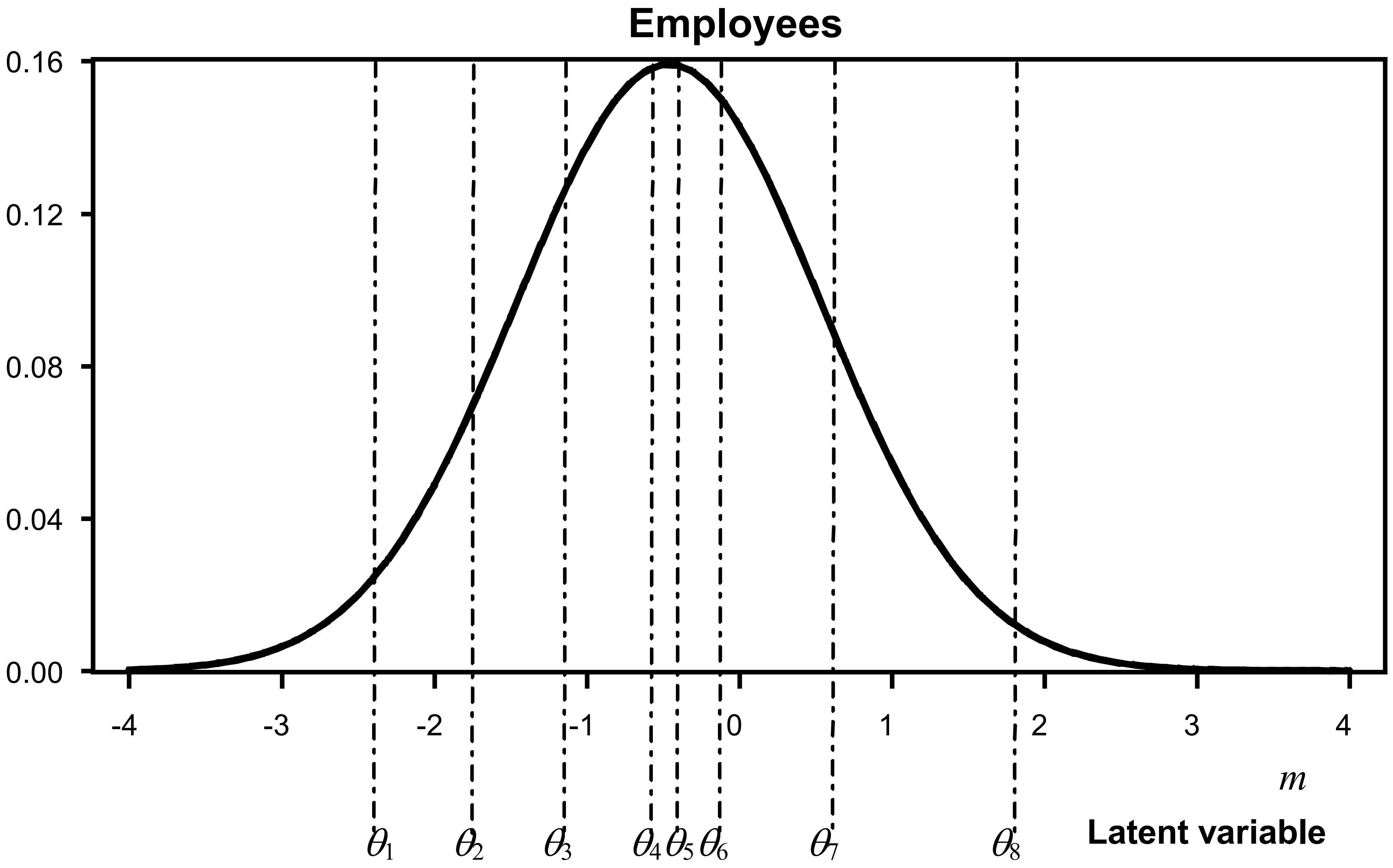


Fig. 1.1: Observed frequencies for snack food scores (Best et al., 2000).


Consumers
0.16
0.12
0.08
0.04
0.00
-4
-3
-2
-1
0
1
2
3
4
m
Latent variable
π_1
π_2
π_3
π_4
π_5
π_6
π_7
π_8
π_9
θ_1
θ_2
θ_3
θ_4
θ_5
θ_6
θ_7
θ_8


Employees
0.16
0.12
0.08
0.04
0.00
-4
-3
-2
-1
0
1
2
3
4
m
Latent variable
θ_1
θ_2
θ_3
θ_4
θ_5
θ_6
θ_7
θ_8

Latent variables

Latent variable	→	observed score
m		y
(quantitative)		(ordered categorical)

"… increasing the energy of the physical stimulus, or the concentration or amount of food ingredient, should result in an increase in how strong something feels, looks, smells, or tastes. We increase the salt in a soup, and it tastes saltier."

(Lawless and Heymann, 1998, p. 208)

Multinomial probabilities

- Threshold values θ_1, θ_2, ...
- Multinomial probabilities π_1, π_2, …

$$\pi_1 = \Phi(\theta_1 - \eta)$$

$$\pi_2 = \Phi(\theta_2 - \eta) - \Phi(\theta_1 - \eta)$$

$$\pi_3 = \Phi(\theta_3 - \eta) - \Phi(\theta_2 - \eta)$$

.

.

.

$$\pi_{I-1} = \Phi(\theta_I - \eta) - \Phi(\theta_{I-1} - \eta)$$

$$\pi_I = 1 - \Phi(\theta_I - \eta)$$

Φ = standard normal c.d.f.

η = mean on the latent scale

I = number of categories

Linear modelling

Expected value on latent scale:

$$\eta_k = \mu + \alpha_k$$

where

η_k = mean for the *k*-th panel
(k = 1: consumers k = 2: employees)

α_k = effect of *k*-th panel

Random variable on latent scale:

$$m_{jk} = \eta_k + e_{jk}$$

$$e_{jk} \sim N(0,1)$$

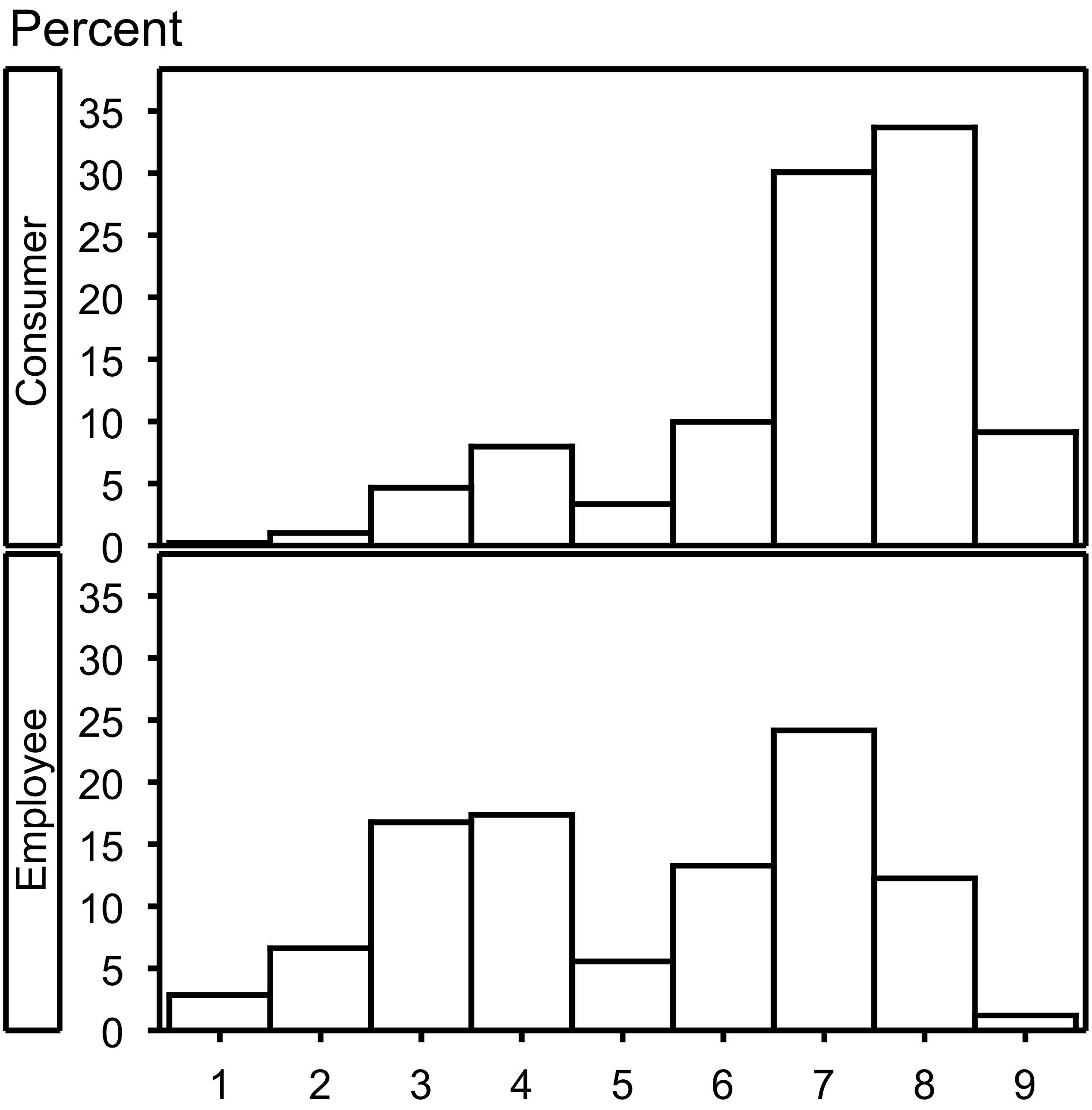


Fig. 1.3: Fitted frequencies of scores for snack food scores.

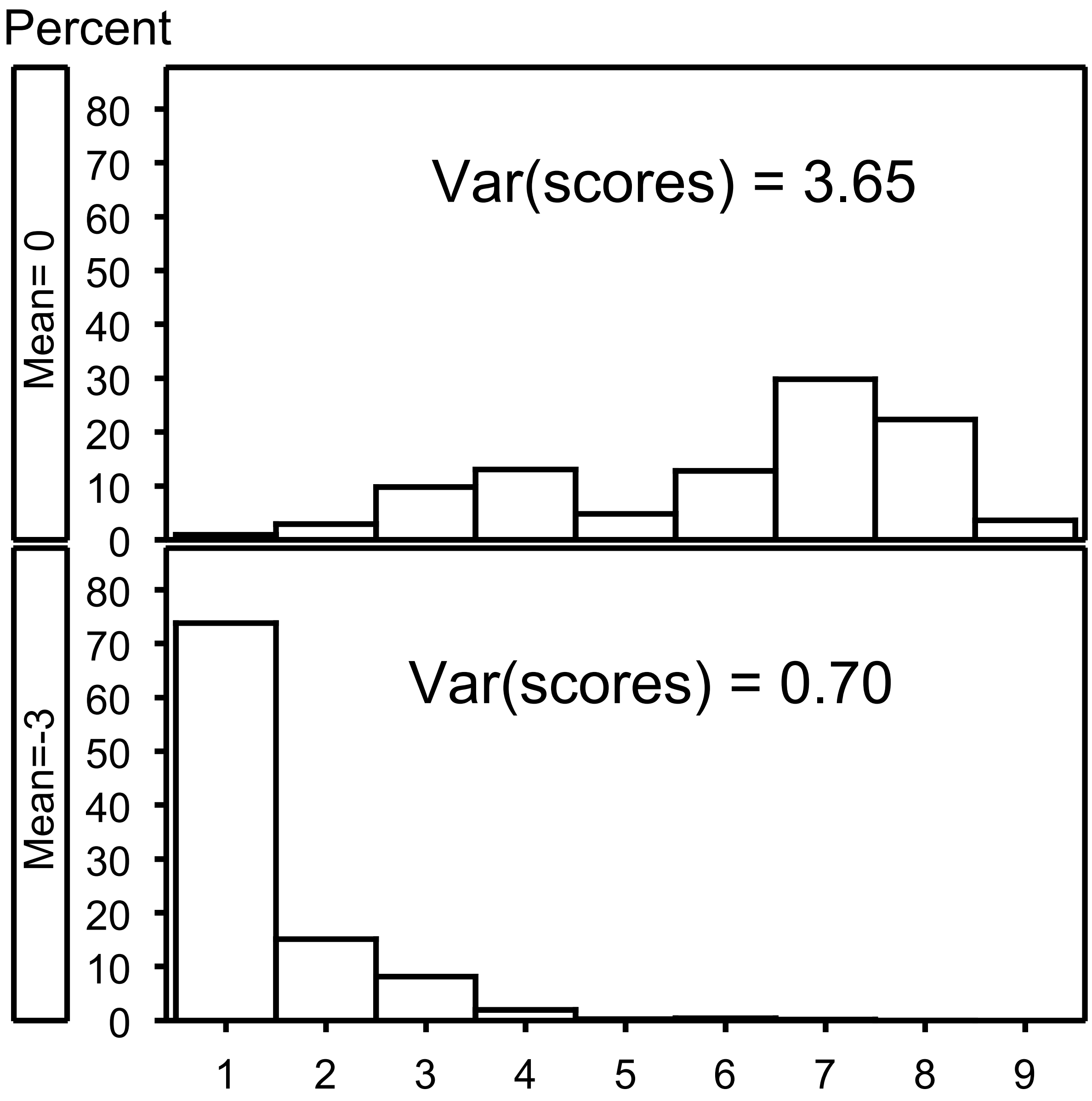


Fig. 1.4: Fitted frequencies based on threshold model for snack food scores, assuming means of $\eta = 0$ and $\eta = -3$ for the latent variable.

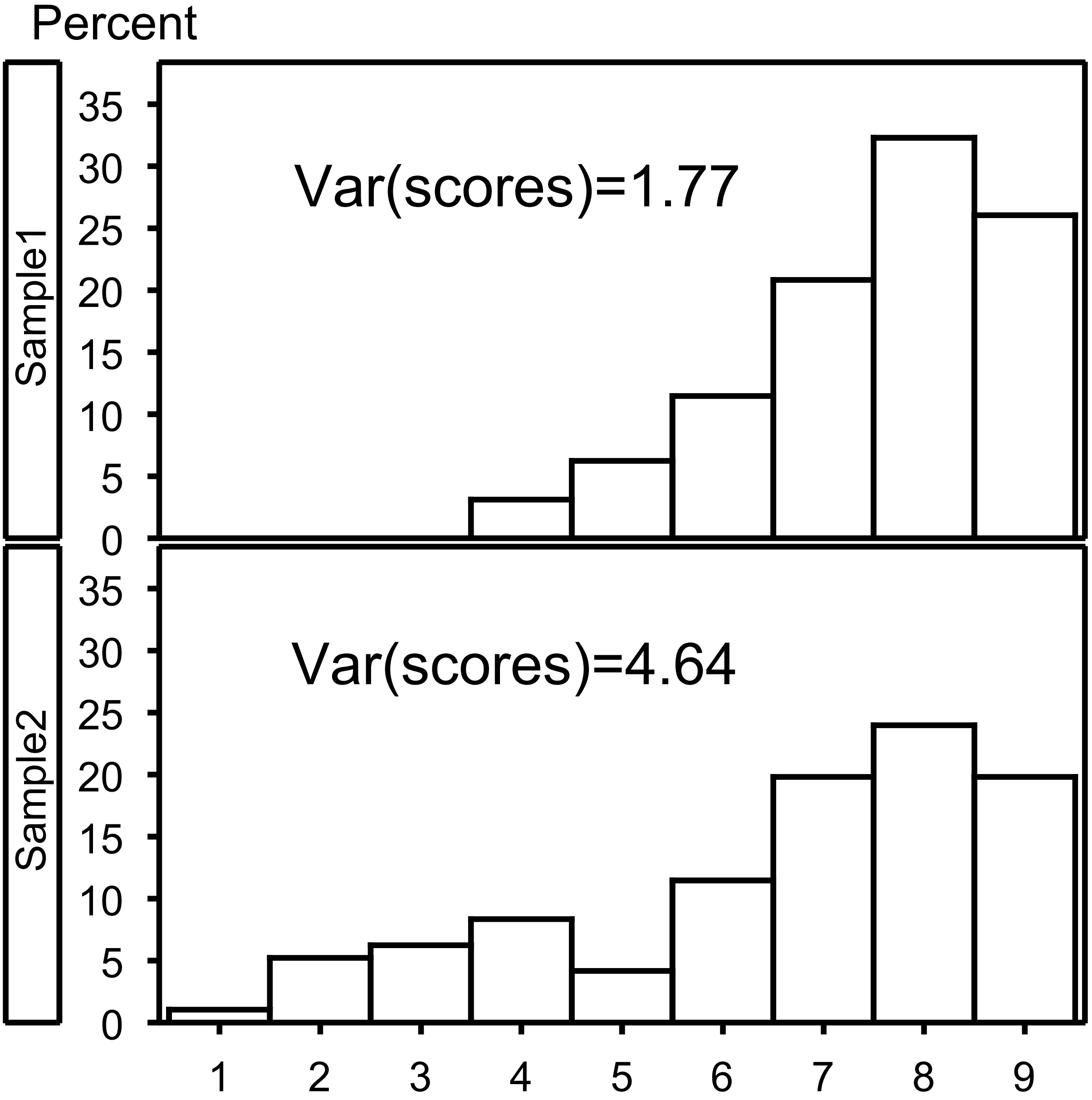


Fig. 1.5: Histogram of observed overall acceptance data (Villanueva et al., 2000). Scores display significant variance heterogeneity by an LR-test, assuming normality ($p = < 0.0001$).

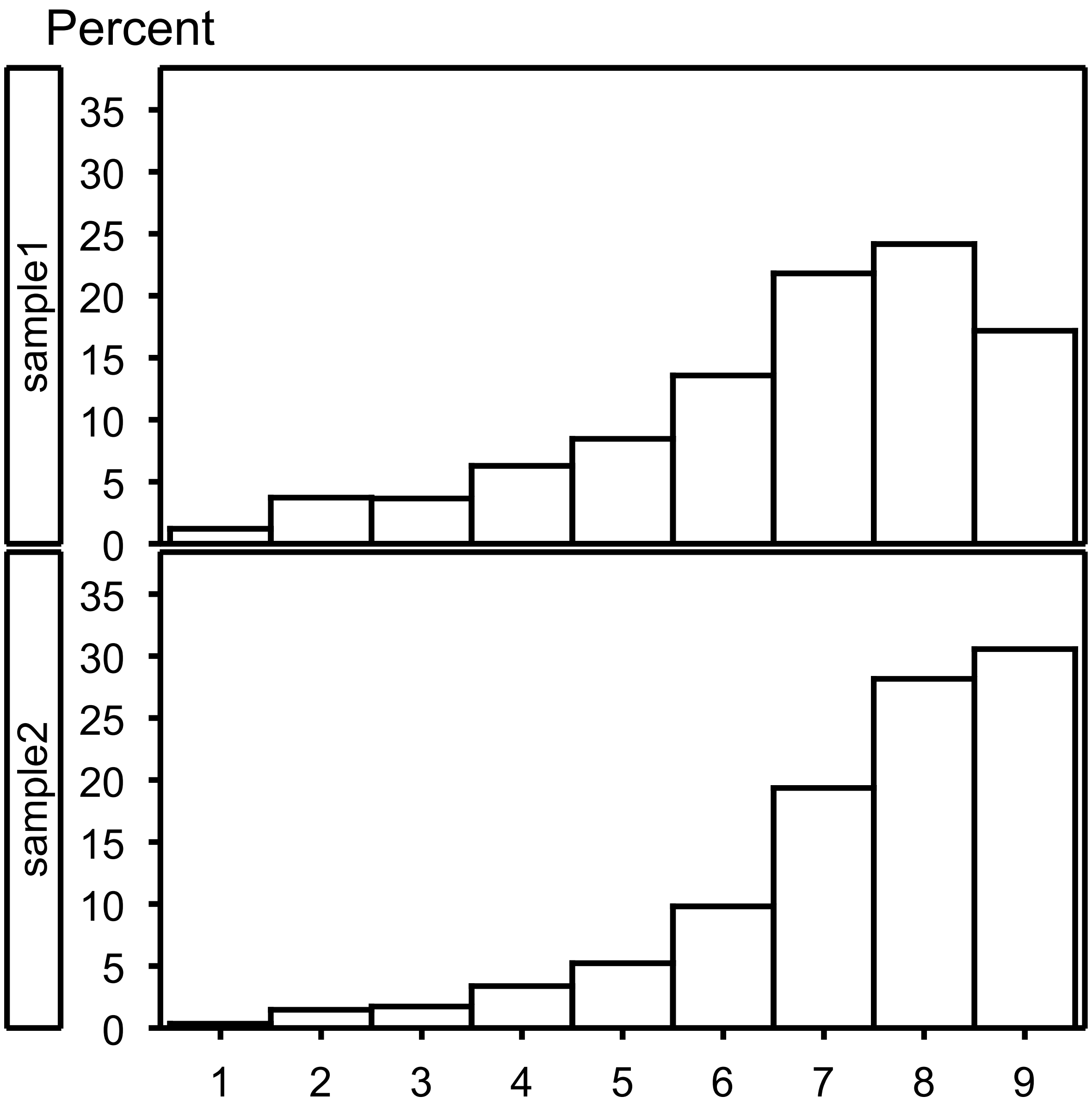


Fig. 1.6: Fitted histogram of overall acceptance data (Villanueva et al., 2000), assuming threshold model with homogeneous variance (–2 log *L* = 1104.0).

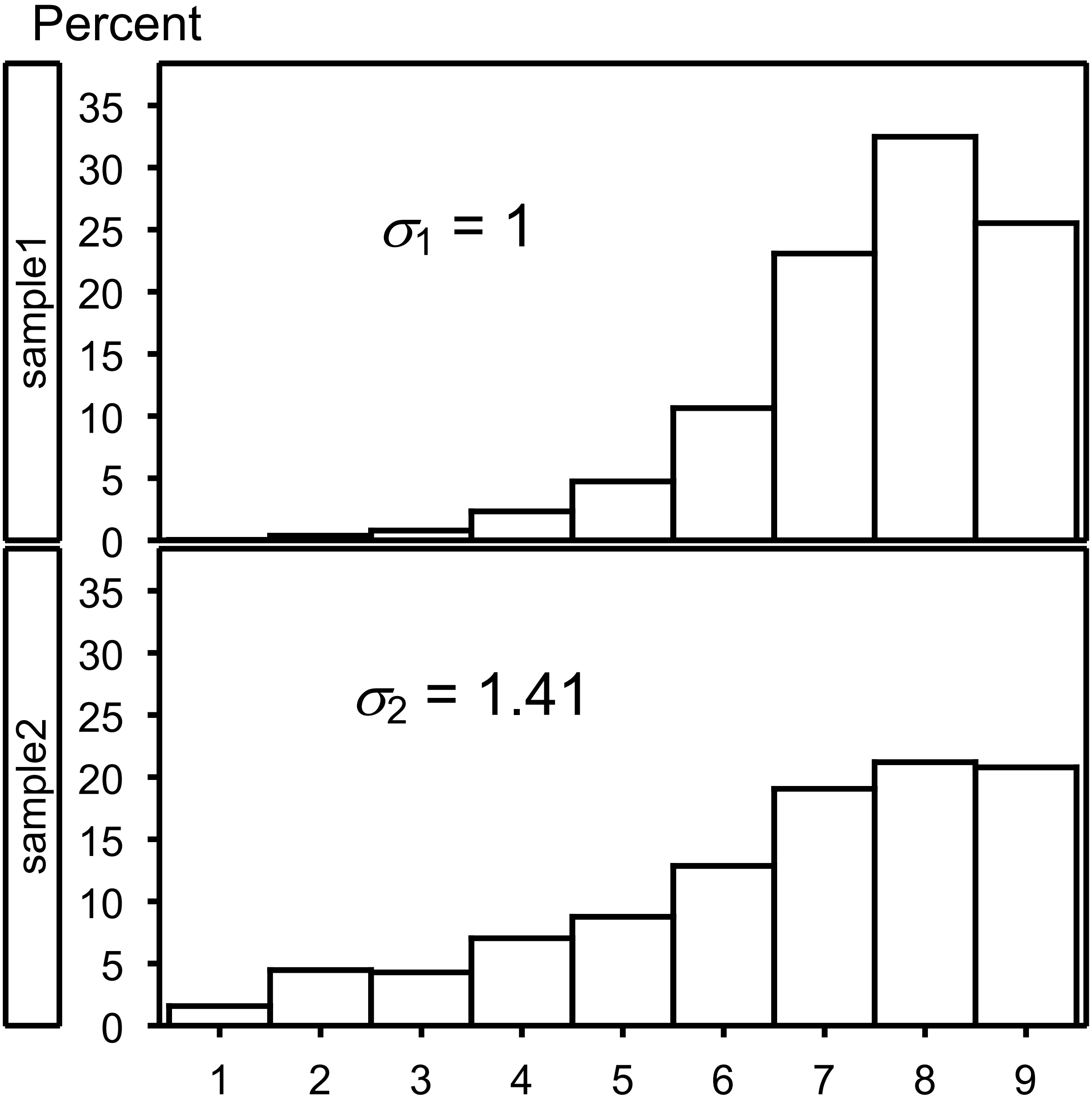


Fig. 1.7: Fitted histogram of overall acceptance data (Villanueva et al., 2000), assuming threshold model with heterogeneous variance ($-2 \log L = 1097.0$). σ_k = standard deviation on latent scale for *k*-th product.

Unequal variance on latent scale

⇒ invalidates ANOVA of observed scores!

Example:

- Date ganerated by threshold model with three categories (θ_1 = −1 and θ_2 = 1)

Parameter	Group 1	Group 2
Latent scale:		
Mean	0.9	1.0
Variance	1.0	4.0
Observed scale:		
Mean	2.4	2.2

2. Case studies

Case study 1:

Consumer preferences for on farm and industrially processed quarg samples

- nine point hedonic scale

  1 = dislike extremely to

  9 = like extremely
- 152 consumers
- Seven quarg products

An initial model

$$m_{ij} = \alpha_i + u_j + e_{ij} \qquad (2.1)$$

where

m_{ij} = latent variable for *i*-th product and *j*-th assessor

α_i = expected value of *i*-th product

u_j = random effect of *j*-th assessor;

$$u_j \sim N(0, \sigma_u^2)$$

e_{ij} = random deviation of *j*-th assessor for *i*-th product;

$$e_{ij} \sim N(0, 1)$$

$$\mathrm{corr}(m_{ij}, m_{i'j}) = \frac{\sigma_u^2}{\sigma_u^2 + 1}$$

Table 2.1: The Johnson System for u_j.

Distribution	§Model for u_j	Support
S_L	$u_j = \exp(w_j)$	$0 < u_j \le \infty$
S_B	$u_j = \phi \dfrac{\exp(w_j)}{1+\exp(w_j)}$	$0 \le u_j \le \phi$
S_U	$u_j = \phi \sinh(w_j)$	$-\infty < u_j \le \infty$

§: $w_j \sim N(\mu_w, \sigma_w^2)$

Factor-analytic models

$$m_{ij} = \alpha_i + \lambda_i u_j + e_{ij}$$

λ_i = factor loading for *i*-th product

u_j = latent value for *j*-th assessor;

$u_j \sim \mathrm{N}(0,1)$

e_{ij} = error; $e_{ij} \sim \mathrm{N}(0,1)$

Model implies heterogeneity:

$$\mathrm{var}(m_{ij}) = \lambda_i^2 + 1$$

$$\mathrm{corr}(m_{ij}, m_{i'j}) = \frac{\lambda_i \lambda_{i'}}{\sqrt{\left(\lambda_i^2 + 1\right)\left(\lambda_{i'}^2 + 1\right)}}$$

Model selection by AIC

Akaike Information Criterion = AIC:

$$AIC = -2 \log L + 2p$$

where

$\log L$ = maximized log-likelihood

p = number of parameters

"Smaller is better"

Table 2.2: Model fits for quarg data.

Model for latent variable	Distribution for random effects	*AIC*
$m_{ij} = \alpha_i + e_{ij}$	--------	4464.7
$m_{ij} = \alpha_i + u_j + e_{ij}$	$u_j \sim N(0, \sigma_u^2)$	4442.8
	$u_j \sim S_L(\phi, \mu_w, \sigma_w^2)$	4443.3
	$u_j \sim S_B(\phi, \mu_w, \sigma_w^2)$	4445.8
	$u_j \sim S_U(\phi, \mu_w, \sigma_w^2)$	4445.0
$m_{ij} = \alpha_i + \lambda_i u_j + e_{ij}$	$u_j \sim N(0,1)$	4439.4
$m_{ij} = \alpha_i + \lambda_i u_{j1} + \lambda_{i2} u_{j2} + e_{ij}$	$u_{jk} \sim N(0,1)$	4407.3
$m_{ij} = \alpha_i + \lambda_i u_{j1} + \lambda_{i2} u_{j2} + \lambda_{i3} u_{j3} + e_{ij}$	$u_{jk} \sim N(0,1)$	4403.6

Factor-analytic threshold model (two multiplicative terms):

Product	Estimate of α_i
1	2.10 bc
2	2.20 c
3	1.57 d
4	1.96 be
5	1.70 de
6	2.11 bc
7	2.88 a
$\overline{SED}$	0.132

ANOVA of scores:

Product	Score mean
1	5.43 b
2	5.60 b
3	4.45 c
4	5.17 b
5	4.63 c
6	5.41 b
7	6.67 a
LSD	0.445

Estimates in a column followed by the same letter are not significantly different.

Artificial example

- Four treatments (T1, T2, T3, T4)
- Comparison of T1 and T4 significant

Initial display:

T1 a
T2 a
T3 a
T4 a

⇒ misses significance of T1-T4 comparison

Insertion:

T1 a b	T1 ~~a~~ b	T1 b
T2 a b	T2 a b	T2 a b
T3 a b	T3 a b	T3 a b
T4 a b	T4 a ~~b~~	T4 a

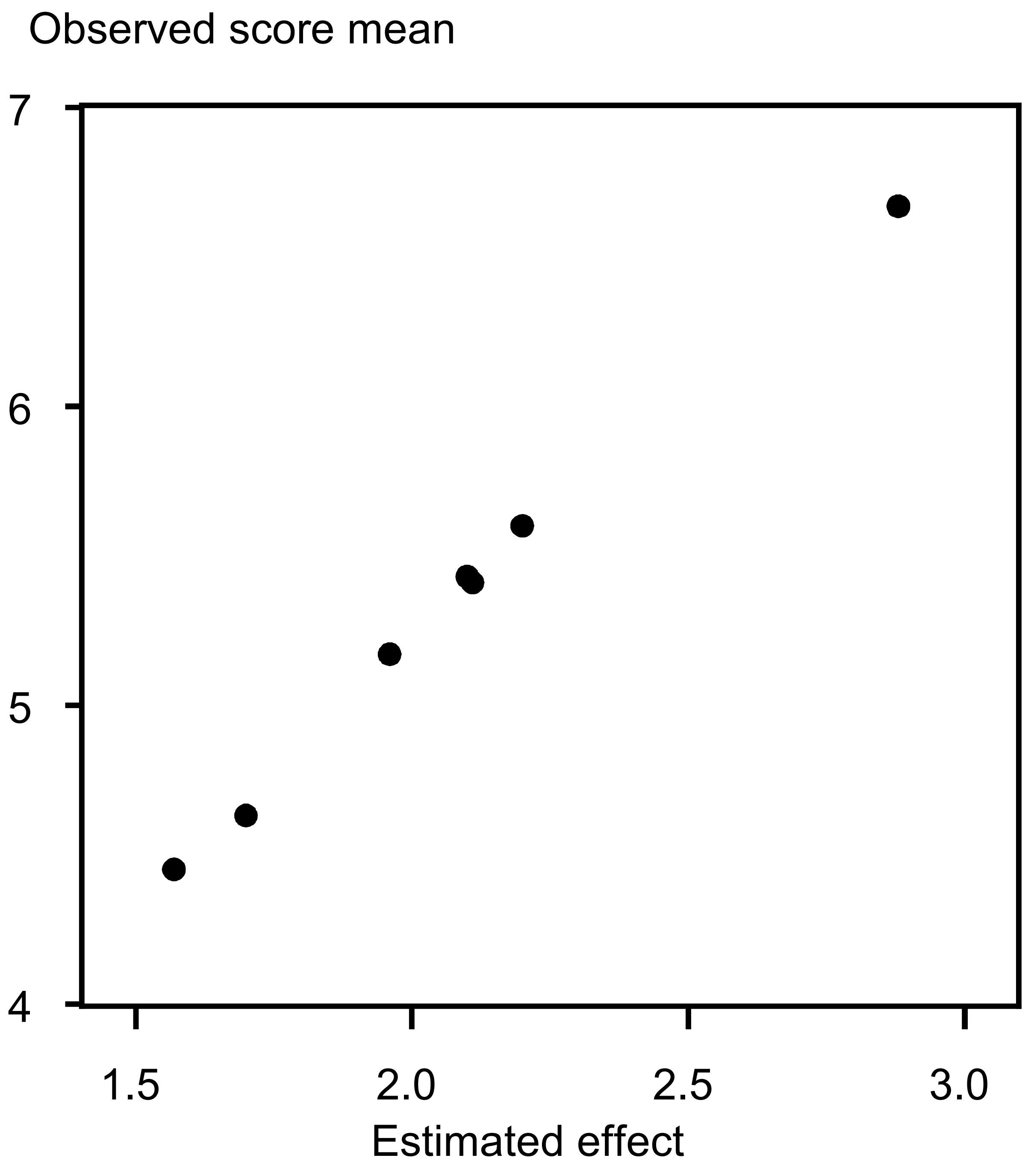


Fig. 2.1: Plot of observed score means (ANOVA) versus estimated effects under threshold model for seven quarg products.

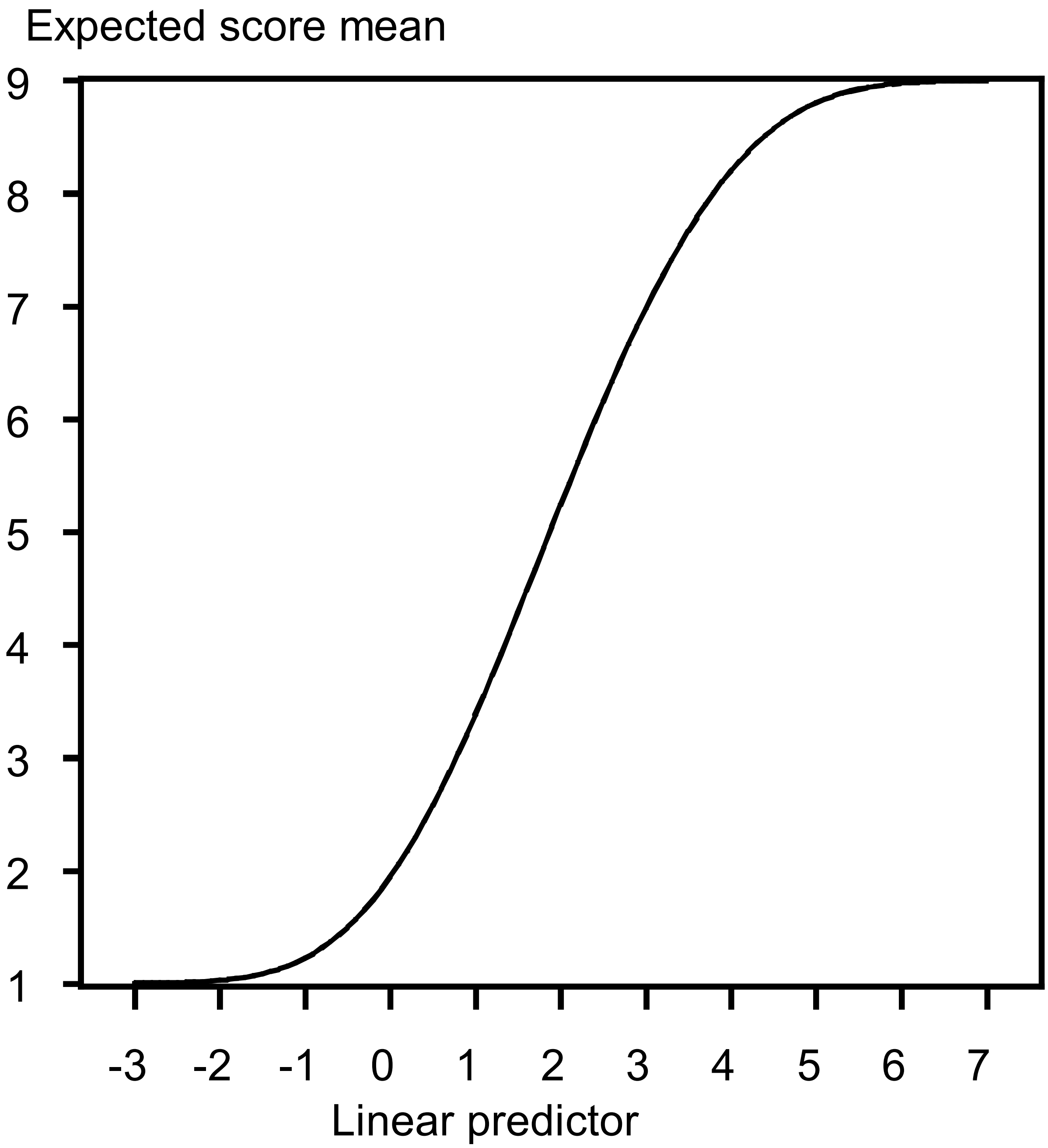


Fig. 2.4: Plot of score mean versus linear predictor under estimated threshold model for quarg data.

Case study 2:

Sensory descriptive analysis of on farm and industrial processed quarg samples by a trained panel

- trained panel of 15 assessors
- 8 quarg samples (B1-B8)
- 12 sensory attributes
- 12 sessions
- ordinal scale

  (0 = non- detectable to 5 = very strong)

Mixed linear model (per attribute):

$$m_{ijk} = u_j + \tau_i + \alpha_k + (\alpha\tau)_{ik} + e_{ijk} \qquad (2.6)$$

where

u_j = effect of *j*-th assessor

τ_i = effect of *i*-th product

α_k = effect of *k*-th session (**random**)

$(\alpha\tau)_{ik}$ = interaction of *i*-th product and *k*-th session (**random**)

Principal component analysis

Two standardizations:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{I}\left(\tau_i - \bar{\tau}_{\bullet}\right)}{\left(I-1\right)}} \tag{2.7a}$$

$$\sigma = \sqrt{\sigma_{\alpha\tau}^{2} + 1} \tag{2.7b}$$

⇒ singular value decomposition of standardized values

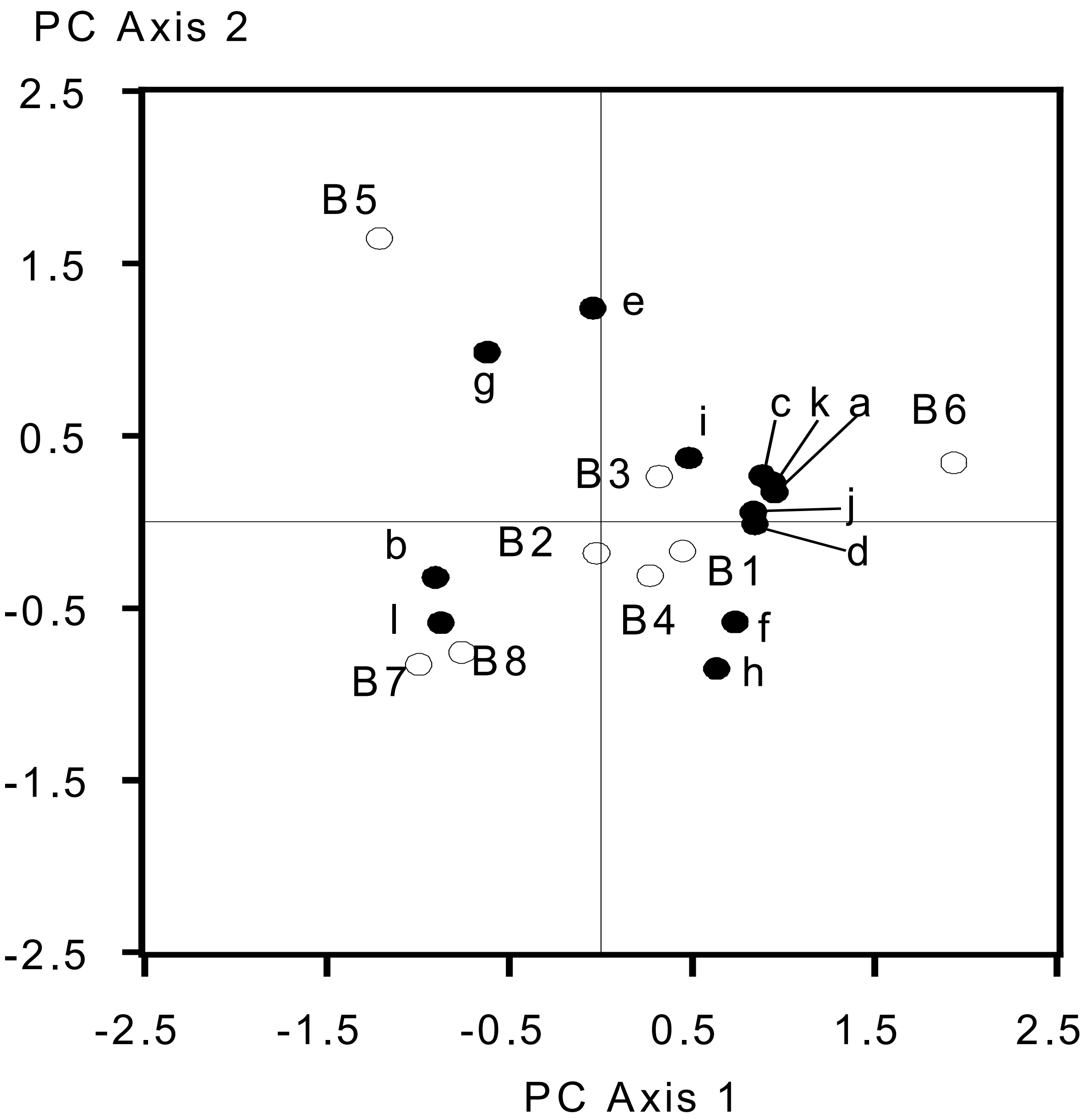


Fig. 2.5: Biplot for PCA of 8 quarg samples and 12 sensory attributes based on standardization (2.7a). 83% of variance are explained. B1-B8: samples; letters: attributes)

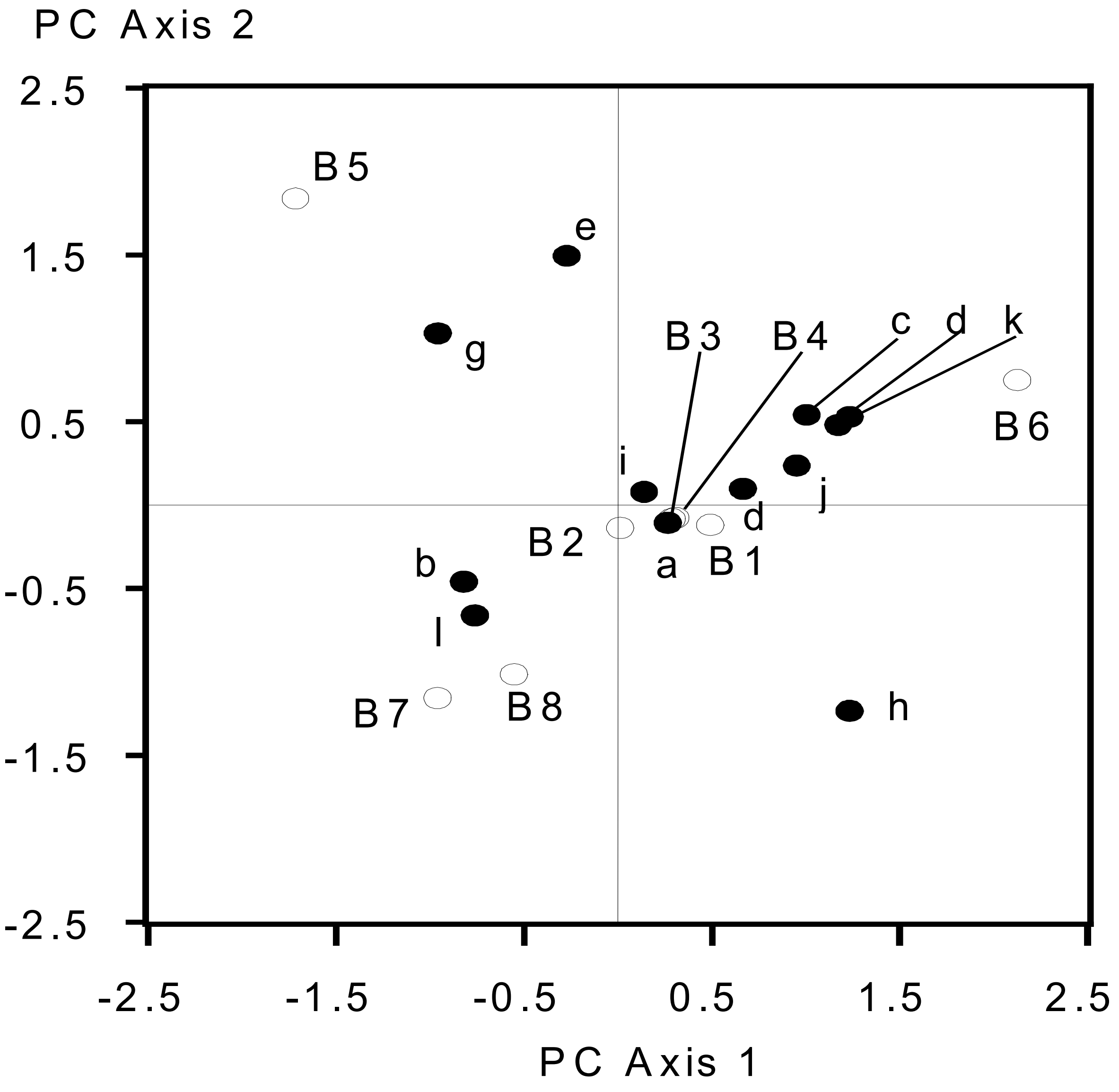


Fig. 2.6: Biplot for PCA of 8 quarg samples and 12 sensory attributes based on standardization (2.7b). 91% of variance are explained. B1-B8: samples; letters: attributes).

Heterogeneity of variance

- We hypothesized that variance was larger for samples B1 through B4
- Fit model with heterogeneity in variance of interaction $(\alpha\tau)_{ik}$.

$\sigma^2_{\alpha\tau(1)}$: variance for products B1 to B4

$\sigma^2_{\alpha\tau(2)}$: variance for products B5 to B8

Table 2.7: Variance components (case study 2). § *: p-value < 0.05.

Attributes	Homogeneous B1-B8 $\sigma^2_{\alpha\tau}$	Heterogeneous B1-B4 $\sigma^2_{\alpha\tau(1)}$	 B5-B8 $\sigma^2_{\alpha\tau(2)}$	LR-test[§]
Granular	0.96	1.01	0.82	ns
Separation of whey	0.71	1.24	0.29	*
Creamy yellow color	0.43	0.90	0.09	*
Bitter	0.04	0.01	0.07	ns
Sour	0.01	0.00	0.07	ns
Flavour of starter cultures	0.12	0.00	0.29	*
Flavour of yoghurt	0.15	0.00	0.37	*
Rancid	0.43	0.06	1.78	*
Other flavour	0.20	0.17	0.25	ns
Firm	1.09	2.37	0.72	*
Gritty	1.40	2.03	0.21	ns
Creamy	0.51	0.32	0.77	ns

Inter-assessor agreement

- Assessed separately for each session

$$m_{ij} = u_j + \tau_i + e_{ij} \tag{2.8}$$

u_j = effect of *j*-th assessor

τ_i = effect of *i*-th product

$$\tau_i \sim N(0, \sigma_\tau^2)$$

e_{ij} = random error; standard normal

Intra-class correlation:

$$\rho_{IC} = \frac{\sigma_\tau^2}{1 + \sigma_\tau^2}$$

Table 2.9: Estimates of the intra-class correlation ρ_{IC} and Kendall's coefficient of concordance for quarg data.

	Attributes		
Session	granular	colour	sour
Kendall's measure of concordance:			
1	0.87	0.71	0.67
3	0.64	0.64	0.80
7	0.68	0.80	0.77
9	0.79	0.56	0.85
11	0.83	0.78	0.74
12	0.64	0.32	0.78
Intraclass correlation (ρ_{IC}):			
1	0.86	0.78	0.60
3	0.64	0.81	0.82
7	0.79	0.76	0.80
9	0.79	0.60	0.74
11	0.81	0.77	0.88
12	0.74	0.34	0.90

Case study 3

Consumer preferences for raspberry beverages

- Two regions (geographies)
- 28 beverages
- Panel 1 (123 assessors):
  1 = dislike extremely, 9 = like extremely
- Panel 2 (227 assessors):
  1 =bad, 5 =good

Questions:

- Difference in accuracy among scales
- Differences in ratings among regions
- Correlation among attributes

Difference in accuracy

Linear model:

$$m_{ij} = \alpha_i + u_{jk} + e_{ijk} \tag{2.9}$$

α_i = effect of *i*-th product

u_{jk} = effect of *j*-th assessor on *k*-th scale

$$u_{jk} \sim N(0, \sigma^2_{u(k)})$$

e_{ijk} = error

$$e_{ijk} \sim N(0, \sigma^2_{e(k)})$$

Separate set of thresholds for both scales!

Hypotheses of interest:

H_{01}: $\sigma^2_{u(1)} = \sigma^2_{u(2)}$

H_{02}: $\sigma^2_{e(1)} = \sigma^2_{e(2)}$

Table 2.9: Likelihood statistics and variance component estimates for liking, raspberry beverages.

Geography 1:

Model restriction	$-2 \log L$	$\sigma^2_{e(1)}$	$\sigma^2_{e(2)}$	$\sigma^2_{u(1)}$	$\sigma^2_{u(2)}$
None	41767.6	1.06	1	0.43	0.21
$\sigma^2_{u(1)} = \sigma^2_{u(2)}$	41782.2	0.94	1	0.28	0.28
$\sigma^2_{e(1)} = \sigma^2_{e(2)}$	41768.5	1	1	0.39	0.21
$\sigma^2_{e(1)} = \sigma^2_{e(2)};\ \sigma^2_{u(1)} = \sigma^2_{u(2)}$	41783.9	1	1	0.30	0.30

Geography 2:

Model restriction	$-2 \log L$	$\sigma^2_{e(1)}$	$\sigma^2_{e(2)}$	$\sigma^2_{u(1)}$	$\sigma^2_{u(2)}$
None	23553.0	1.19	1	0.53	0.19
$\sigma^2_{u(1)} = \sigma^2_{u(2)}$	23569.3	1.03	1	0.29	0.29
$\sigma^2_{e(1)} = \sigma^2_{e(2)}$	23558.8	1	1	0.38	0.19
$\sigma^2_{e(1)} = \sigma^2_{e(2)};\ \sigma^2_{u(1)} = \sigma^2_{u(2)}$	23569.5	1	1	0.28	0.28

Differences in ratings among regions

$$m_{ijh} = \alpha_i + \gamma_{ih} + u_{jh} + e_{ijh} \tag{2.11}$$

where

α_i = main effect of *i*-th product
γ_{ih} = deviation of *i*-th product in *h*-th geography (h = 1, 2)
u_{jh} = effect of *j*-th assessor in *h*-th geography,
$u_{jh} \sim N(0, \sigma_u^2)$
e_{ijh} = error (standard normal)

Ratings the same?

H_0: $\gamma_{i1} = \gamma_{i2}$ for every product *i*

Table 2.12: Comparison of mixed threshold models for testing H$_0$: $\gamma_{1i} = \gamma_{2i}$ (consumers behave identically in both regions). Attribute: liking (case study 3).

Scale	Model for m_{ijh}	*AIC*	Test of H$_0$: $\gamma_{1i} = \gamma_{2i}$	
			Wald-χ^2	p-value
5 point	$\alpha_i + \gamma_{ih} + u_{jh} + e_{ijh}$	27360.4	2.20	0.0006
	$\alpha_i + \gamma_{ih} + \lambda_i u_{jh} + e_{ijh}$	27309.0	2.21	0.0006
9 point	$\alpha_i + \gamma_{ih} + u_{jh} + e_{ijh}$	37957.8	3.09	<0.0001
	$\alpha_i + \gamma_{ih} + \lambda_i u_{jh} + e_{ijh}$	37895.5	3.05	<0.0001

Correlation among attributes

- liking and appearance
- geography 1
- five-point ordinal scale

Bivariate mixed model:

$$m_{ij1} = \alpha_{i1} + u_{j1} + e_{ij1} \qquad (2.13)$$
$$m_{ij2} = \alpha_{i2} + u_{j2} + e_{ij2}$$

where

α_{ih} = effect of *i*-th product for *h*-th attribute
u_{jh} = effect of *j*-th assessor for *h*-th attribute
e_{ijh} = residual corresponding to m_{ijh}

Distributional assumptions:

$$\begin{pmatrix} u_{j1} \\ u_{j2} \end{pmatrix} \sim BVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u1}^2 & \rho_u \sigma_{u1} \sigma_{u2} \\ \rho_u \sigma_{u1} \sigma_{u2} & \sigma_{u2}^2 \end{pmatrix}\right]$$

$$\begin{pmatrix} e_{ij1} \\ e_{ij2} \end{pmatrix} \sim BVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_e \\ \rho_e & 1 \end{pmatrix}\right]$$

where

BVN(.,.) = bivariate normal distribution
ρ_u = between-product correlation
ρ_e = within-product correlation

Bivariate multinomial probabilities of observed scores:

$$\pi_{ab} = P(y_1 = a, y_2 = b) = \Phi\left(\theta_{1a} - \eta_1, \theta_{2b} - \eta_2, \rho_e\right) - \sum_{\substack{r=1 \\ rs<ab}}^{a} \sum_{s=1}^{b} \pi_{rs} \tag{2.14}$$

$$(a = 1, \ldots, 5;\ b = 1, \ldots, 5)$$

with

y_1, y_2 = observed scores for liking and appearance

Φ = bivariate standard normal c.d.f

$\theta_{1a},\ \theta_{2b}$ = thresholds; $\theta_{15} = \theta_{25} = \infty$

$\eta_1,\ \eta_2$ = linear predictors (means) on latent scale

Result: $\hat{\rho}_e = 0.588 \quad (s.e. = 0.011)$ (within-product analysis)

Case study 4

Recovery of inter-assessor information

- Sensory panels of 33 assessors
- Three groups of cheese products
- Hedonic 9-point scale
- Tests in incomplete blocks: five or six products per assessor

Type of analysis	Assumption about assessor effects
Intra-assessor analysis	fixed
Recovery of inter-assessor information	random

Table 2.14: Average standard errors of a difference (s.e.d.).

	Average s.e.d.		
Use of inter-assessor Information	Hard cheese	Fresh cheese with additives	Fresh cheese
yes	0.375	0.407	0.369
no	0.385	0.423	0.380

3. Numerical issues

y = observed scores
u = random effects

Joint density of u and y:

$$\phi_{yu}(y, u) = \phi_{yu}(y|u)\phi_u(u) \tag{3.1}$$

where

$\phi_{yu}(y|u)$ = conditional density of y, given u (multinomial)
$\phi_u(u)$ = marginal density of u (normal)

Estimation ⇒ maximize

$$\phi_y(y) = \int \phi_{yu}(y \mid v)\phi_u(v)dv \tag{3.2}$$

Numerical integration

- Approximations of the likelihood

$\Rightarrow$ poor performance for small samples

- Gaussian quadrature

$\Rightarrow$ Computationally demanding
$\Rightarrow$ Handles only one subject level

All methods rely on asymptotic theory!

4. Simulations

4.1 No random effects

Hypothesis:

Show that ANOVA may be problematic with ordinal data

Simulate according to:

$$m_{ijk} = \alpha_i + u_j + e_{ijk} \qquad (4.1)$$

where

α_i = product effect (*i* = 1, …, *I*)
u_j = random assessor effect (*j* = 1, …, *J*),
e_{ijk} = error (standard normal distribution)
(*k* = 1, .., *R*)

Three (nine) categories

Simulations under:

- global null hypothesis
- departure from global null

Teste evaluated:

Global H_0:
- ANOVA F-test
- LR test for threshold model

H_0: $\alpha_1 = \alpha_2$:
- ANOVA-LSD
- paired t-test
- Wald-test for threshold model

Table 4.5: Simulation results for threshold model with fixed effects, Type I error at nominal significance level of 5%, Global H_0 false, nine categories ($\theta_1 = -0.5$, $\theta_2 = -0.4$, $\theta_3 = -0.3$, $\theta_4 = -0.1$, $\theta_5 = 0.1$, $\theta_6 = 0.2$, $\theta_7 = 0.3$, $\theta_8 = 0.5$, $\sigma_u = 0$). Test of H_0: $\alpha_1 = \alpha_2$.

R	J	I	α_1	α_2	$\alpha_{i>2}$	ANOVA LSD	Threshold Wald	Paired t-test
1	50	20	0	0	2	0.420	0.082	0.061
1	50	20	2	2	0	0	0.019	0.036
4	20	20	0	0	2	0.410	0.068	0.058
4	20	20	2	2	0	0	0.039	0.049

4.2 Pairwise comparison with heteroscedasticity on latent scale

Hypothesis:

Heteroscedasticity on latent scale of threshold model invalidates ANOVA

Simulated according to:

$$m_{ij} = \alpha_i + e_{ij} \tag{4.2}$$

where

α_i = product effect (*i* = 1, 2)
e_{ij} = error of *j*-th replicate for *i*-th product ($j = 1, \ldots, J_i$)
σ_1 = standard deviation of e_{1j} (product 1)
σ_2 = standard deviation of e_{2j} (product 2)

Nine categories.

Tests evaluated:

- Simple t-test
- Satterthwaite t-test (heteroscedastic)
- Standard threshold model
- Threshold model with heteroscedastic errors

Table 4.6: Simulation results for heteroscedastic threshold model, Type I error at nominal significance level of 5%, $\alpha_1 = \alpha_2 = 0$. Test of H_0: $\alpha_1 = \alpha_2$.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	σ_2	t-test	Threshold model $\sigma_1 = \sigma_2$	Threshold model $\sigma_1 \neq \sigma_2$
$J_1 = 50$; $J_2 = 150$:											
–2.0	–1.8	1.0	1.2	1.4	1.6	1.8	2.0	2	0.624	0.023	0.063
–2.0	–1.7	–1.5	0.0	1.0	1.5	1.7	2.0	2	0.085	0.005	0.045
–2.0	–1.3	0.5	0.9	1.2	1.5	1.8	2.0	2	0.409	0.008	0.050
–2.0	–1.5	–1.0	0.0	0.5	1.0	1.5	2.0	2	0.065	0.007	0.038
–2.0	–1.8	1.0	1.2	1.4	1.6	1.8	2.0	1.4	0.260	0.018	0.057
–2.0	–1.7	–1.5	0.0	1.0	1.5	1.7	2.0	1.4	0.072	0.029	0.055
–2.0	–1.3	0.5	0.9	1.2	1.5	1.8	2.0	1.4	0.209	0.022	0.052
–2.0	–1.5	–1.0	0.0	0.5	1.0	1.5	2.0	1.4	0.058	0.024	0.057

										Threshold model	
θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	σ_2	t-test	$\sigma_1 = \sigma_2$	$\sigma_1 \neq \sigma_2$
J_1 = 150; J_2 = 50:											
–2.0	–1.8	1.0	1.2	1.4	1.6	1.8	2.0	2	0.336	0.151	0.053
–2.0	–1.7	–1.5	0.0	1.0	1.5	1.7	2.0	2	0.071	0.132	0.068
–2.0	–1.3	0.5	0.9	1.2	1.5	1.8	2.0	2	0.242	0.107	0.048
–2.0	–1.5	–1.0	0.0	0.5	1.0	1.5	2.0	2	0.041	0.094	0.042
–2.0	–1.8	1.0	1.2	1.4	1.6	1.8	2.0	1.4	0.130	0.110	0.057
–2.0	–1.7	–1.5	0.0	1.0	1.5	1.7	2.0	1.4	0.056	0.099	0.054
–2.0	–1.3	0.5	0.9	1.2	1.5	1.8	2.0	1.4	0.092	0.088	0.059
–2.0	–1.5	–1.0	0.0	0.5	1.0	1.5	2.0	1.4	0.052	0.077	0.051

4.3 One random effect

Hypothessis:

Asymptotics are even more critical for mixed models

Simulated according to:

$$m_{ijk} = \alpha_i + u_j + f_{ij} + e_{ijk} \qquad (4.3)$$

where

f_{ij} = random assessor × product interaction

Table 4.7: Simulation results for threshold model with fixed and random effects (4.3), Type I errors at nominal significance level of 5%. Global H$_0$ true, R = 4 replicates, three categories. Test of H$_0$: $\alpha_1 = \alpha_2$. \$: Wald-Test, §: LR-test.

						Overall ANOVA		Threshold model		Paired ANOVA
θ_1	θ_2	σ_u	σ_f	J	I	global\$	$\alpha_1 = \alpha_2$\$	global§	$\alpha_1 = \alpha_2$\$	$\alpha_1 = \alpha_2$
–0.5	0.5	0	0.2	10	4	0.059	0.052	0.078	0.064	0.052
–0.5	0.5	0	0.2	10	8	0.051	0.045	0.064	0.067	0.051
–1.0	1.0	0	0.2	10	4	0.034	0.051	0.092	0.087	0.051
–1.0	1.0	0	0.2	10	8	0.049	0.053	0.072	0.062	0.043
–0.5	0.5	0	0.5	10	4	0.051	0.052	0.140	0.108	0.055
–0.5	0.5	0	0.5	10	8	0.064	0.051	0.122	0.066	0.054
–1.0	1.0	0	0.5	10	4	0.034	0.045	0.122	0.081	0.047
–1.0	1.0	0	0.5	10	8	0.045	0.040	-	-	0.047

5. Conclusion

- ANOVA can fail badly with ordinal data
- Threshold model well suited to analyse ordinal data
- Many experimental settings call for mixed models
- Threshold model can be extended to cover random effects
- Allows great flexibility in answering relevant research questions
- Maximum likelihood estimation computationally demanding
- Asymptotics work for large samples only
- Tests in particular need to be taken with a gain of salt
- More simulations needed