Statistical analysis of discrete outcomes in longitudinal studies

Ivonne Solis-Trapala

NCRM Lancaster-Warwick-Stirling Node

ESRC Research Festival 2010

Discrete outcomes in longitudinal studies

The title explained:

- Discrete outcomes: Discrete outcomes having only integer values, for example:
  Number of heart attacks (0,1,2...),
  Failure (0) or success (1) in a psychological test item.

- Longitudinal studies: One or more variables for each of a number of subjects are measured a number of different time points.

Longitudinal data is not:

- Time series data: Single long series of measurements,
- Multivariate data: Single outcome of two or more different kinds of measurements on each subject;

BUT:

a large number of short time series

Example of time series data

Cardiovascular events during the FIFA world cup in 2006

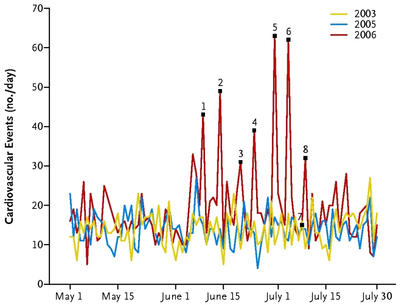


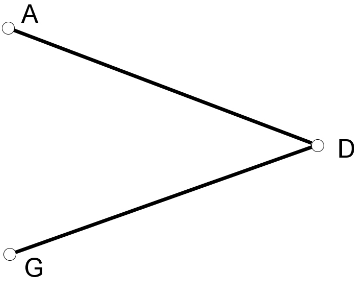
Figure 1. Daily Cardiovascular Events in the Study Population from May 1 to July 31 in 2003, 2005, and 2006.

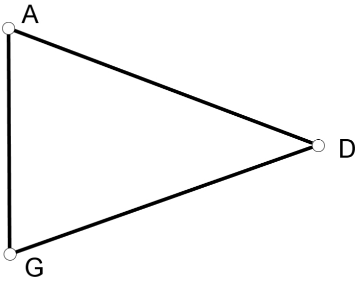
The FIFA World Cup 2006 in Germany started on June 9, 2006, and ended on July 9, 2006. The 2006 World Cup matches with German participation are indicated by numbers 1 through 7: match 1, Germany versus Costa Rica; match 2, Germany versus Poland; match 3, Germany versus Ecuador; match 4, Germany versus Sweden; match 5, Germany versus Argentina; match 6, Germany versus Italy; and match 7, Germany versus Portugal (for third-place standing). Match 8 was the final match, Italy versus France.

Source: Wilbert-Lampen *et al.* (2008) N. Engl. J. Med. **358** 5 475–483

Example of multivariate data
Fair admission process to a university

$A =$ student is admitted (yes/no)
$G =$ student's gender (female/male)
$D =$ department (Mathematics, Medicine, Engineering, Biology)





Fair: female admission rates similar to male admission rates at each department

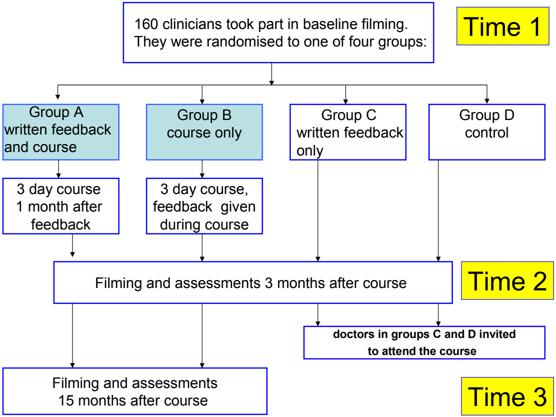
Unfair: otherwise

Example of longitudinal data
Improving communication skills of oncologists

"Of course I'm listening to your expression of spiritual suffering. Don't you see me making eye contact, striking an open posture, leaning towards you and nodding empathetically?"

Example of longitudinal data (cont.)

A randomised controlled trial

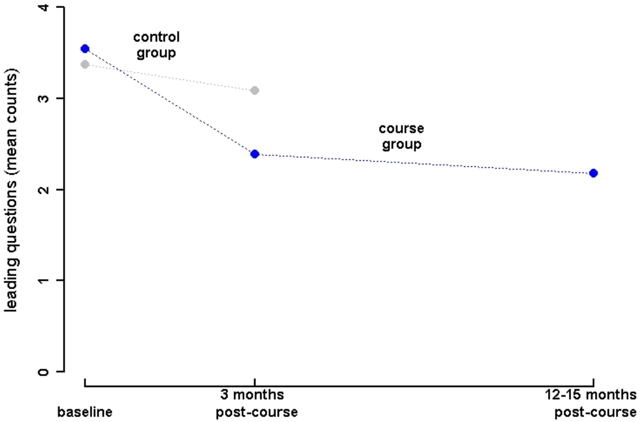


Reference: Fallowfield *et al.* (2002) The Lancet, **359**, 650–656

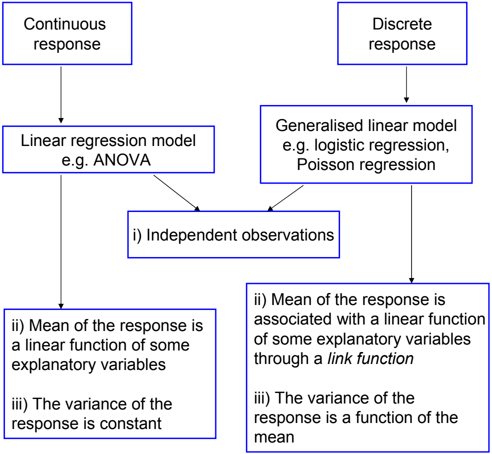
Example of longitudinal data (cont.)
The MIPS data

- MIPS = Medical Interaction Process
- DATA: COUNTS of primary outcomes, i.e. leading questions, expressions of empathy, focused questions
- Participants: 160 doctors
- 2 consultations filmed for each doctor at TIMES 1, 2, 3

Example of longitudinal data (cont.)
Longitudinal performance



Statistical aspects
Modelling independent discrete data



Why the model assumptions are important

Consider the following four data sets

x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Source: Anscombe, F.J. (1973) The American Statistician, **27**, 17–21.

Fitting a linear regression model

Same for four data sets

Dependent variable: y

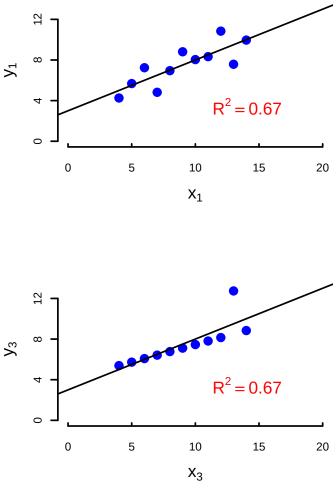
Coefficient	Estimate	Std. Error	t value	p-value
(Intercept)	3.0001	1.1247	2.67	0.0257
x	0.5001	0.1179	4.24	0.0022

Multiple R-Squared: 0.6665

F-statistic: 17.99 on 1 and 9 DF, p-value: 0.002170

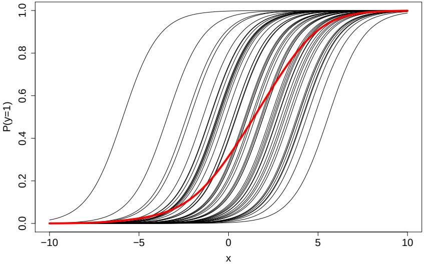
Equation of regression line: $y = 3 + 0.5x$

Assess the linearity assumption!



Target of inference: mean response (—) *vs.* mean response of an individual (—)

- Marginal models
- Random effects models
- Transition models



Example of marginal models

Generalised estimating equations (GEE)

- describe the relationship between *response variable* and *explanatory variables* with a *population average* regression model
- the approach provides consistent regression coefficient estimates even if the correlation structure is mis-specified

How is GEE implemented?

1. specify the mean regression
2. make a plausible guess of the covariance matrix
3. fit the model
4. use the residuals to adjust standard errors

Can you raed tihs?

I cdn'uolt blveiee taht I cluod aulaclty uesdnatnrd waht I was rdanieg: the phaonmneal pweor of the hmuan mnid. Aoccdrnig to a rsceearch taem at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoatnt tihng is taht the frist and lsat ltteer be in the rghit pclae. The rset can be a taotl mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe. Such a cdonition is arppoiatrely cllaed Typoglycemia

An example of misleading inferences when standard errors of regression parameters estimates are not adjusted:

MIPS data revisited

Robust conditional Poisson regression models comparing T_2 (3 month post-course) to T_1 (baseline) assessment

Behaviour	$\hat{\beta}_c$	naive SE	robust SE
Leading questions	-0.30	0.13	0.18
Focused questions	0.23	0.077	0.13
Focused and open questions	0.16	0.067	0.10
Expressions of empathy	0.41	0.14	0.25
Summarising of information	0.054	0.11	0.24
Interruptions	-0.15	0.30	0.41
Checking understanding	-0.18	0.15	0.22

Reference: Solis-Trapala & Farewell (2005) Biometrical Journal, **47**, 1–14

Final remarks

- Repeated observations on the same person generally produce correlated outcomes.
- Regression models are useful when the objective is to relate an outcome variable to other variables;
- however, traditional regression models assume that all outcomes are independent.
- Therefore, we may resource to generalizations of GLM, such as GEE, that account for correlated outcomes.