Using modern statistical methodology for validating and reporting Patient Reported Outcomes

Karl Bang Christensen

Dept. of Biostatistics, Univ. of Copenhagen
joint DSBS/FMS Meeting
October 2, 2014, Copenhagen

Agenda

1. Indirect measurement

Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
   - Item Response Theory (IRT) models
   - Probit models

Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
   - Item Response Theory (IRT) models
   - Probit models
3. Example data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program:

Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
   - Item Response Theory (IRT) models
   - Probit models
3. Example data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program:

   Vitality (VT) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)

Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
   - Item Response Theory (IRT) models
   - Probit models
3. Example data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program:

   Vitality (VT) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)

   Physical Functioning (PF) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)

Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
   - Item Response Theory (IRT) models
   - Probit models
3. Example data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program:

   Vitality (VT) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)

   Physical Functioning (PF) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)
4. Summary

Indirect measurement

1. Binet-Simon scale, first intelligence test
   - Designed for children, measured a childs 'mental age' which compared the child with his or her age group (Binet, L'Année Psychologique, vol. 12, 191-244, 1905)

Indirect measurement

1. Binet-Simon scale, first intelligence test
    - Designed for children, measured a childs 'mental age' which compared the child with his or her age group (Binet, L'Année Psychologique, vol. 12, 191-244, 1905)
    - latent variable, can be changed

Indirect measurement

1. Karnofsky performance status
   - 0 ('Dead'), 10 ('Moribund'), 20 ('Very sick'), ..., 90 ('Able to carry on normal activity; minor symptoms'), 100

     (Karnofsky, Burchenal JH, In: MacLeod CM (ed.): Evaluation of chemotherapeutic agents. New York: Columbia University Press, pp 191-205, 1949)

Indirect measurement

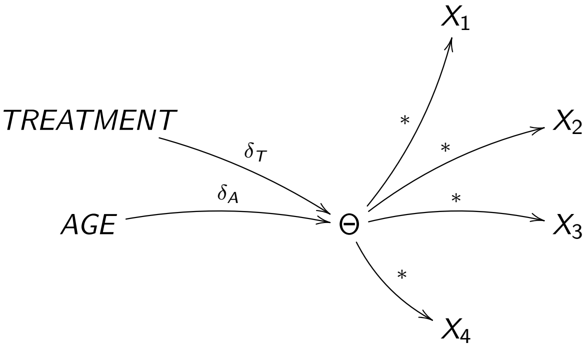
1. Karnofsky performance status
   - 0 ('Dead'), 10 ('Moribund'), 20 ('Very sick'), ..., 90 ('Able to carry on normal activity; minor symptoms'), 100

     (Karnofsky, Burchenal JH, In: MacLeod CM (ed.): Evaluation of chemotherapeutic agents. New York: Columbia University Press, pp 191-205, 1949)
   - latent variable, can be changed

Indirect measurement

1. PRO's
   - Generic (e.g. SF36) or disease specific (e.g. FACT)
   - Sub scales consisting of ordinal items
   - Often reported as standardized (e.g. zero to 100) mean scores
   - Often analyzed using linear models or nonparametric statistics

Indirect measurement



Latent variable Θ, $X_1, \ldots, X_4$ items. [*: monotone association]

Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
   - Item Response Theory (IRT) models
   - Probit models
3. Example data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program:

   Vitality (VT) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)

   Physical Functioning (PF) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)
4. Summary

IRT model for item response X_i

Item parameters $\bar{\beta}_i = (\beta_{i0}, \beta_{i1}, \beta_{i2}, \ldots)$

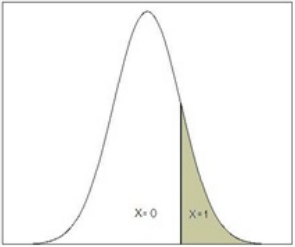
$$P(X_i = x | \Theta = \theta) = \frac{\exp(x\theta + \beta_{ix})}{K_i(\theta)} \qquad (x = 0, 1, \ldots, m_i), \quad (1)$$

where $\beta_{i0} = 0$ for convenience and

$$K_i(\theta) = \sum_{h=0}^{m_i} \exp(h\theta + \beta_{ih})$$

Alternative models for item response X_i

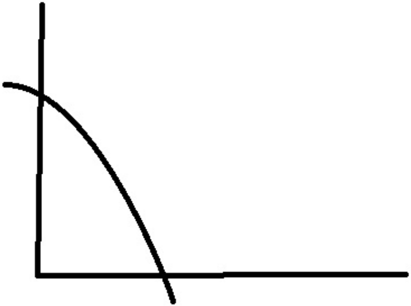
- Change linear predictor: replace $x\theta + \beta_{ix}$ with $\alpha_i(x\theta + \beta_{ix})$
- Probit model



- Regardless of these choices
  - Conditional probability of each response option given θ

Conditional probability of each response option given θ

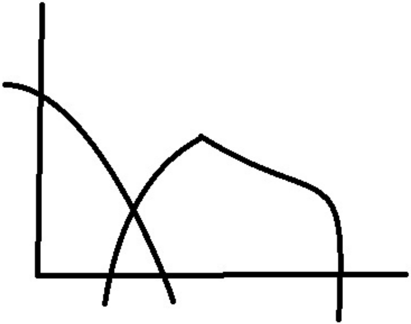
$$P(X = 0|\Theta = \theta)$$



θ

Conditional probability of each response option given θ

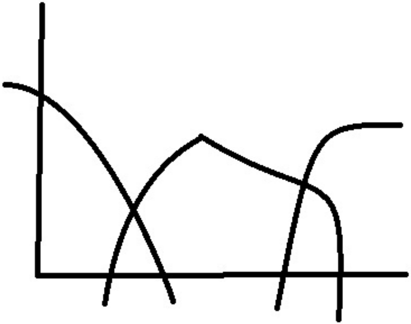
$$P(X = 1 | \Theta = \theta)$$



θ

Conditional probability of each response option given θ

$$P(X = 2|\Theta = \theta)$$



θ

Model for response vector $\bar{X} = (X_i)_{i \in I}$

Probability of observing $\bar{x} = (x_i)_{i \in I}$ is

$$P(\bar{X} = \bar{x} | \Theta = \theta) \overset{*}{=} \prod_{i \in I} P(X_i = x | \Theta = \theta) \quad (2)$$

$$= \frac{\exp(\theta \sum_{i \in I} X_i + \sum_{i \in I} \beta_{ix_i})}{\prod_{i \in I} K_i(\theta)} \quad (3)$$

Marginal probability

$$P(\bar{X} = \bar{x}) = \int P(\bar{X} = \bar{x} | \Theta = \theta) \phi(\theta) d\theta \quad (4)$$

$$= \int \frac{\exp(\theta \sum_{i \in I} X_i + \sum_{i \in I} \beta_{ix_i})}{\prod_{i \in I} K_i(\theta)} \varphi(\theta) d\theta \quad (5)$$

[*: assumption: 'local' independence]

Subjects $v = 1, \ldots, N$, T_v treatment indicator, A_v age

$$\theta_v = \delta_0 + \delta_T T_v [+\delta_A A_v] + \epsilon_v \quad (6)$$

$$L(\bar{\bar{\beta}}, \delta_0, \delta_T, \delta_A) = \prod_{v=1}^{N} \int P(\bar{X} = \bar{x} | \Theta = \delta_0 + \delta_T T_v + \epsilon_v) \varphi(\epsilon) d\theta \quad (7)$$

Implementation: PROC NLMIXED

Set of items split into two disjoint sets

$$I = I_1 \cup I_2, \tag{8}$$

with items in I_1 and I_2 measuring latent variables θ_1 and θ_2 respectively.

$$P(\overline{X} = \overline{x} | \bar{\Theta} = \bar{\theta}) \quad = \quad \frac{\exp\left(r_1\theta_1 + r_2\theta_2 + \sum_{i \in I} \eta_{ix_i}\right)}{K(\bar{\theta})} \tag{9}$$

where $r_d = \sum_{i \in I_d} x_i$, $d = 1, 2$ and

$$K(\bar{\theta}) = \left(\prod_{i \in I_1} K_i(\theta_1)\right)\left(\prod_{i \in I_2} K_i(\theta_2)\right), \quad \bar{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

Two-dimensional model, cont.

(i) θ_1 and θ_2 are distinct, but correlated latent variables,

(ii) θ_1 and θ_2 represent repeated measurements of the same latent variable

Test invariance: $(\bar{\beta}_i)_{i \in I_1}$ and $(\bar{\beta}_i)_{i \in I_2}$ identical.

Example of population models

- μ average shift in latent variable, latent correlation:

$$\bar{\theta} \sim N\left(\begin{bmatrix} 0 \\ \mu \end{bmatrix}, \Sigma\right) \quad (10)$$

- baseline means constrained to be equal across randomization groups (Liu et al, Statist. Med. vol. 28, 2509-2530, 2009):

$$\bar{\theta}_v = \begin{bmatrix} \delta_0 \\ \delta_0 \end{bmatrix} + \begin{bmatrix} 0 \\ \delta_T T_v \end{bmatrix} + \begin{bmatrix} \epsilon_{v1} \\ \epsilon_{v2} \end{bmatrix}, \quad \begin{bmatrix} \epsilon_{v1} \\ \epsilon_{v2} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma\right) \quad (11)$$

- Implementation: PROC NLMIXED

Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
    - Item Response Theory (IRT) models
    - Probit models
3. Example data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program:

    Vitality (VT) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)

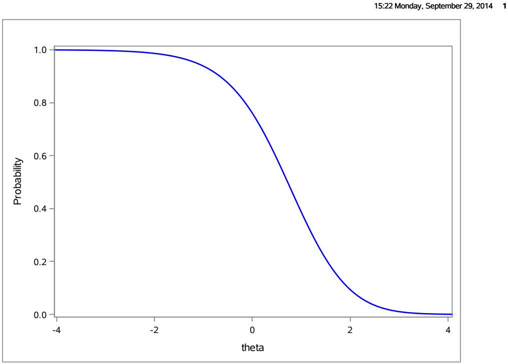
    Physical Functioning (PF) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)
4. Summary

PF sub scale from SF36

- 10-item scale, measures limitations
- Items endorsed on 3-point scale ('Yes, Limited A Lot', 'Yes, Limited A Little', 'No, Not Limited At All')
  - Vigorous Activities

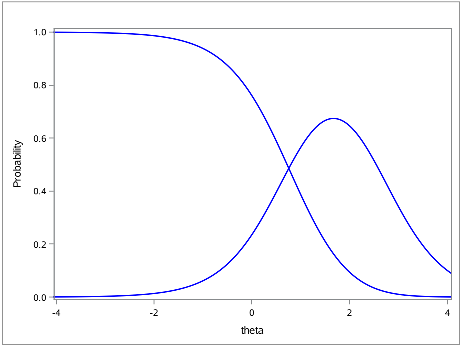
    ⋮
  - Walking one block
  - Bathing or dressing yourself

PF: Conditional probability given θ



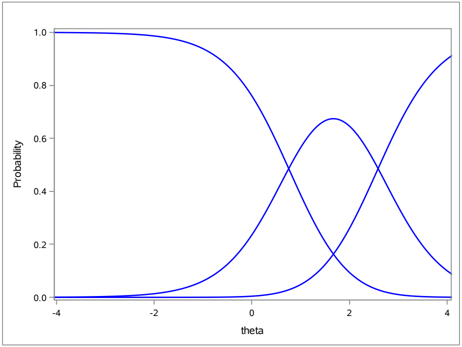
PF: Conditional probability given θ

15:22 Monday, September 29, 2014 1

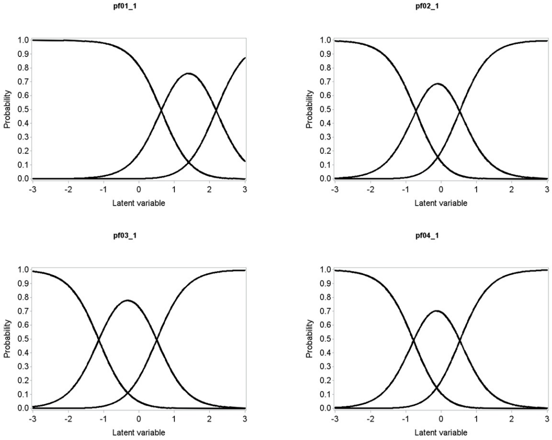


PF: Conditional probability given θ

15:22 Monday, September 29, 2014 1



Conditional probability given θ



VT sub scale from SF36

- Widely used measure of vitality
- 4-item scale, (two positively, two negatively-worded items)
  - Did you feel full of pep?
  - Did you have a lot of energy?
  - Did you feel worn out?
  - Did you feel tired?
- Items endorsed on 6-point scale ranging from 'None of the time' to 'All of the time'

PF and VT sub scales from SF36 - linear models

Data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program.

		Baseline			Twelve weeks			Change baseline to twelve weeks				Group comparison			
		N	Mean	SD	N	Mean	SD	N	D	95%	CI	D	95%	CI	p
SF36 Physical Functioning (PF)	A	34	61.9	27.5	32	82.0	21.8	32	19.1	10.0	28.1	10.4	-3.8	24.6	0.15
	B	36	66.7	24.2	30	72.7	26.3	30	8.6	-2.8	20.1				
SF36 Vitality (VT)	A	34	45.4	21.4	32	63.4	19.4	32	17.7	9.8	25.5	11.6	-0.6	23.8	0.06
	B	36	54.6	21.9	30	58.1	25.1	30	6.1	-3.7	15.9				

PF sub scale from SF36

	T-test			IRT		
	D	(95% CI)	p	$\hat{\delta}_T$	(95% CI)	p
PF	10.4	(-3.8 to 24.6)	0.15	0.56	(-0.11 to 1.22)	0.10

PF sub scale from SF36

	T-test			IRT		
	D	(95% CI)	p	$\hat{\delta}_T$	(95% CI)	p
PF	10.4	(-3.8 to 24.6)	0.15	0.56	(-0.11 to 1.22)	0.10

Power
N=70
N=100
N=100

PF sub scales from SF36 - linear models

	T-test			IRT		
	D	(95% CI)	p	$\hat{\delta}_T$	(95% CI)	p
PF	10.4	(-3.8 to 24.6)	0.15	0.56	(-0.11 to 1.22)	0.10

Power		
N=70	$\sim$ 35%	$\sim$ 42%
N=100		
N=100		

PF sub scales from SF36 - linear models

	T-test			IRT		
	D	(95% CI)	p	$\hat{\delta}_T$	(95% CI)	p
PF	10.4	(-3.8 to 24.6)	0.15	0.56	(-0.11 to 1.22)	0.10

Power	T-test	IRT
N=70	$\sim$ 35%	$\sim$ 42%
N=100	$\sim$ 42%	$\sim$ 49%
N=130		

PF sub scales from SF36 - linear models

	T-test			IRT		
	D	(95% CI)	p	$\hat{\delta}_T$	(95% CI)	p
PF	10.4	(-3.8 to 24.6)	0.15	0.56	(-0.11 to 1.22)	0.10

Power		
N=70	$\sim$ 35%	$\sim$ 42%
N=100	$\sim$ 42%	$\sim$ 49%
N=130	$\sim$ 53%	$\sim$ 64%

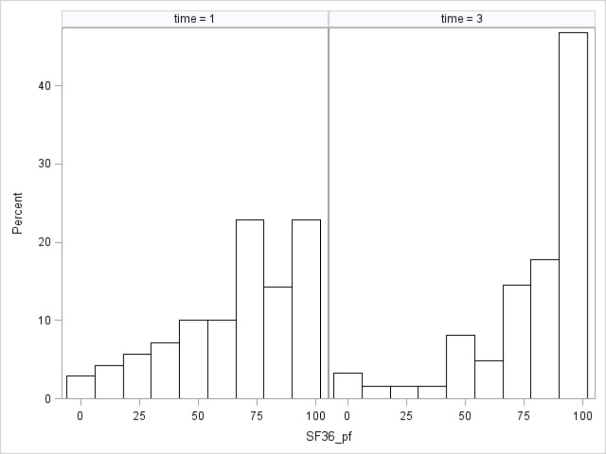
What about the VT sub scale?

	T-test			IRT		
	D	(95% CI)	p	$\hat{\delta}_T$	(95% CI)	p
PF	10.4	(-3.8 to 24.6)	0.15	0.56	(-0.11 to 1.22)	0.10
VT	11.6	(-0.6 to 23.8)	0.06			

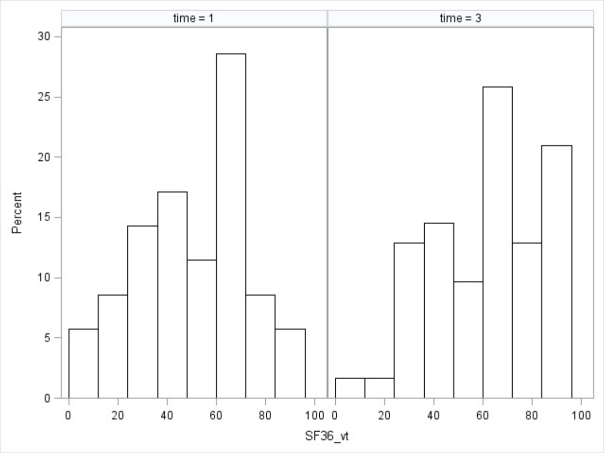
What about the VT sub scale?

	T-test			IRT		
	D	(95% CI)	p	$\hat{\delta}_T$	(95% CI)	p
PF	10.4	(-3.8 to 24.6)	0.15	0.56	(-0.11 to 1.22)	0.10
VT	11.6	(-0.6 to 23.8)	0.06	0.37	(-0.12 to 0.86)	0.14

PF sub scale from SF36



VT sub scale from SF36



Agenda

1. Indirect measurement
2. Ordinal regression models with random effects
    - Item Response Theory (IRT) models
    - Probit models
3. Example data from RCT to determine if patients with acute leukemia can benefit by a structured and supervised counseling and exercise program:

    Vitality (VT) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)

    Physical Functioning (PF) sub scale in the short form 36 (SF36) of the health status survey of the Medical Outcomes Study (MOS)
4. Summary

Summary

- Nonlinear random effects models yield increased power for skewed data
- Can specify model for the (unobserved) variable of interest
- Can quantify the uncertainty on change scores estimated for individual patients
- Implementation: PROC NLMIXED