Today

- HW 1: due February 7, 2 pm.

  January 31, 4-5 pm reserved for questions re HW

- Aspects of Design — CD Chapter 2, Placebo/migraine study

- Generalized linear models: fitting, scale parameter, over-dispersion, examples

- In the News: neuroscience reading study,

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews, such as the Cochrane reviews
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews, such as the Cochrane reviews
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews, such as the Cochrane reviews
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews, such as the Cochrane reviews
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews, such as the Cochrane reviews
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

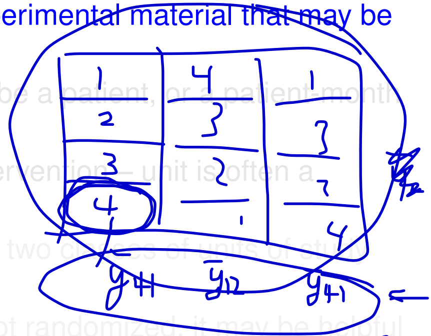
- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study



two or more sources of variation

info on trt 4

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis

usually 2 "tmts"
one applied to large unit
other to subunits

- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Table S5. Structure of the eight treatment sequences and assignment of subjects to treatment sequences

Treatment sequence[a]	Treatment conditions						Number of subjects		
	Attack 1	Attack 2	Attack 3	Attack 4	Attack 5	Attack 6	Recruited	Dropped out	Analyzed
5	M – M	M – P	P – M	P – P	U – M	U – P	10	1	9
7	P – M	P – P	M – M	M – P	U – M	U – P	9	2	7
1	U – M	U – P	M – M	M – P	P – M	P – P	9	2	7
3	U – M	U – P	P – M	P – P	M – M	M – P	10	0	10
2	U – P	U – M	M – P	M – M	P – P	P – M	9	2	7
4	U – P	U – M	P – P	P – M	M – P	M – M	9	2	7
6	M – P	M – M	P – P	P – M	U – P	U – M	10	1	9
8	P – P	P – M	M – P	M – M	U – P	U – M	10	0	10
						Totals	**76**	**10**	**66**

The 6 pill/label combinations are abbreviated as follows: the first letter (in *italic*) denotes the label (*M* for 'Maxalt', *P* for 'Placebo', *U* for the unspecified 'Maxalt *or* Placebo'); the second letter (in color) denotes the actual pill (M for maxalt, P for placebo). [a]Sequence numbers correspond to the order they were entered in the GLMM analyses (cf. table S6).

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

... avoidance of systematic error

CD §2.4

Table : Illustration: a comparison of T and C

Day	1	2	3	4	5	6	7	8
morning	T	T	T	T	T	T	T	T
afternoon	C	C	C	C	C	C	C	C

Day	1	2	3	4	5	6	7	8
morning	T	T	T	C	T	T	C	T
afternoon	C	C	C	T	C	C	T	C

Day	1	2	3	4	5	6	7	8
morning	T	T	C	T	C	T	C	C
afternoon	C	C	~~C~~ T	C	T	C	T	T

... avoidance of systematic error

- sometimes systematic error can be removed by modelling
-
$$y_{ij} = \mu + \tau x_{ij} + \delta z_j + \epsilon_{ij}, \quad j = 1, 2; i = 1, \dots n$$
-
$$x_{ij} = \begin{cases} +1 & \text{if } T \text{ used} \\ -1 & \text{if } C \text{ used} \end{cases}$$
$$\begin{aligned} z_1 &= 1 \quad \text{morning} \\ z_2 &= -1 \quad \text{afternoon} \end{aligned}$$
- find least squares estimate $\hat{\tau}$ of τ
- if T used pn times in morning, $\text{var}(\hat{\tau}) = \sigma^2/\{8p(1-p)n\}$
- minimized at $p = 1/2$ compare (b) and (c) on previous slide
- in (a) systematic error cannot be adjusted for;
  in (b) it can be adjusted for with some loss of precision;
  in (c) treatment comparison is unaffected by systematic differences between morning and afternoon

... avoidance of systematic error

- sometimes systematic error can be removed by modelling
- $$y_{ij} = \mu + \tau x_{ij} + \delta z_j + \epsilon_{ij}, \quad j = 1, 2; i = 1, \dots n$$
- $$x_{ij} = \begin{cases} +1 & \text{if } T \text{ used} \\ -1 & \text{if } C \text{ used} \end{cases}$$
  $$\begin{aligned} z_1 &= 1 \quad \text{morning} \\ z_2 &= -1 \quad \text{afternoon} \end{aligned}$$

$E\varepsilon_{\cdot j} = 0$

- find least squares estimate $\hat{\tau}$ of τ
- if T used pn times in morning, $\text{var}(\hat{\tau}) = \sigma^2 / \{8p(1-p)n\}$
- minimized at $p = 1/2$ compare (b) and (c) on previous slide
- in (a) systematic error cannot be adjusted for;
  in (b) it can be adjusted for with some loss of precision;
  in (c) treatment comparison is unaffected by systematic differences between morning and afternoon

... avoidance of systematic error

- sometimes systematic error can be removed by modelling
- $$y_{ij} = \mu + \tau x_{ij} + \delta z_j + \epsilon_{ij}, \quad j = 1, 2; i = 1, \ldots n$$
- $$x_{ij} = \begin{cases} +1 & \text{if } T \text{ used} \\ -1 & \text{if } C \text{ used} \end{cases}$$
  $$\begin{aligned} z_1 &= 1 \quad \text{morning} \\ z_2 &= -1 \quad \text{afternoon} \end{aligned}$$
- find least squares estimate $\hat{\tau}$ of τ
- if T used pn times in morning, $\text{var}(\hat{\tau}) = \sigma^2/\{8p(1-p)n\}$
- minimized at $p = 1/2$ compare (b) and (c) on previous slide
- in (a) systematic error cannot be adjusted for;
  in (b) it can be adjusted for with some loss of precision;
  in (c) treatment comparison is unaffected by systematic differences between morning and afternoon

... avoidance of systematic error

- sometimes systematic error can be removed by modelling
- $$y_{ij} = \mu + \tau x_{ij} + \delta z_j + \epsilon_{ij}, \quad j = 1, 2; i = 1, \dots n$$
- $$x_{ij} = \begin{cases} +1 & \text{if } T \text{ used} \\ -1 & \text{if } C \text{ used} \end{cases}$$
  $$\begin{aligned} z_1 &= 1 \quad \text{morning} \\ z_2 &= -1 \quad \text{afternoon} \end{aligned}$$
  $E \epsilon_{ij} = 0$
- find least squares estimate $\hat{\tau}$ of τ
- if T used pn times in morning, $\text{var}(\hat{\tau}) = \sigma^2/\{8p(1-p)n\}$
- minimized at $p = 1/2$ compare (b) and (c) on previous slide
- in (a) systematic error cannot be adjusted for;
  in (b) it can be adjusted for with some loss of precision;
  in (c) treatment comparison is unaffected by systematic differences between morning and afternoon

... avoidance of systematic error

- sometimes systematic error can be removed by modelling
- $$y_{ij} = \mu + \tau x_{ij} + \delta z_j + \epsilon_{ij}, \quad j = 1, 2; i = 1, \dots n$$
- $$x_{ij} = \begin{cases} +1 & \text{if } T \text{ used} \\ -1 & \text{if } C \text{ used} \end{cases}$$

  $\sigma^2 = E\varepsilon_{ij}^2$

  $$\begin{aligned} z_1 &= 1 \quad \text{morning} \\ z_2 &= -1 \quad \text{afternoon} \end{aligned}$$
- find least squares estimate $\hat{\tau}$ of τ (3rd)
- if T used pn times in morning, $\text{var}(\hat{\tau}) = \sigma^2/\{8p(1-p)n\}$
- minimized at $p = 1/2$ compare (b) and (c) on previous slide
- in (a) systematic error cannot be adjusted for;
  in (b) it can be adjusted for with some loss of precision;
  in (c) treatment comparison is unaffected by systematic differences between morning and afternoon

... avoidance of systematic error

- sometimes systematic error can be removed by modelling
- $$y_{ij} = \mu + \tau x_{ij} + \delta z_j + \epsilon_{ij}, \quad j = 1, 2; i = 1, \dots n$$
- $$x_{ij} = \begin{cases} +1 & \text{if } T \text{ used} \\ -1 & \text{if } C \text{ used} \end{cases}$$

  $$\begin{aligned} z_1 &= 1 \quad \text{morning} \\ z_2 &= -1 \quad \text{afternoon} \end{aligned}$$
- find least squares estimate $\hat{\tau}$ of τ
- if T used pn times in morning, $\text{var}(\hat{\tau}) = \sigma^2/\{8p(1-p)n\}$
- minimized at $p = 1/2$ 3rd compare (b) and (c) on previous slide
- in (a) systematic error cannot be adjusted for;
  in (b) it can be adjusted for with some loss of precision;
  in (c) treatment comparison is unaffected by systematic differences between morning and afternoon

... avoidance of systematic error

- sometimes systematic error can be removed by modelling
- $$y_{ij} = \mu + \tau x_{ij} + \delta z_j + \epsilon_{ij}, \quad j = 1, 2; i = 1, \dots n$$
- $$x_{ij} = \begin{cases} +1 & \text{if } T \text{ used} \\ -1 & \text{if } C \text{ used} \end{cases}$$
  $$\begin{aligned} z_1 &= 1 \quad \text{morning} \\ z_2 &= -1 \quad \text{afternoon} \end{aligned}$$
- find least squares estimate $\hat{\tau}$ of τ
- if T used pn times in morning, $\mathrm{var}(\hat{\tau}) = \sigma^2/\{8p(1-p)n\}$
- minimized at $p = 1/2$ compare (b) and (c) on previous slide
- in (a) systematic error cannot be adjusted for;
  in (b) it can be adjusted for with some loss of precision;
  in (c) treatment comparison is unaffected by systematic differences between morning and afternoon

Control and estimation of random error CD §2.5

- statistical analysis is particularly important in investigations in which haphazard variation plays an important role
- we can lessen the impact of haphazard variation by
  - use of artificially uniform material
  - arranging that the comparisons of main interest compare like with like
  - inclusion of background variables
  - replication
- these may impact generalizability, so depend on the context

Control and estimation of random error CD §2.5

- statistical analysis is particularly important in investigations in which haphazard variation plays an important role
- we can lessen the impact of haphazard variation by
  - use of artificially uniform material
  - arranging that the comparisons of main interest compare like with like
  - inclusion of background variables
  - replication
- these may impact generalizability, so depend on the context

Control and estimation of random error CD §2.5

- statistical analysis is particularly important in investigations in which haphazard variation plays an important role
- we can lessen the impact of haphazard variation by
  - use of artificially uniform material
  - arranging that the comparisons of main interest compare like with like
  - inclusion of background variables
  - replication
- these may impact generalizability, so depend on the context

Control and estimation of random error CD §2.5

- statistical analysis is particularly important in investigations in which haphazard variation plays an important role
- we can lessen the impact of haphazard variation by
  - use of artificially uniform material
  - arranging that the comparisons of main interest compare like with like
  - inclusion of background variables
  - replication
- these may impact generalizability, so depend on the context

Control and estimation of random error

CD §2.5

- statistical analysis is particularly important in investigations in which haphazard variation plays an important role
- we can lessen the impact of haphazard variation by
  - use of artificially uniform material
  - arranging that the comparisons of main interest compare like with like
  - inclusion of background variables
  - replication
- these may impact generalizability, so depend on the context

Control and estimation of random error

CD §2.5

- statistical analysis is particularly important in investigations in which haphazard variation plays an important role
- we can lessen the impact of haphazard variation by
  - use of artificially uniform material
  - arranging that the comparisons of main interest compare like with like
  - inclusion of background variables
  - replication
- these may impact generalizability, so depend on the context

Control and estimation of random error CD §2.5

- statistical analysis is particularly important in investigations in which haphazard variation plays an important role
- we can lessen the impact of haphazard variation by
  - use of artificially uniform material
  - arranging that the comparisons of main interest compare like with like
  - inclusion of background variables
  - replication
- these may impact generalizability, so depend on the context

Scale of effort

CD §2.6

- how big should my sample be?
- key observation: $\text{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma^2/m$
- set a bound on the standard error of the most important comparison, say c
- then want $2\sigma^2/m \approx c^2$
- i.e. $m \approx 2\sigma^2/c^2$
- c will be to some extent determined by the magnitude of differences of interest
- this requires fewer quantities to be set than usual power calculations

Scale of effort

CD §2.6

- how big should my sample be?
- key observation: $\text{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma^2/m$

- set a bound on the standard error of the most important comparison, say c
- then want $2\sigma^2/m \approx c^2$
- i.e. $m \approx 2\sigma^2/c^2$

- c will be to some extent determined by the magnitude of differences of interest
- this requires fewer quantities to be set than usual power calculations

Scale of effort

CD §2.6

- how big should my sample be?
- key observation: $\text{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma^2/m$

- set a bound on the standard error of the most important comparison, say c
- then want $2\sigma^2/m \approx c^2$
- i.e. $m \approx 2\sigma^2/c^2$

- c will be to some extent determined by the magnitude of differences of interest
- this requires fewer quantities to be set than usual power calculations

Scale of effort

CD §2.6

- how big should my sample be?
- key observation: $\text{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma^2/m$

- set a bound on the standard error of the most important comparison, say c
- then want $2\sigma^2/m \approx c^2$
- i.e. $m \approx 2\sigma^2/c^2$

- c will be to some extent determined by the magnitude of differences of interest
- this requires fewer quantities to be set than usual power calculations

Scale of effort

CD §2.6

- how big should my sample be?
- key observation: $\text{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma^2/m$

- set a bound on the standard error of the most important comparison, say c
- then want $2\sigma^2/m \approx c^2$
- i.e. $m \approx 2\sigma^2/c^2$

- c will be to some extent determined by the magnitude of differences of interest
- this requires fewer quantities to be set than usual power calculations

Scale of effort

CD §2.6

- how big should my sample be?
- key observation: $\text{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma^2/m$
- set a bound on the standard error of the most important comparison, say c
- then want $2\sigma^2/m \approx c^2$
- i.e. $m \approx 2\sigma^2/c^2$
- c will be to some extent determined by the magnitude of differences of interest
- this requires fewer quantities to be set than usual power calculations

Scale of effort

CD §2.6

- how big should my sample be?
- key observation: $\mathrm{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma^2/m$

- set a bound on the standard error of the most important comparison, say c
- then want $2\sigma^2/m \approx c^2$
- i.e. $m \approx 2\sigma^2/c^2$

- c will be to some extent determined by the magnitude of differences of interest
- this requires fewer quantities to be set than usual power calculations

Migraine study revisited

- 7 “conditions”, or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st ’treatment’ always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating *M* and *P* for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

Migraine study revisited

- 7 “conditions”, or treatments
- unit of analysis? "attack" (within pt.)
- within patients, each attack assigned one of the 7 treatments; 1st ’treatment’ always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating M and P for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

Migraine study revisited

- 7 “conditions”, or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st ’treatment’ always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating *M* and *P* for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

Migraine study revisited

- 7 "conditions", or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st 'treatment' always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating *M* and *P* for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

B				
1	2	3	4	1
2	1	2	3	4
3	4	3	1	2

4! arr.
choose 3
at random

Migraine study revisited

- 7 “conditions”, or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st ’treatment’ always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating *M* and *P* for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

1 C

2 C

⋮ ⋮

66 C

Migraine study revisited

- 7 “conditions”, or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st ’treatment’ always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating M and P for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

sequences

1
2
⋮
8

Migraine study revisited

- 7 “conditions”, or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st ’treatment’ always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating *M* and *P* for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

Migraine study revisited

- 7 “conditions”, or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st ’treatment’ always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating M and P for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

Migraine study revisited

- 7 "conditions", or treatments
- unit of analysis?
- within patients, each attack assigned one of the 7 treatments; 1st 'treatment' always C
- small subset of 6! choices used for each patient/block
- balanced on order, since attacks are sequential in time
- alternating *M* and *P* for for pill; repeat each envelope label twice
- several observations in each unit, corresponding to different patients
- model

$f(y_{ijt} \mid b_i) \sim N(\mu_{ijt}, \sigma^2)$

$f(y_{ijt}) = \int \cdots \, db_i$

$$\log \mu_{ijt} = \beta_1 + \text{cond}_j + \text{time}_t + \text{cond} \times \text{time}_{jt} + b_i$$

$$y_{ijt} = \mu_{ijt} + \epsilon_{ijt}$$

- `family = gaussian, link = log`

Table S5. Structure of the eight treatment sequences and assignment of subjects to treatment sequences

Treatment sequence[a]	Treatment conditions						Number of subjects		
	Attack 1	Attack 2	Attack 3	Attack 4	Attack 5	Attack 6	Recruited	Dropped out	Analyzed
5	*M* – M	*M* – P	*P* – M	*P* – P	*U* – M	*U* – P	10	1	9
7	*P* – M	*P* – P	*M* – M	*M* – P	*U* – M	*U* – P	9	2	7
1	*U* – M	*U* – P	*M* – M	*M* – P	*P* – M	*P* – P	9	2	7
3	*U* – M	*U* – P	*P* – M	*P* – P	*M* – M	*M* – P	10	0	10
2	*U* – P	*U* – M	*M* – P	*M* – M	*P* – P	*P* – M	9	2	7
4	*U* – P	*U* – M	*P* – P	*P* – M	*M* – P	*M* – M	9	2	7
6	*M* – P	*M* – M	*P* – P	*P* – M	*U* – P	*U* – M	10	1	9
8	*P* – P	*P* – M	*M* – P	*M* – M	*U* – P	*U* – M	10	0	10
						Totals	**76**	**10**	**66**

The 6 pill/label combinations are abbreviated as follows: the first letter (in *italic*) denotes the label (*M* for 'Maxalt', *P* for 'Placebo', *U* for the unspecified 'Maxalt *or* Placebo'); the second letter (in color) denotes the actual pill (M for maxalt, P for placebo). [a]Sequence numbers correspond to the order they were entered in the GLMM analyses (cf. table S6).

ijb:

$$y_{ijt} = \mu_{ijt} + \varepsilon_{ijt}$$

$E\varepsilon_{ijt} = 0$

$E\varepsilon_{ijt}^2 = \sigma^2$

$$\log y_{ijt} = \beta_0 + \text{cond}_i + \ldots$$

$$\log E(y_{ijt}) = \beta_0 + \text{cond}_i + \ldots$$

$$E \log X \neq \log EX$$

var^n to also be X

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$

- $E(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j

- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates

- $g(\cdot)$ is the link function; η_j is the linear predictor

- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$

- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$
- $E(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j;\mu_j,\phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j;\phi_j)\}$$
- $E(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$
- $E(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j;\mu_j,\phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j;\phi_j)\}$$
- $E(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$

  j j
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j;\mu_j,\phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j;\phi_j)\}$$
- $E(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi_j b''(\theta_j) = \phi_j V(\mu_j)$ $\qquad$ ϕ_j scale par.
- $V(\cdot)$ is the variance function

Examples

- Normal
- Binomial
- Poisson
- Gamma/Exponential
- Inverse Gaussian

$||| \; r_j/m_j$

$e^{\theta_j} = \frac{p_j}{1-p_j}$ $\quad E y_j = \mu_j = p_j$

$\boxed{\text{link}: \log \frac{p}{1-p} = \log \frac{\mu}{1-\mu}}$

$R_j \sim Bin(m_j, p_j)$ $\quad f(r_j) = \binom{m_j}{r_j} p_j^{r_j} (1-p_j)^{m_j - r_j}$

$= \exp\left[r_j \log\left(\frac{p_j}{1-p_j}\right) + m_j \log(1-p_j) + \log\binom{m_j}{r_j} \right]$

$= \exp\left[r_j x_j^T\beta + m_j \log(1+e^{x_j^T\beta}) + \log\binom{m_j}{r_j} \right]$

$= \exp\left[m_j \{ y_j x_j^T\beta + \log(1+e^{x_j^T\beta}) \} + \log\binom{m_j}{y_j m_j} \right]$

$$f(y_j;\mu_j,\phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j;\phi_j)\}, \quad E(y_j) = \mu_j, \quad \text{var}(y_j) = \phi V(\mu_j)$$

$\phi_j = \frac{1}{m_j}$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known

- Normal distribution, $\phi =$

- Binomial distribution $\phi_j =$

- Gamma distribution, $\phi =$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$

Inference

- $\ell(\beta; y) = \sum_{j=}^{n} \{ \frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j, \phi_j) \}$

- $b'(\theta_j) = \mu_j; \quad g(\mu_j) = \eta_j = x_j^{\mathrm{T}}\beta$

- $\ell(\beta; y) = \sum \ell_j\{\eta_j(\beta), y_j\}$, say

- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_k} = \sum \dfrac{\partial \ell_j}{\partial \eta_j}\dfrac{\partial \eta_j}{\partial \beta_k} = \sum \dfrac{\partial \ell_j}{\partial \eta_j} x_{jk}$

- $\dfrac{\partial \ell_j}{\partial \eta_j} = \dfrac{\partial \ell_j}{\partial \theta_j}\dfrac{\partial \theta_j}{\partial \eta_j} = \dfrac{y_j - \mu_j}{\phi_j g'(\mu_j) V(\mu_j)}$

- matrix notation:

$$\frac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \frac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j, \phi_j)\}$

- $b'(\theta_j) = \mu_j; \quad g(\mu_j) = \eta_j = x_j^{\mathrm{T}}\beta$

- $\ell(\beta; y) = \sum \ell_j\{\eta_j(\beta), y_j\}, \quad$ say

- $\dfrac{\partial\ell(\beta; y)}{\partial\beta_k} = \sum \dfrac{\partial\ell_j}{\partial\eta_j}\dfrac{\partial\eta_j}{\partial\beta_k} = \sum \dfrac{\partial\ell_j}{\partial\eta_j}x_{jk}$

- $\dfrac{\partial\ell_j}{\partial\eta_j} = \dfrac{\partial\ell_j}{\partial\theta_j}\dfrac{\partial\theta_j}{\partial\eta_j} = \dfrac{y_j - \mu_j}{\phi_j g'(\mu_j)V(\mu_j)}$

- matrix notation:

$$\frac{\partial\ell(\beta)}{\partial\beta} = X^{\mathrm{T}}u(\beta), \quad X = \frac{\partial\eta}{\partial\beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j, \phi_j)\}$

- $b'(\theta_j) = \mu_j; \quad g(\mu_j) = \eta_j = x_j^{\mathrm{T}}\beta$

- $\ell(\beta; y) = \sum \ell_j\{\eta_j(\beta), y_j\}, \quad$ say

- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_k} = \sum \dfrac{\partial \ell_j}{\partial \eta_j}\dfrac{\partial \eta_j}{\partial \beta_k} = \sum \dfrac{\partial \ell_j}{\partial \eta_j} x_{jk}$

- $\dfrac{\partial \ell_j}{\partial \eta_j} = \dfrac{\partial \ell_j}{\partial \theta_j}\dfrac{\partial \theta_j}{\partial \eta_j} = \dfrac{y_j - \mu_j}{\phi_j g'(\mu_j) V(\mu_j)}$

- matrix notation:

$$\frac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \frac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j, \phi_j)\}$
- $b'(\theta_j) = \mu_j; \quad g(\mu_j) = \eta_j = x_j^{\mathrm{T}}\beta$
- $\ell(\beta; y) = \sum \ell_j\{\eta_j(\beta), y_j\}, \quad$ say
- $$\frac{\partial \ell(\beta; y)}{\partial \beta_k} = \sum \frac{\partial \ell_j}{\partial \eta_j}\frac{\partial \eta_j}{\partial \beta_k} = \sum \frac{\partial \ell_j}{\partial \eta_j} x_{jk}$$
- $$\frac{\partial \ell_j}{\partial \eta_j} = \frac{\partial \ell_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \eta_j} = \frac{y_j - \mu_j}{\phi_j g'(\mu_j) V(\mu_j)}$$
- matrix notation:

$$\frac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \frac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j, \phi_j)\}$
- $b'(\theta_j) = \mu_j; \quad g(\mu_j) = \eta_j = x_j^{\mathrm{T}}\beta$
- $\ell(\beta; y) = \sum \ell_j\{\eta_j(\beta), y_j\}, \quad$ say
- $\frac{\partial \ell(\beta; y)}{\partial \beta_k} = \sum \frac{\partial \ell_j}{\partial \eta_j}\frac{\partial \eta_j}{\partial \beta_k} = \sum \frac{\partial \ell_j}{\partial \eta_j} x_{jk}$
- $\frac{\partial \ell_j}{\partial \eta_j} = \frac{\partial \ell_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \eta_j} = \frac{y_j - \mu_j}{\phi_j g'(\mu_j) V(\mu_j)}$
- matrix notation:

$$\frac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \frac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$$

$b'(\theta_j) = \underline{\mu_j}$

$b''(\theta_j) = \phi_j V(\mu_j)$

Inference

- $\ell(\beta; y) = \sum \{ \frac{y_j \theta_j - b(\theta_j)}{\phi_j} + c(y_j, \phi_j) \}$

- $b'(\theta_j) = \mu_j; \quad g(\mu_j) = \eta_j = x_j^{\mathrm{T}} \beta$

- $\ell(\beta; y) = \sum \ell_j \{ \eta_j(\beta), y_j \}, \quad$ say

- $$\frac{\partial \ell(\beta; y)}{\partial \beta_k} = \sum \frac{\partial \ell_j}{\partial \eta_j} \frac{\partial \eta_j}{\partial \beta_k} = \sum \frac{\partial \ell_j}{\partial \eta_j} x_{jk}$$

- $$\frac{\partial \ell_j}{\partial \eta_j} = \frac{\partial \ell_j}{\partial \theta_j} \frac{\partial \theta_j}{\partial \eta_j} = \boxed{\frac{y_j - \mu_j}{\phi_j g'(\mu_j) V(\mu_j)}}$$

- matrix notation:

$$\underset{p \times 1}{\frac{\partial \ell(\beta)}{\partial \beta}} = \underset{p \times n}{X^{\mathrm{T}}} \underset{n \times 1}{u(\beta)}, \quad X = \frac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad \underline{u = (u_1, \ldots, u_n)}, \quad u_j =$$

Maximum likelihood estimation

- $\dfrac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \dfrac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$
- linearization: $X^{\mathrm{T}} u(\hat\beta) = 0 \doteq X^{\mathrm{T}} u(\beta) + (\hat\beta - \beta) X^{\mathrm{T}} \dfrac{\partial u(\beta)}{\partial \beta^{\mathrm{T}}}$
- re-arrange: $\hat\beta = \beta + I(\beta)^{-1} X^{\mathrm{T}} u(\beta)$
- ntbc:
  $I(\beta) = X^{\mathrm{T}} W X, \quad W = \mathrm{diag}(w_j), \quad w_j = 1/\{g'(\mu_j)^2 \phi_j V(\mu_j)\}$

$$
\begin{aligned}
\hat\beta &= \beta + (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} u(\beta) = (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W X \beta + X^{\mathrm{T}} u(\beta)\} \\
&= (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W (X\beta + W^{-1} u(\beta)\} \\
&= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z
\end{aligned}
$$

Maximum likelihood estimation

- $\dfrac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \dfrac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$
- linearization: $X^{\mathrm{T}} u(\hat\beta) = 0 \doteq X^{\mathrm{T}} u(\beta) + (\hat\beta - \beta) X^{\mathrm{T}} \dfrac{\partial u(\beta)}{\partial \beta^{\mathrm{T}}}$

$$(\hat\beta - \beta)\cancel{X}^{T} \frac{\partial u}{\partial \beta} = -X^{T} u(\beta)$$

- re-arrange: $\hat\beta \doteq \beta + \{I(\beta)\}^{-1} X^{\mathrm{T}} u(\beta)$
- ntbc:

$I(\beta) = X^{\mathrm{T}} W X, \quad W = \mathrm{diag}(w_j), \quad w_j = 1/\{g'(\mu_j)^2 \phi_j V(\mu_j)\}$

$$(\hat\beta - \beta) = [-\ ]^{-1} X^{T} u(\beta)$$

$$\begin{aligned}
\hat\beta &= \beta + (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} u(\beta) = (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W X \beta + X^{\mathrm{T}} u(\beta)\} \\
&= (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W (X\beta + W^{-1} u(\beta))\} \\
&= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z
\end{aligned}$$

$$\hat\beta \doteq \beta + [-\ ]^{-1} X^{T} u(\beta)$$

Maximum likelihood estimation

- $\dfrac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \dfrac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_j =$
- linearization: $X^{\mathrm{T}} u(\hat\beta) = 0 \doteq X^{\mathrm{T}} u(\beta) + (\hat\beta - \beta) \boxed{X^{\mathrm{T}} \dfrac{\partial u(\beta)}{\partial \beta^{\mathrm{T}}}}$
- re-arrange: $\hat\beta = \beta + I(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

$-I(\beta)$

- ntbc:
  $I(\beta) = X^{\mathrm{T}} W X, \quad W = \mathrm{diag}(w_j), \quad w_j = 1/\{g'(\mu_j)^2 \phi_j V(\mu_j)\}$

$$\begin{aligned}
\hat\beta &= \beta + (X^{\mathrm{T}}WX)^{-1} X^{\mathrm{T}} u(\beta) = (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}WX\beta + X^{\mathrm{T}}u(\beta)\} \\
&= (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}W(X\beta + W^{-1}u(\beta)\} \\
&= (X^{\mathrm{T}}WX)^{-1} X^{\mathrm{T}} W z
\end{aligned}$$

Maximum likelihood estimation

- $\dfrac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}} u(\beta), \quad X = \dfrac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \dots, u_n), \quad u_j =$

- linearization: $X^{\mathrm{T}} u(\hat\beta) = 0 \doteq X^{\mathrm{T}} u(\beta) + (\hat\beta - \beta) X^{\mathrm{T}} \dfrac{\partial u(\beta)}{\partial \beta^{\mathrm{T}}}$

- re-arrange: $\hat\beta = \beta + I(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

- ntbc:
  $I(\beta) = X^{\mathrm{T}} W X, \quad W = \mathrm{diag}(w_j), \quad w_j = 1/\{g'(\mu_j)^2 \phi_j V(\mu_j)\}$

st. pt $\quad \hat\beta = (X^T \hat W X)^{-1} X^T \hat W \hat z$

$$\begin{aligned}
\hat\beta &= \beta + (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} u(\beta) = (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W X \beta + X^{\mathrm{T}} u(\beta)\} \\
&= (X^{\mathrm{T}} W X)^{-1} \{\underline{X^{\mathrm{T}} W} \underbrace{(X\beta + W^{-1} u(\beta))}_{z}\} \\
&= \underline{(X^{\mathrm{T}} W X)}^{-1} \underline{X^{\mathrm{T}} W} z
\end{aligned}$$

derived response

... maximum likelihood estimation

- $$\begin{aligned}\hat{\beta} &= \beta + (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}u(\beta) = (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}WX\beta + X^{\mathrm{T}}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}W(X\beta + W^{-1}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz\end{aligned}$$
- does not involve ϕ_j
- if unknown (e.g. normal distribution or gamma distribution), must be estimated
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p}\sum_{j=1}^{n}\frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

... maximum likelihood estimation

- $$\begin{aligned}\hat{\beta} &= \beta + (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}u(\beta) = (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}WX\beta + X^{\mathrm{T}}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}W(X\beta + W^{-1}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz\end{aligned}$$
- does not involve ϕ_j
- if unknown (e.g. normal distribution or gamma distribution), must be estimated
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p}\sum_{j=1}^{n}\frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

... maximum likelihood estimation

$E Y_j = \mu_j$

- $$\begin{aligned}\hat\beta &= \beta + (X^{\mathrm T}WX)^{-1}X^{\mathrm T}u(\beta) = (X^{\mathrm T}WX)^{-1}\{X^{\mathrm T}WX\beta + X^{\mathrm T}u(\beta)\}\\ &= (X^{\mathrm T}WX)^{-1}\{X^{\mathrm T}W(X\beta + W^{-1}u(\beta)\}\\ &= (X^{\mathrm T}WX)^{-1}X^{\mathrm T}Wz\end{aligned}$$

$\operatorname{var} Y_j = \phi_j V(\mu_j)$

often known or fn of j

- does not involve ϕ_j

$w_j = \dfrac{1}{g'(\mu_j)^2 \phi_j V(\mu_j)}$

$u_j = (y_j - \mu_j)/\phi_j g'(\mu_j) V(\mu_j)$

- if unknown (e.g. normal distribution or gamma distribution), must be estimated

$z = X\beta + W^{-1}u$

$z_j = x_j^{\mathrm T}\beta + w_j^{-1}u_j$

- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)

- $$\hat\phi = \frac{1}{n-p}\sum_{j=1}^{n}\frac{(y_j - \hat\mu_j)^2}{V(\hat\mu_j)}$$

N $\phi_j = \sigma^2$

Pois $\phi_j = 1$ $\quad E Y_j = \operatorname{var} Y_j$

Bin $\phi_j = \dfrac{1}{m_j}$

Gamma $\phi_j = \dfrac{1}{\nu}$ shape

... maximum likelihood estimation

-
$$\begin{aligned}\hat{\beta} &= \beta + (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}u(\beta) = (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}WX\beta + X^{\mathrm{T}}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}W(X\beta + W^{-1}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz\end{aligned}$$

- does not involve ϕ_j
- if unknown (e.g. normal distribution or gamma distribution), must be estimated
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
-
$$\hat{\phi} = \frac{1}{n-p}\sum_{j=1}^{n}\frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

... maximum likelihood estimation

- $$\begin{aligned}\hat{\beta} &= \beta + (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}u(\beta) = (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}WX\beta + X^{\mathrm{T}}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}W(X\beta + W^{-1}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz\end{aligned}$$
- does not involve ϕ_j
- if unknown (e.g. normal distribution or gamma distribution), must be estimated
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p}\sum_{j=1}^{n}\frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

Chimp data

SM, Ex 10.16

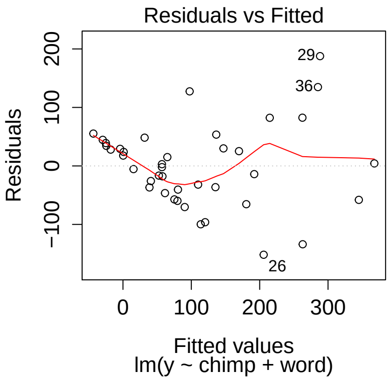
Table 10.5 Times in minutes taken by four chimpanzees to learn ten words (Brown and Hollander, 1977, p. 257).

	Word									
Chimpanzee	1	2	3	4	5	6	7	8	9	10
1	178	60	177	36	225	345	40	2	287	14
2	78	14	80	15	10	115	10	12	129	80
3	99	18	20	25	15	54	25	10	476	55
4	297	20	195	18	24	420	40	15	372	190

- "when a linear model is fitted, the *F*-statistic for non-additivity (8.27) strongly indicates and change of scale" (p.485,6); eq. (8.27) is on p.391
- linear model: $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$
- non-additivity: $y_{ij} = \mu + \alpha_i + \beta_j + \delta(\alpha_i\beta_j) + \epsilon_{ij}$
- special type of non-additivity with just 1 parameter to estimate δ

```
chimp.lm = lm(y ~ chimp + word, data = chimps)
anova(update(chimp.lm, . ~ . + I(chimp.lm$fitted.values*chimp.lm$fitted.values)))
```

... chimp data


Residuals vs Fitted
Residuals
200
100
0
−100
0
100
200
300
29
36
26
Fitted values
lm(y ~ chimp + word)

... chimp data

- change to a model more suitable for a response that measures time
- suggestion: Gamma model with mean $\mu_{cw} = \exp(\alpha_c + \gamma_w)$
- $$f(y_{cw}; \mu_{cw}, \nu) = \frac{1}{\Gamma(\nu)} y_{cw}^{\nu-1} \left(\frac{\nu}{\mu_{cw}}\right)^{\nu} \exp(-\nu y_{cw}/\mu_{cw})$$
- $$E(y_{cw}) = \mu_{cw}; \qquad \text{var}(y_{cw}) = \mu_{cw}^2/\nu$$
- linear predictor
$$\eta_{cw} = \alpha_c + \gamma_w$$
- link function
$$\eta = \log(\mu); \qquad \mu = \exp(\eta)$$

... chimp data

486 *10 · Nonlinear Regression Models*

Term	df	Deviance reduction	Term	df	Deviance reduction
Chimp (unadj. for Word)	3	6.95	Chimp (adj. for Word)	3	6.22
Word (adj. for Chimp)	9	38.46	Word (unadj. for Chimp)	9	39.19

Table 10.6 Analysis of deviance for models fitted to chimpanzee data.

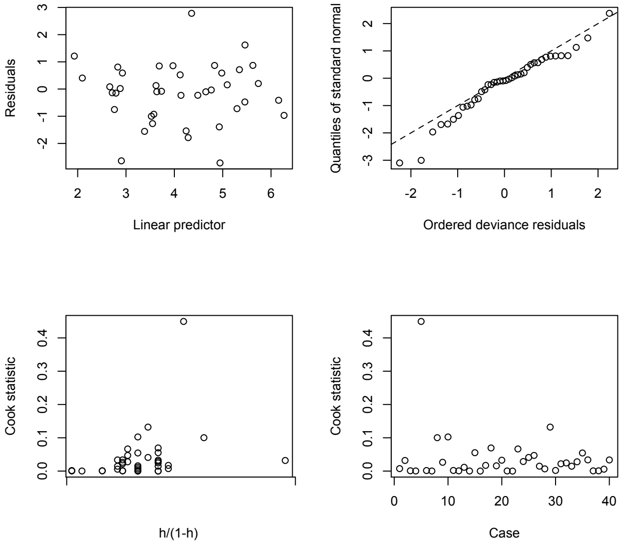
```
chimp.glm = glm(y ~ chimp + word,  family = Gamma(link = "log"), data = chimps)
> anova(chimp.glm)
Analysis of Deviance Table

Model: Gamma, link: log

Response: y

Terms added sequentially (first to last)


      Df Deviance Resid. Df Resid. Dev
NULL                     39     60.378
chimp  3    6.948        36     53.430
word   9   38.459        27     14.972
> summary(fit7)
(Dispersion parameter for Gamma family taken to be 0.4336663)

    Null deviance: 60.378  on 39  degrees of freedom
Residual deviance: 14.972  on 27  degrees of freedom
```

... chimp data

- "the signficance of the deviance reductions ... is gauged by F-tests" (p.486)
- see Eq (10.2), but note a few lines above "for now we suppress ϕ"
- see Example 10.3: $D_B - D_A = \phi^{-1} \sum\{...\} \dot{\sim} \chi^2_{p-q}$
- in this example we are estimating ϕ not needed for binary data
- p.483, 2nd paragraph: "when ϕ is unknown, the scaled deviance is replaced by the deviance"
- net result: deviance reduction for `chimp`, adjusted for `word` is 6.22 on 3 d.f.
- this is scaled by the estimate of ϕ, using (10.20), which is 0.4336 from `R` code; 0.432 in text
- refer $(6.22/3)/0.433$ to $F_{3,27}$ distribution; p-value is `pf(4.788,3,27,lower.tail=F) # 0.0084`

... chimp data

```
plot.glm.diag(chimps.glm)
```



... chimp data

- the canonical link is $\eta_{cw} = 1/\mu_{cw}$
- interpretation as the speed at which a word is learned
- non-additivity test for this link has p-value 0.11
- how to compare inverse link to log link?

Example 10.29

City	Rain	r/m	City	Rain	r/m	City	Rain	r/m	City	Rain	r/m
1	1735	2/4	11	2050	7/24	21	1756	2/12	31	1780	8/13
2	1936	3/10	12	1830	0/1	22	1650	0/1	32	1900	3/10
3	2000	1/5	13	1650	15/30	23	2250	8/11	33	1976	1/6
4	1973	3/10	14	2200	4/22	24	1796	41/77	34	2292	23/37
5	1750	2/2	15	2000	0/1	25	1890	24/51			
6	1800	3/5	16	1770	6/11	26	1871	7/16			
7	1750	2/8	17	1920	0/1	27	2063	46/82			
8	2077	7/19	18	1770	33/54	28	2100	9/13			
9	1920	3/6	19	2240	4/9	29	1918	23/43			
10	1800	8/10	20	1620	5/18	30	1834	53/75			

Table 10.19 Toxoplamosis data: rainfall (mm) and the numbers of people testing positive for toxoplasmosis, r, our of m people tested, for 34 cities in El Salvador (Efron, 1986).

Terms	df	Deviance
Constant	33	74.21
Linear	32	74.09
Quadratic	31	74.09
Cubic	30	62.63

Table 10.20 Analysis of deviance for polynomial logistic models fitted to the toxoplasmosis data.

- incidence of toxoplasmosis as a function of rainfall
- residual deviances approximately twice the degrees of freedom

Example 10.29

City	Rain	r/m	City	Rain	r/m	City	Rain	r/m	City	Rain	r/m
1	1735	2/4	11	2050	7/24	21	1756	2/12	31	1780	8/13
2	1936	3/10	12	1830	0/1	22	1650	0/1	32	1900	3/10
3	2000	1/5	13	1650	15/30	23	2250	8/11	33	1976	1/6
4	1973	3/10	14	2200	4/22	24	1796	41/77	34	2292	23/37
5	1750	2/2	15	2000	0/1	25	1890	24/51			
6	1800	3/5	16	1770	6/11	26	1871	7/16			
7	1750	2/8	17	1920	0/1	27	2063	46/82			
8	2077	7/19	18	1770	33/54	28	2100	9/13			
9	1920	3/6	19	2240	4/9	29	1918	23/43			
10	1800	8/10	20	1620	5/18	30	1834	53/75			

Table 10.19 Toxoplamosis data: rainfall (mm) and the numbers of people testing positive for toxoplasmosis, r, our of m people tested, for 34 cities in El Salvador (Efron, 1986).

Terms	df	Deviance
Constant	33	74.21
Linear	32	74.09
Quadratic	31	74.09
Cubic	30	62.63

Table 10.20 Analysis of deviance for polynomial logistic models fitted to the toxoplasmosis data.

- incidence of toxoplasmosis as a function of rainfall
- residual deviances approximately twice the degrees of freedom

... example 10.29

```
> data(toxo)
  rain  m  r
1 1620 18  5
2 1650 30 15
3 1650  1  0
4 1735  4  2
> toxo.glm0 = glm(cbind(r,m-r) ~ rain + I(rain^2) + I(rain^3), data = toxo,
family = binomial)

> anova(toxo.glm0)
 ...
          Df Deviance Resid. Df Resid. Dev
NULL                         33     74.212
rain       1   0.1244        32     74.087
I(rain^2)  1   0.0000        31     74.087
I(rain^3)  1  11.4529        30     62.635
> toxo.glm1 = glm(cbind(r,m-r) ~ poly(rain,3), data = toxo, family = binomial)

> summary(toxo.glm1)
...
Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              0.02427    0.07693   0.315 0.752401
poly(rain, degree = 3)1 -0.08606    0.45870  -0.188 0.851172
poly(rain, degree = 3)2 -0.19269    0.46739  -0.412 0.680141
poly(rain, degree = 3)3  1.37875    0.41150   3.351 0.000806 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 74.212  on 33  degrees of freedom
Residual deviance: 62.635  on 30  degrees of freedom
AIC: 161.33
```

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \dots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
-

$$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$

- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \dots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
-

$$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$

- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0, 1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0;$ otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if } \ldots$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$\log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0, 1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0;$ otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if} \ldots$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \dots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \dots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\mathrm{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\mathrm{var}(\hat\beta_{1Z}) = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\mathrm{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \mathrm{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\mathrm{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\mathrm{var}(\hat\beta_{1Y}) = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = 1/\sum(x_i - \bar{x})^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat{\beta}_Z$ is least squares estimator from original data
- $\text{cov}(\hat{\beta}_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Z}) = 1/\sum(x_i - \bar{x})^2$

- $\hat{\beta}_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat{\beta}_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat{\beta}_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Y}) = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = 1/\sum(x_i - \bar{x})^2$
- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\mathrm{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\mathrm{var}(\hat\beta_{1Z}) = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\mathrm{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \mathrm{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\mathrm{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\mathrm{var}(\hat\beta_{1Y}) = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat{\beta}_Z$ is least squares estimator from original data
- $\text{cov}(\hat{\beta}_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Z}) = 1/\sum(x_i - \bar{x})^2$

- $\hat{\beta}_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat{\beta}_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat{\beta}_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Y}) = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat{\beta}_Z$ is least squares estimator from original data
- $\text{cov}(\hat{\beta}_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Z}) = 1/\sum(x_i - \bar{x})^2$

- $\hat{\beta}_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat{\beta}_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat{\beta}_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Y}) = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- Figure 10.6 (right) plots $\beta_1/\sqrt{\sum(x_j - \bar{x})^2}$ on the x-axis, and $\beta_1/\sqrt{\qquad}$ on the y-axis
- trying to compare v_Z and v_Y, as well as indicate behaviour of $\beta_{1Y}/\sqrt{v_Y}$ as $\beta_1 \to \infty$

10.4 · Proportion Data 489

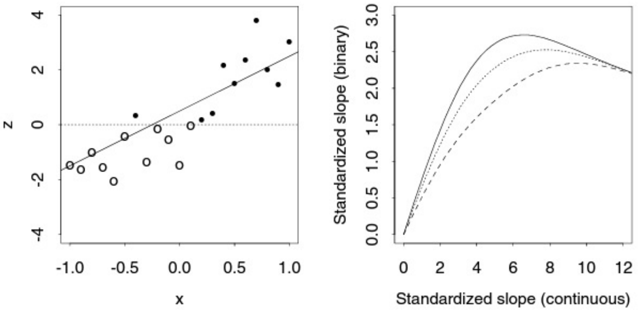


Figure 10.6 Efficiency loss due to reducing continuous variables to binary ones. Left panel: simulated data. Blobs above the dotted line are counted as successes, with zeros below it as failures; the solid line is $0.5 + 2x$. Right panel: Comparison of asymptotic t statistics when continuous data are dichotomized, for normal error distribution, when $\beta_0 = 0.5, 1, 1.5$ (solid, dots, dashes).