Bayesian workshop

- welcome
- 9:00 – ~5:00
- lunch – 1 hour
- a.m. & p.m. breaks
- bathrooms
- security

THE UNIVERSITY OF
MELBOURNE

Bayesian workshop road-map

- https://github.com/goldingn/bayes_ecology_course

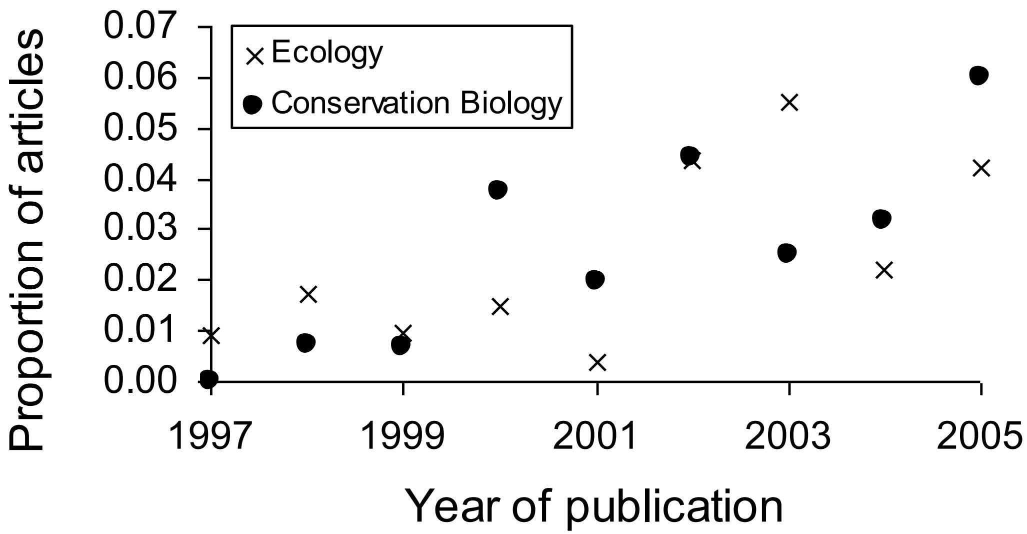
- introduction
- dive into an analysis of a mean
- regression
- logistic and Poisson regression
- ANOVA
- detectability

What we could discuss

- random effects
- posterior model probabilities
- Bayesian model averaging
- mark-recapture
- survival/failure analysis
- Bayesian networks
- proportions
- selecting models
- your own data

Statistics in Ecology

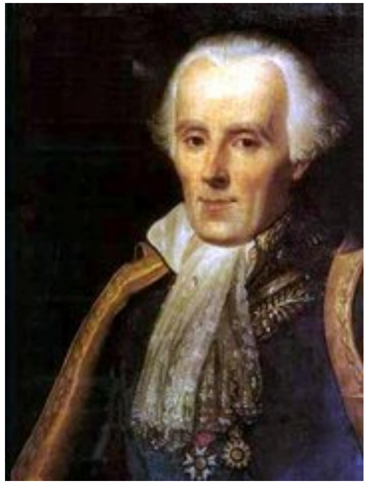
- 80-90% of papers use null hypothesis significance testing (NHST)
- up to ~5% of papers mention Bayesian methods



Alternatives to NHST

- frequentist methods
  - use confidence intervals to focus on the size of effects
  - information theoretic methods
- Bayesian methods

Bayesian inference



Steps in Bayesian inference:

1. Model formulation

2. Prior distributions for model parameters

   - represent prior (before data) belief about parameters

3. Observe data

4. Apply Bayes' rule => Obtain posterior distributions for parameters

   - represents posterior (after data) belief about parameters

Bayes' Theorem

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

$$\Pr(H_1|D) = \frac{\Pr(D|H_1) \times \text{Prior}(H_1)}{\sum \Pr(D|H_i) \times \text{Prior}(H_i)}$$

Constant of proportionality

Difficulties of Bayesian methods

- determining constant of proportionality
  - solved (MCMC, WinBUGS)
- determining priors
  - not resolved
  - need to be as careful with priors as with data
  - “This is what I believe, because of these lines of evidence…”

Benefits of Bayesian methods

- propagate uncertainty into predictions easily
- update belief in hypotheses or parameters coherently
- make results numerically identical to MLE by using uninformative priors
- handle small sample sizes without approximation
- analyse arbitrarily complex models
- interpretation and intuition

Value of Prior Information
Survival of European Dippers

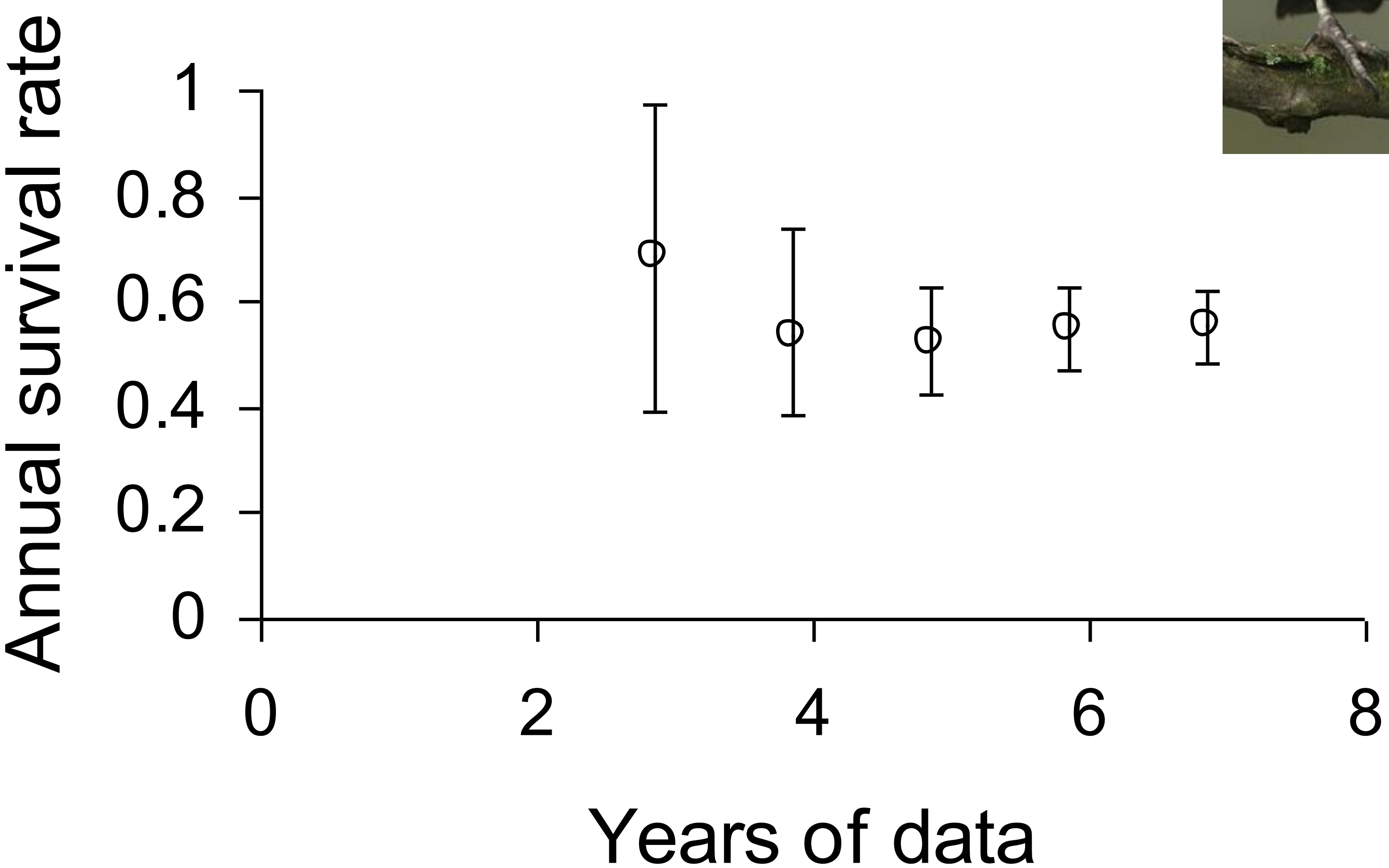
- 7 years of mark-recapture data

(Marzolin 1988)

- 130 female birds marked

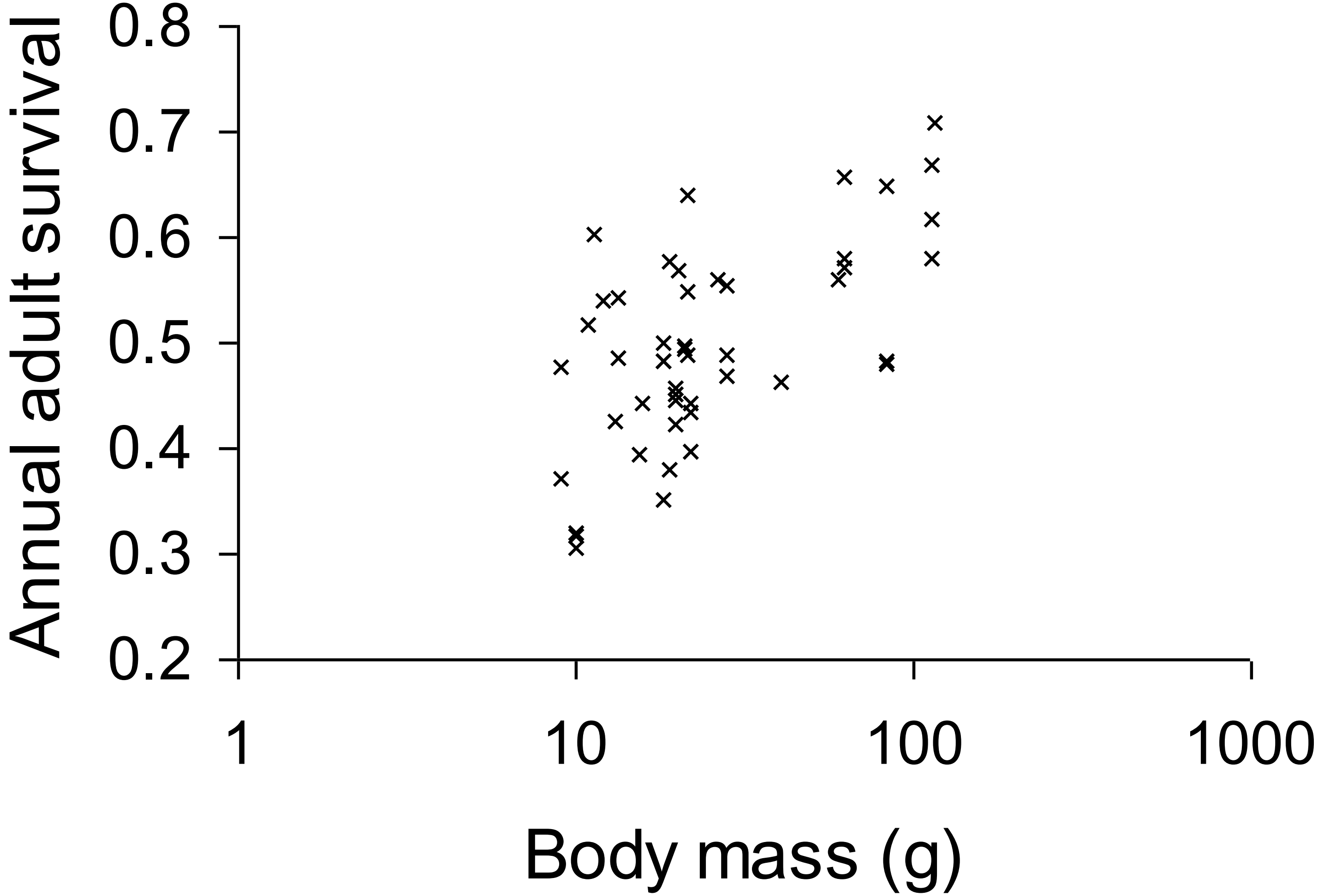
- Annual survival of females
  - 0.55 (95% CI: 0.48 – 0.62)

No prior information

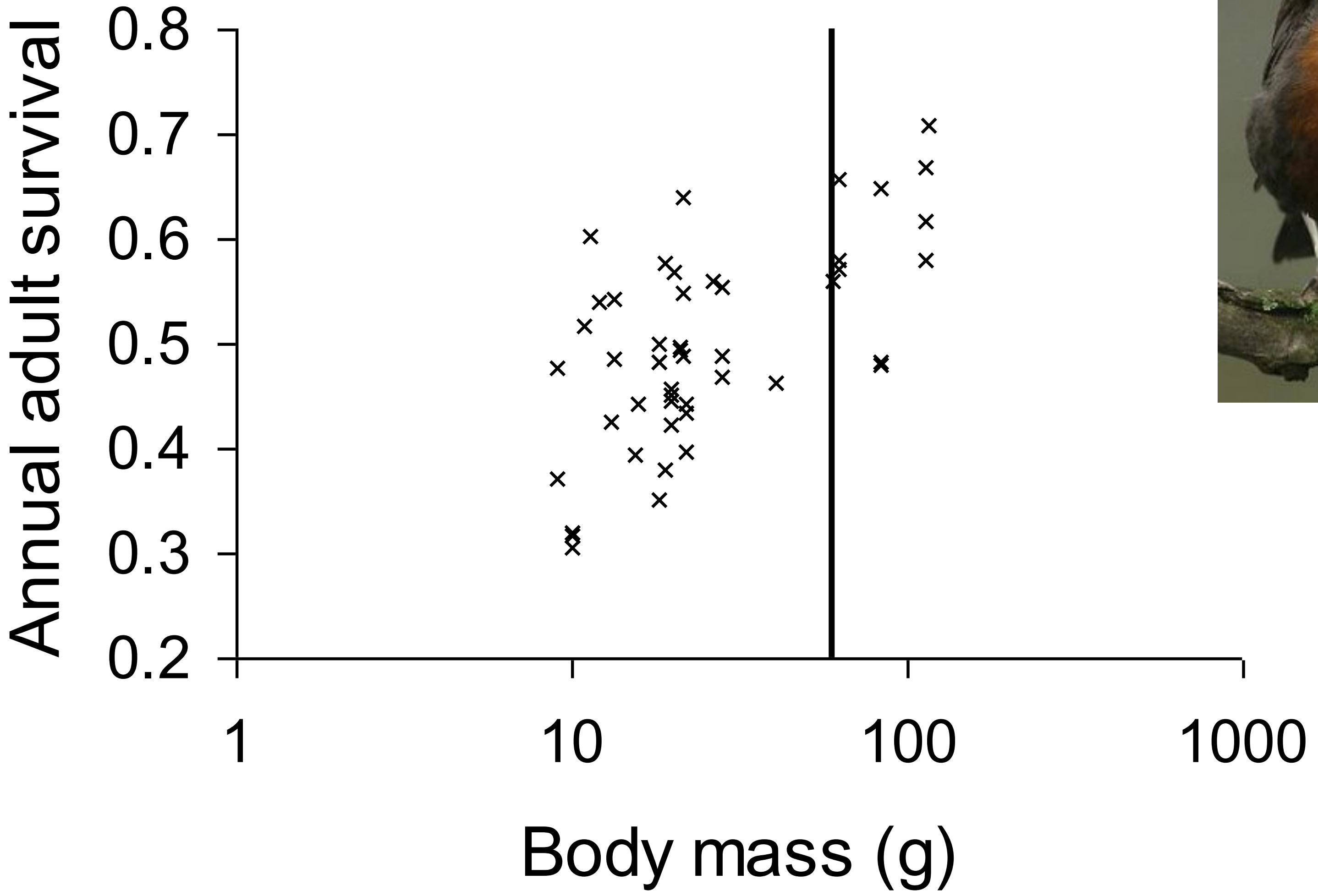


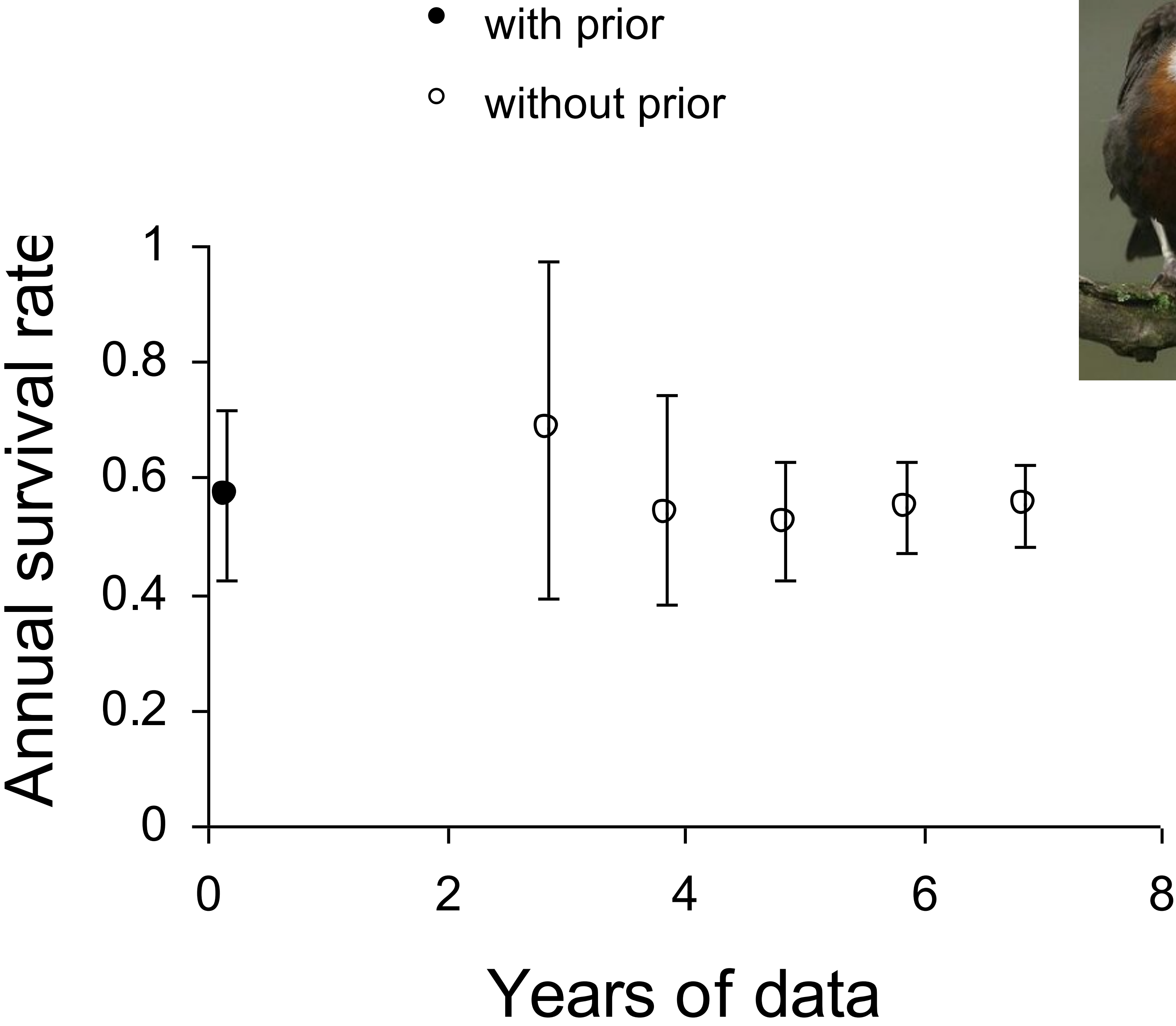
Survival of European Passerines

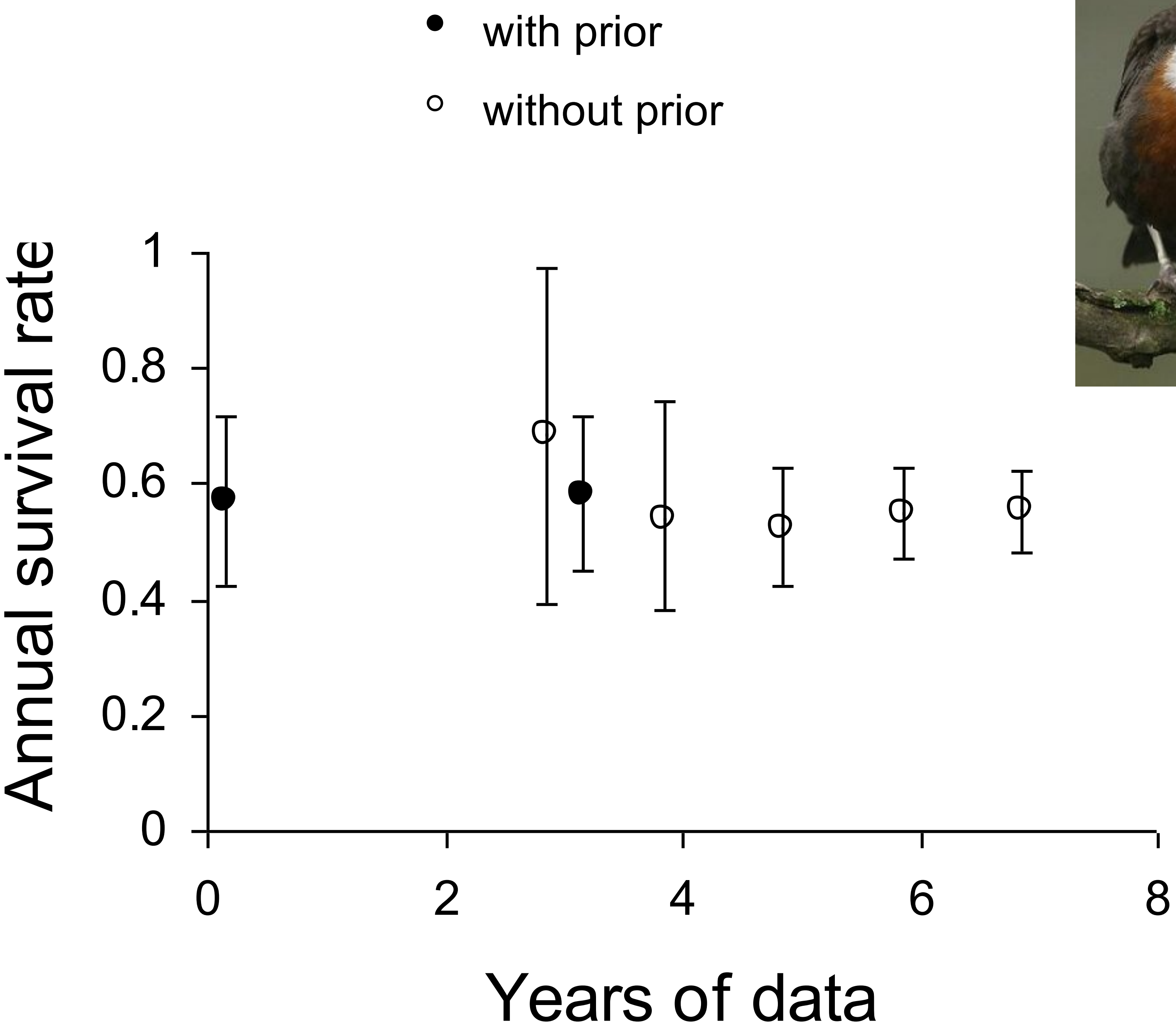
Data from Johnston et al. 1993. American Naturalist

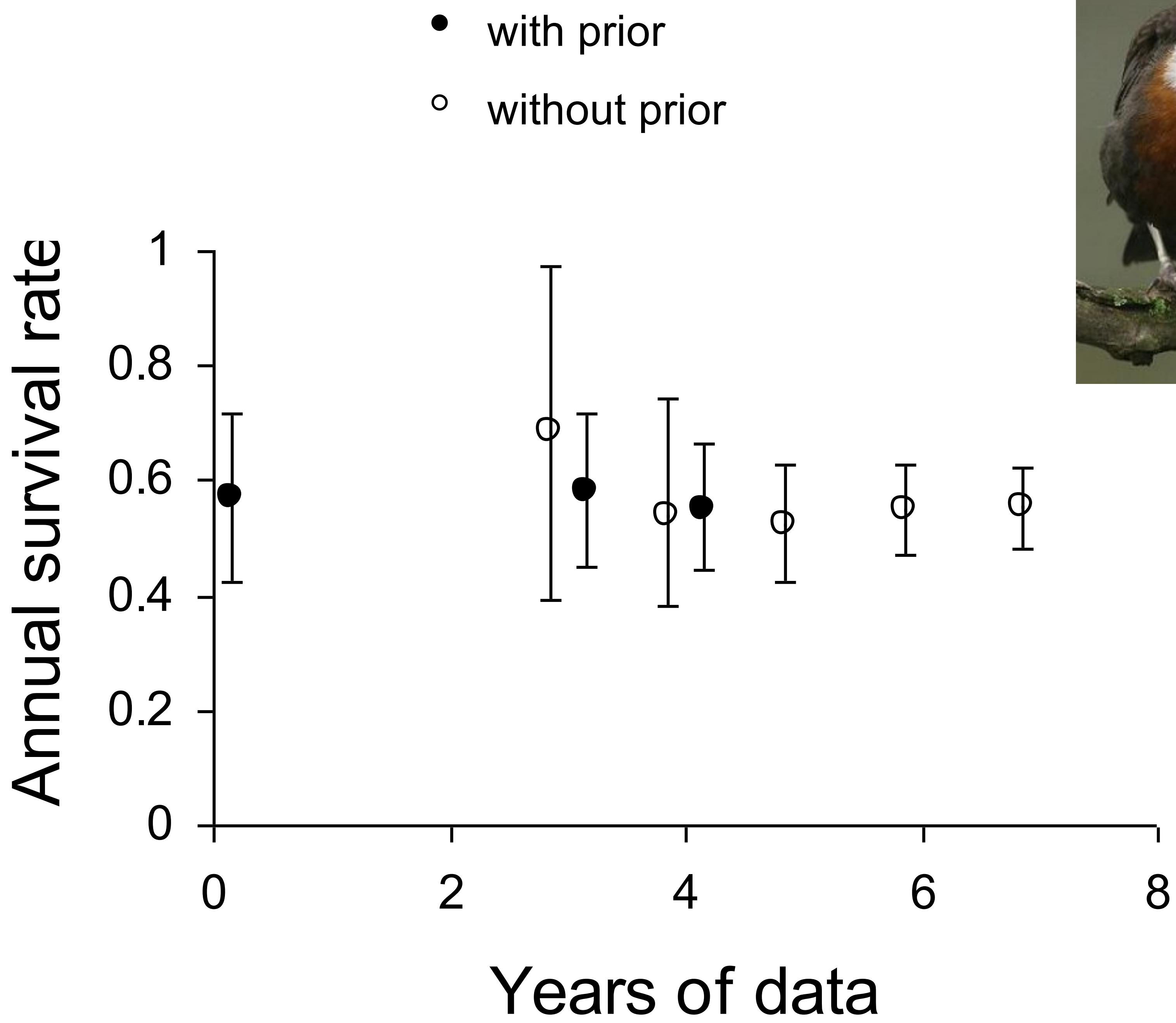


Survival of European Passerines

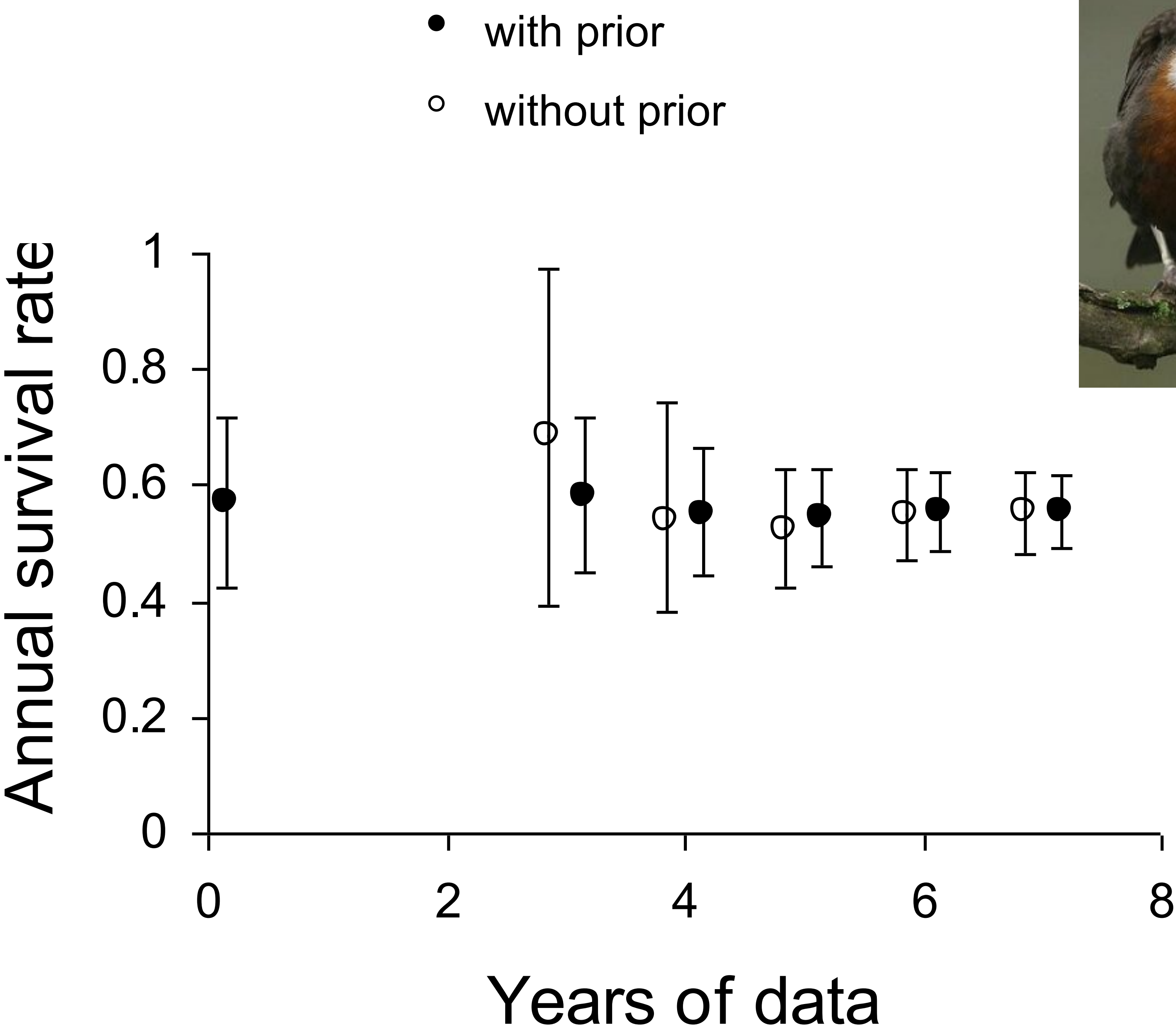



with prior
without prior
Annual survival rate
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
Years of data


with prior
without prior
Annual survival rate
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
Years of data



with prior
without prior
Annual survival rate
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
Years of data


with prior
without prior
Annual survival rate
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
Years of data

Estimating a mean

- mean diameter of remnant trees in a park
- sample size of 10 trees
- 42, 43, 58, 70, 47, 51, 85, 63, 58, 46 cm
- what is the mean diameter of trees in the park?



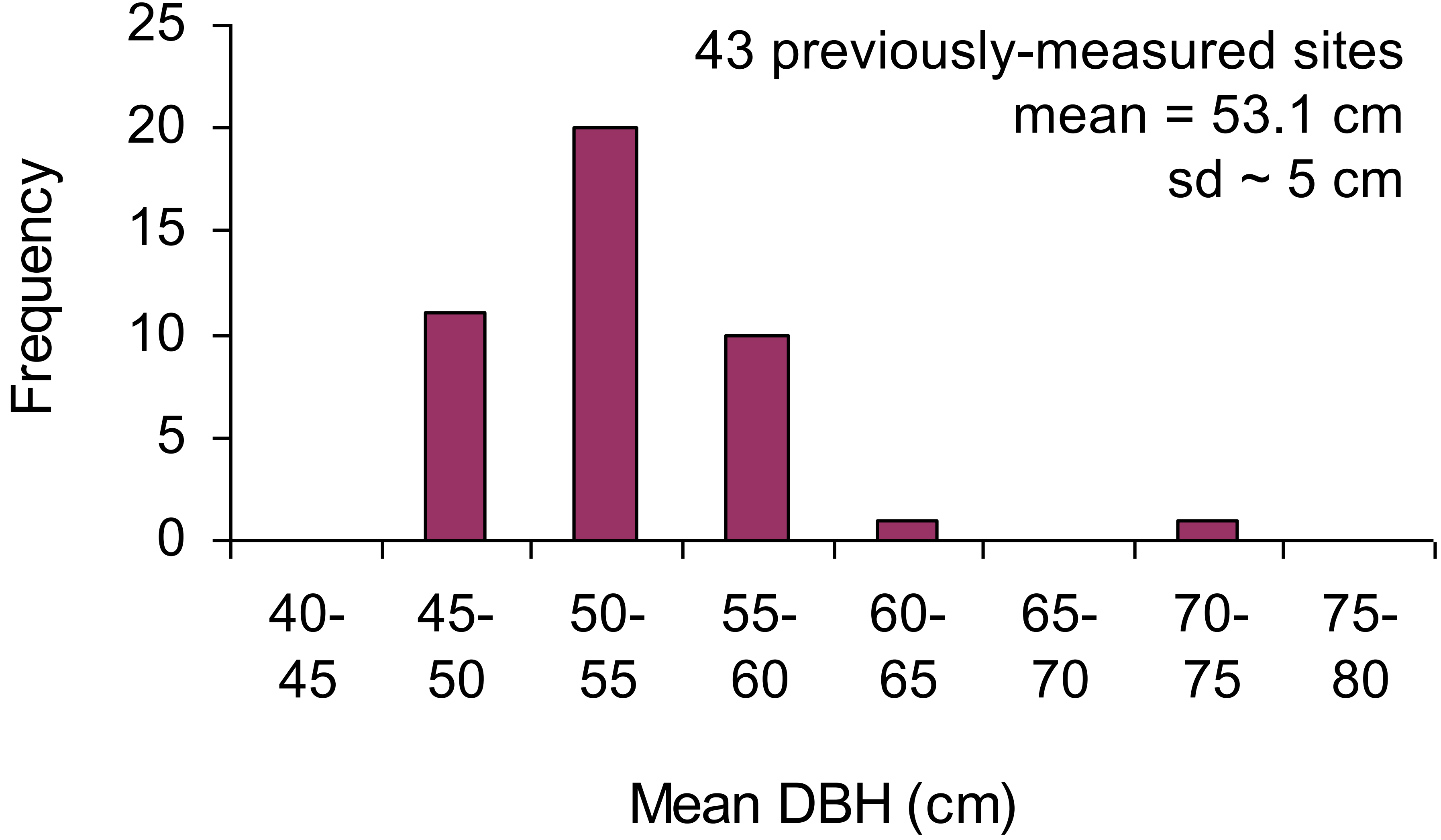

Ignoring prior information

- sample mean = 56.3 cm
- sample variance = 184.9 cm^2
- assume normal distribution

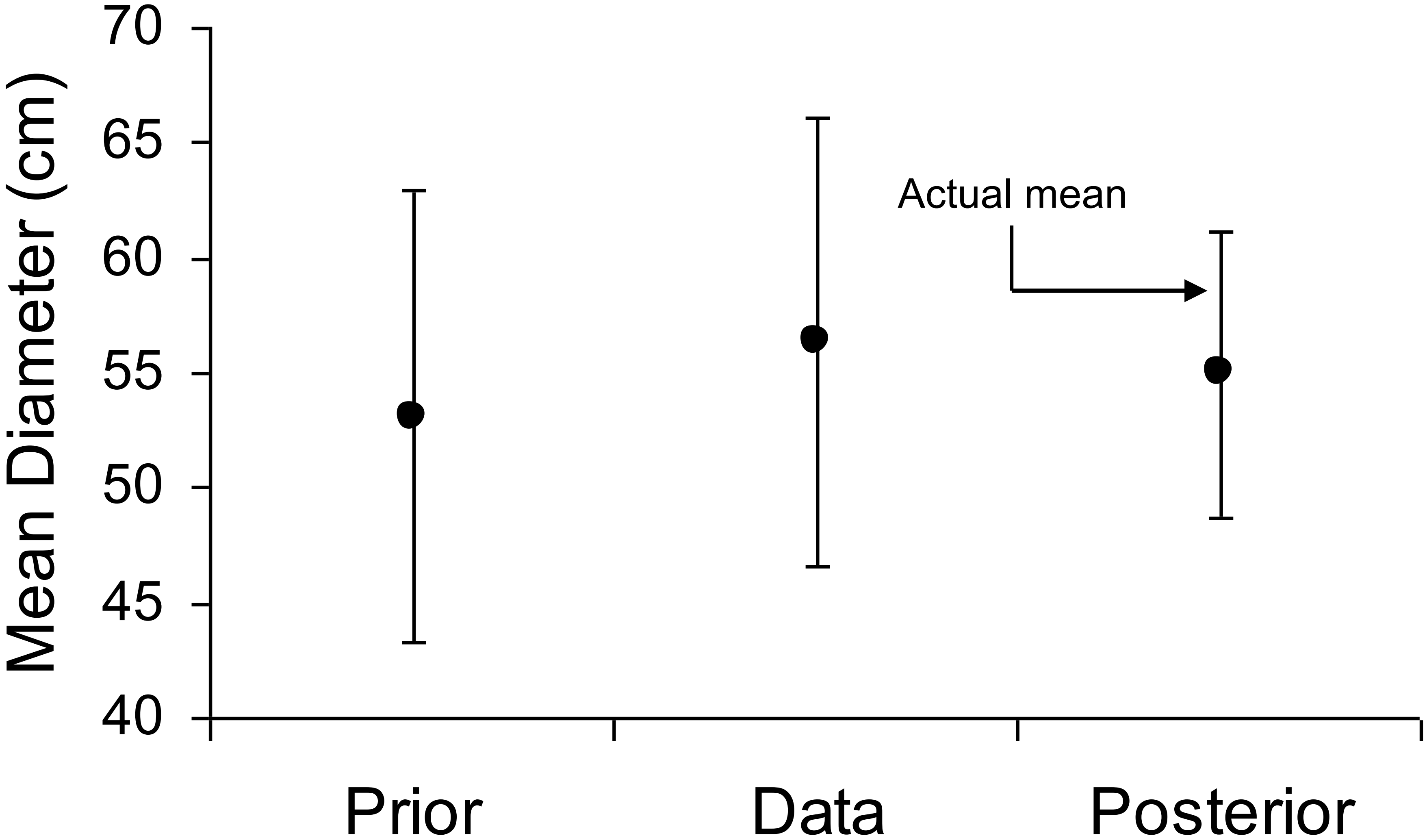
a model

- standard error = 4.3 cm
- t-value (95% CI, df=9) = 2.26
- 95% confidence interval = [46.6 , 66.0] cm

The prior information



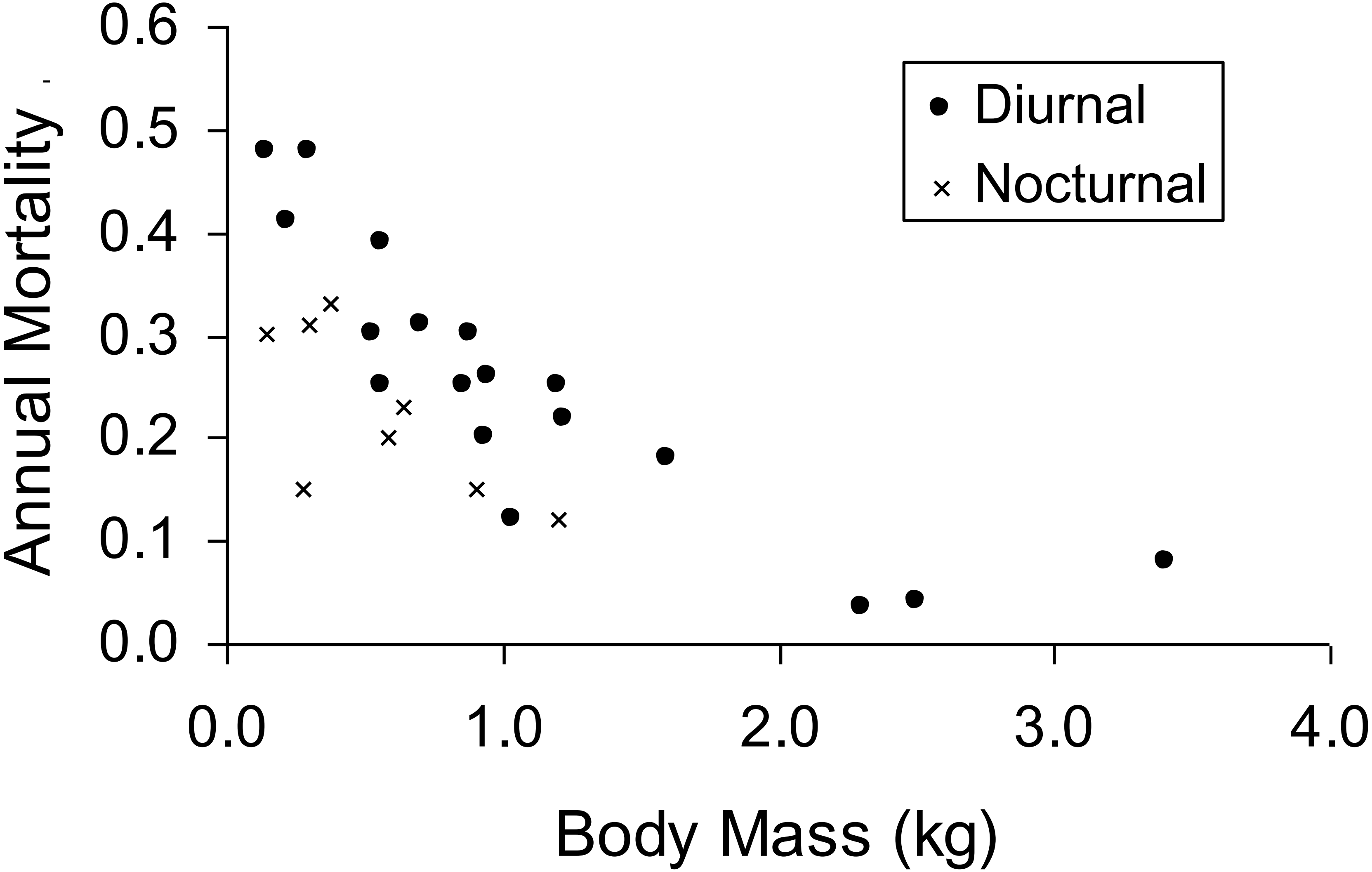
A more precise estimate



Exercise

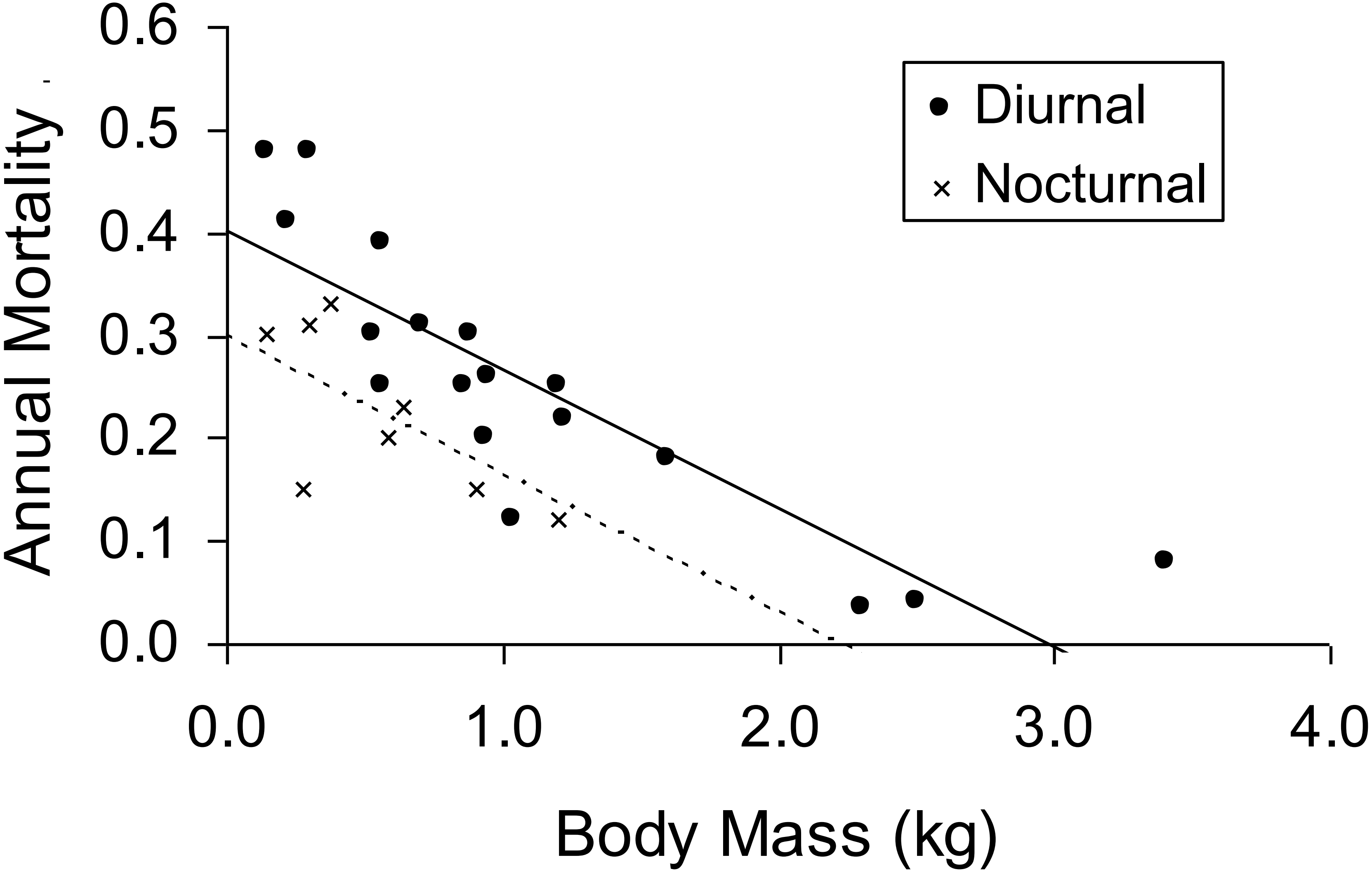
- mean spider web diameter

Regression



Annual Mortality
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Diurnal
Nocturnal
0.0
1.0
2.0
3.0
4.0
Body Mass (kg)

Linear Regression



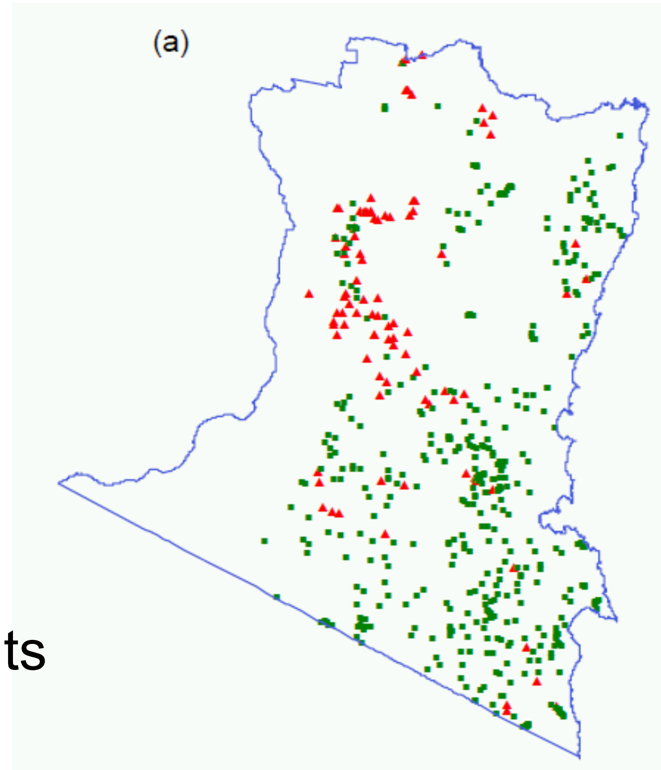
Exercise

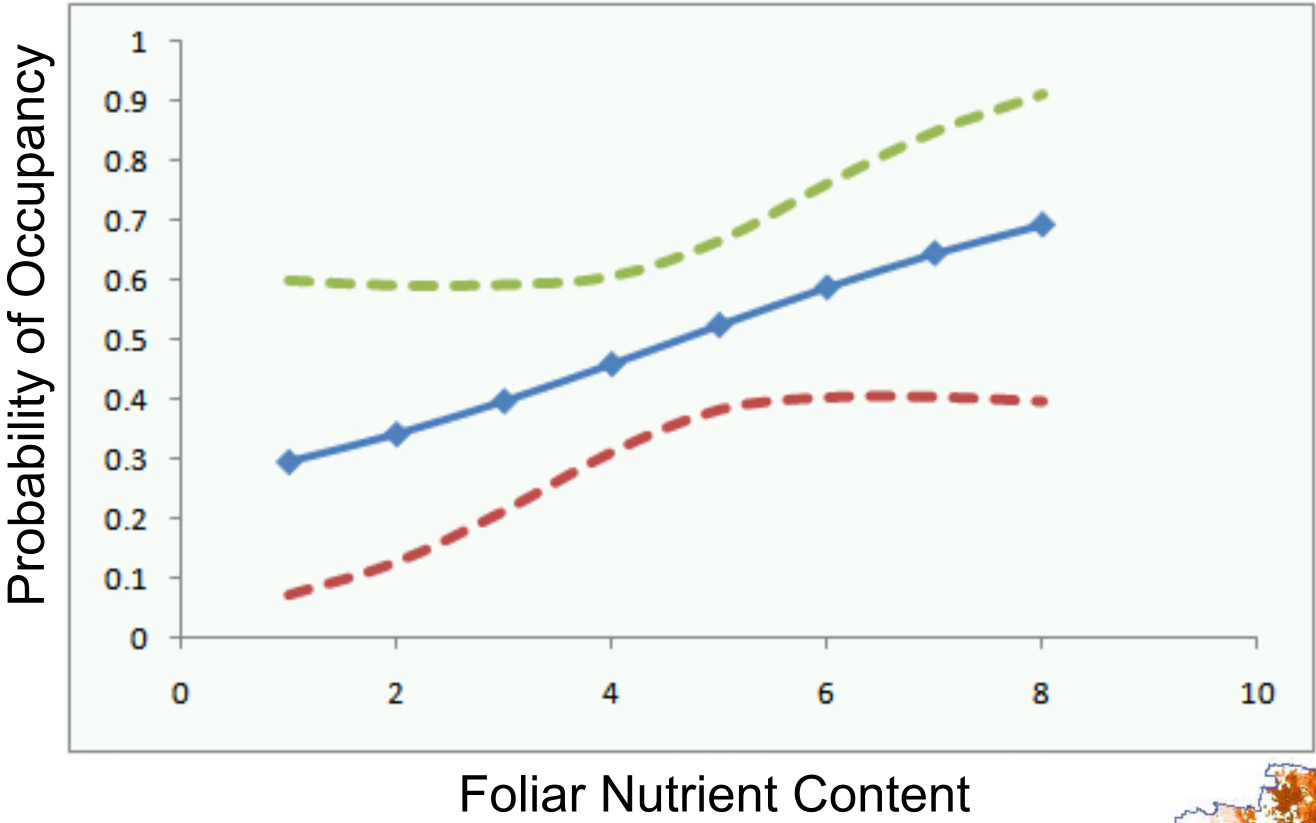
- linear regression CWD

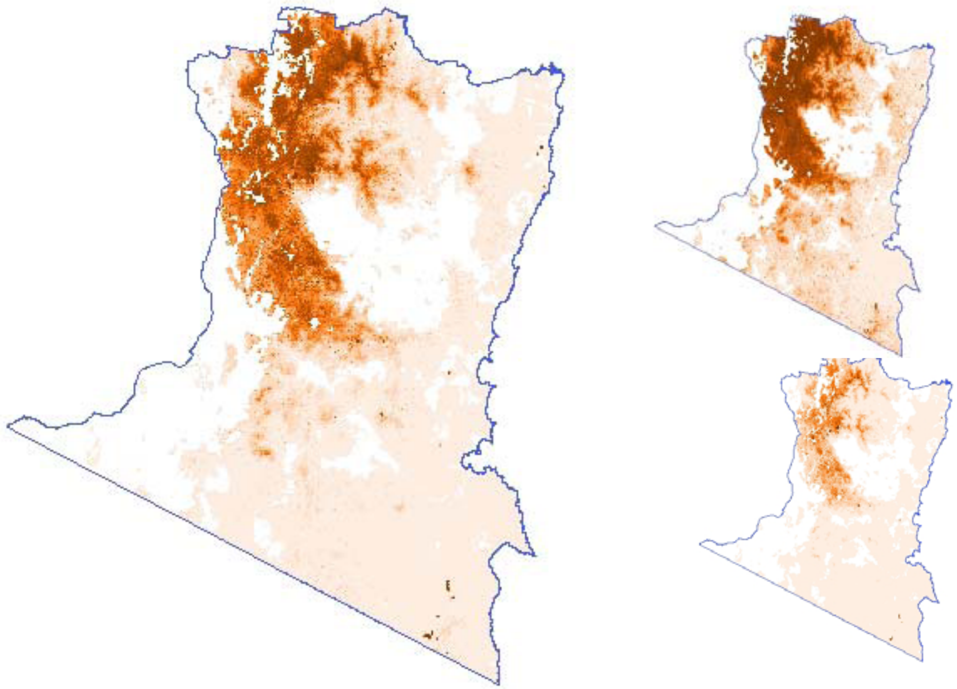
Logistic regression

- $\text{logit}(p) = \ln[p/(1-p)] = a + b_1x_1 + b_2x_2 + \ldots$ **OR**
- $P(Y=1) = \exp(a + b_1x_1 + b_2x_{2.}) / (1 + \exp(a + b_1x_1 + b_2x_{2.}))$
- observe presence (Y=1) or absence (Y=0)
- probability of greater glider presence in forests

Foliar Nutrients




Probability of Occupancy
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
Foliar Nutrient Content

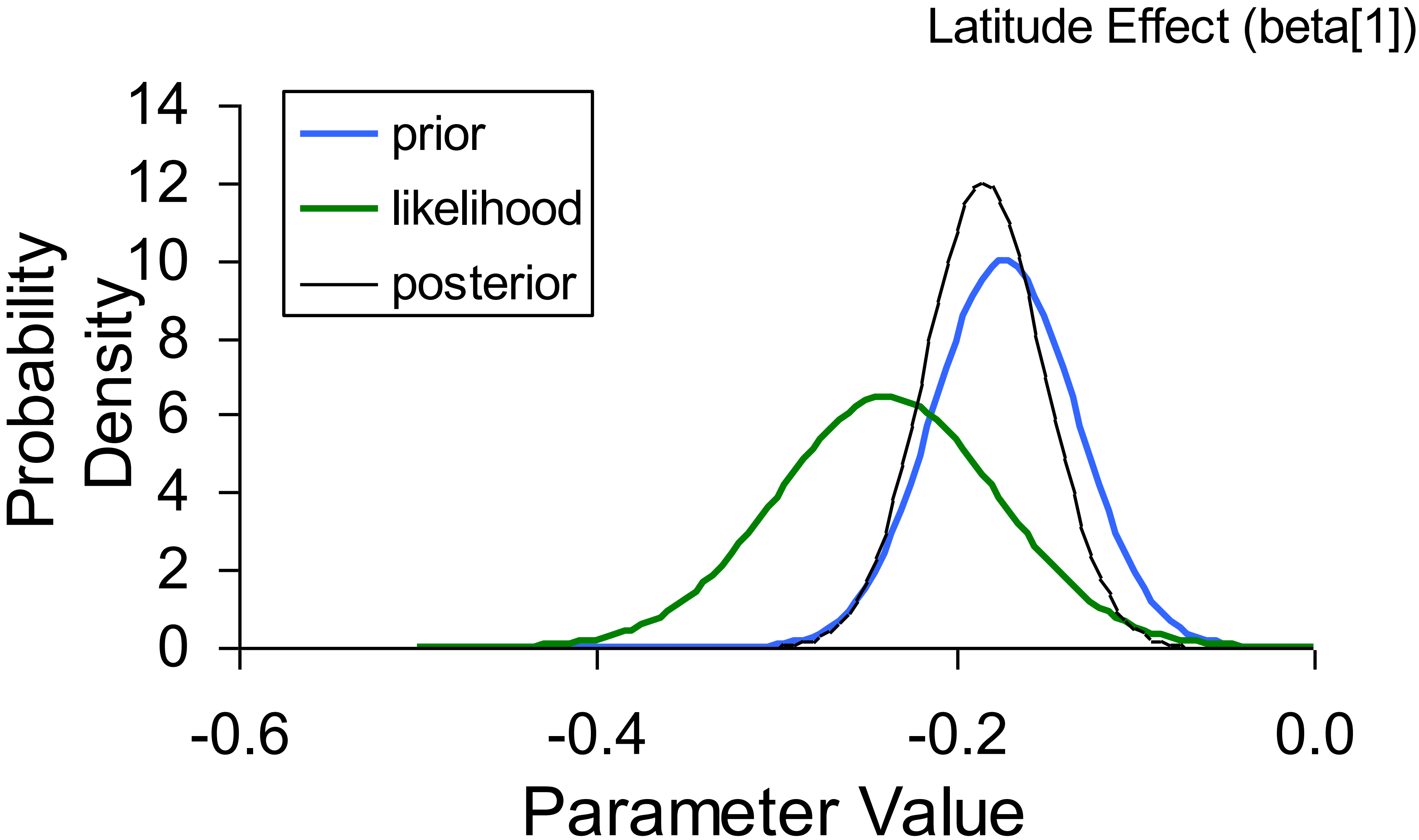


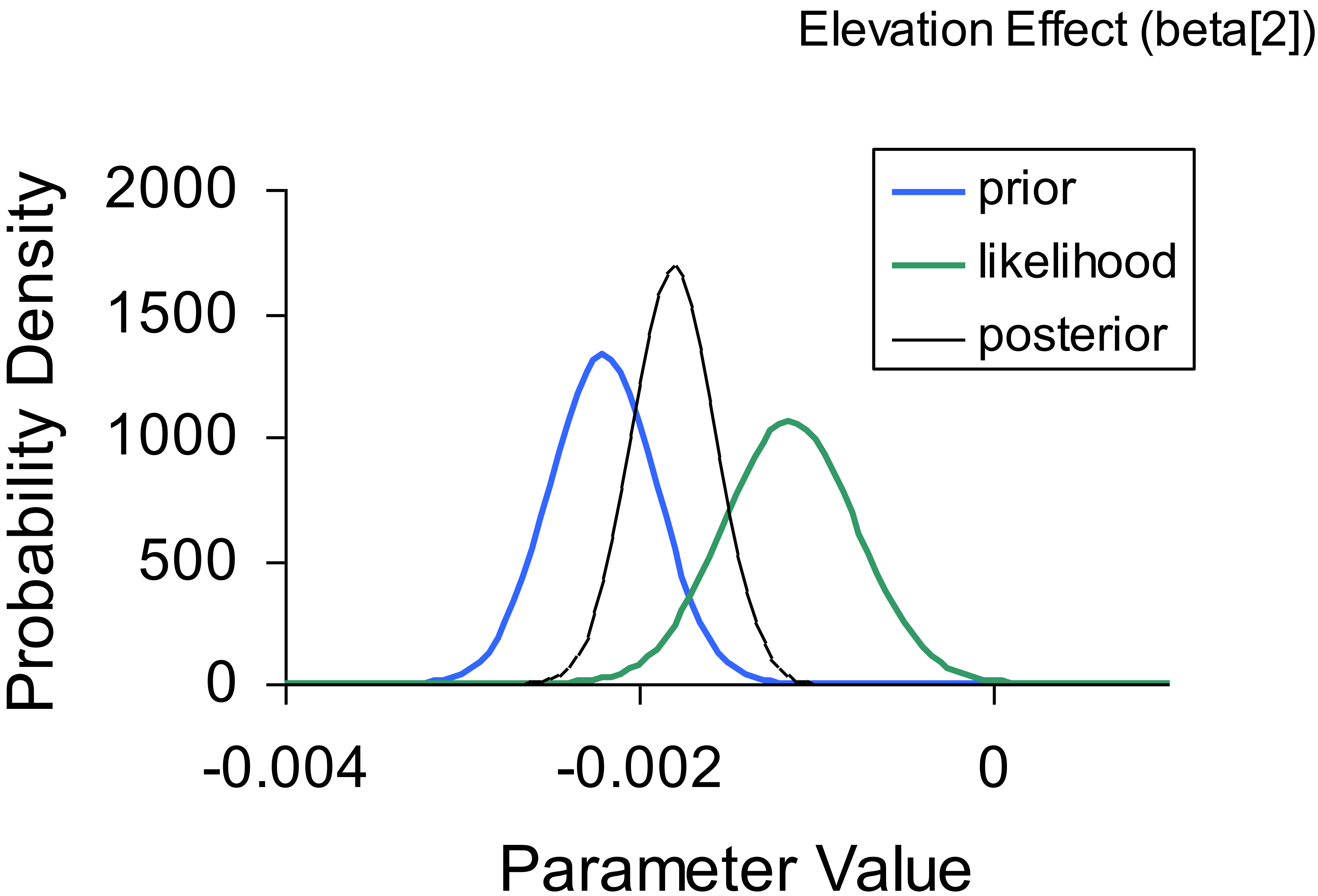
Logistic regression

- Centring data to reduce correlation (numerical stability).
- Exercise… Kirsten Parris' frog surveys.

Poisson regression

- $\log(m) = a + b_1x_1 + b_2x_2 + \ldots$
- $m = \exp(a + b_1x_1 + b_2x_2 + \ldots)$
- observe count ($Y = 0, 1, 2, \ldots$)
- ant species richness (Ellison 2004)
  - habitat (forest / bog)
  - elevation
  - latitude
  - l.EllisonAnts - code

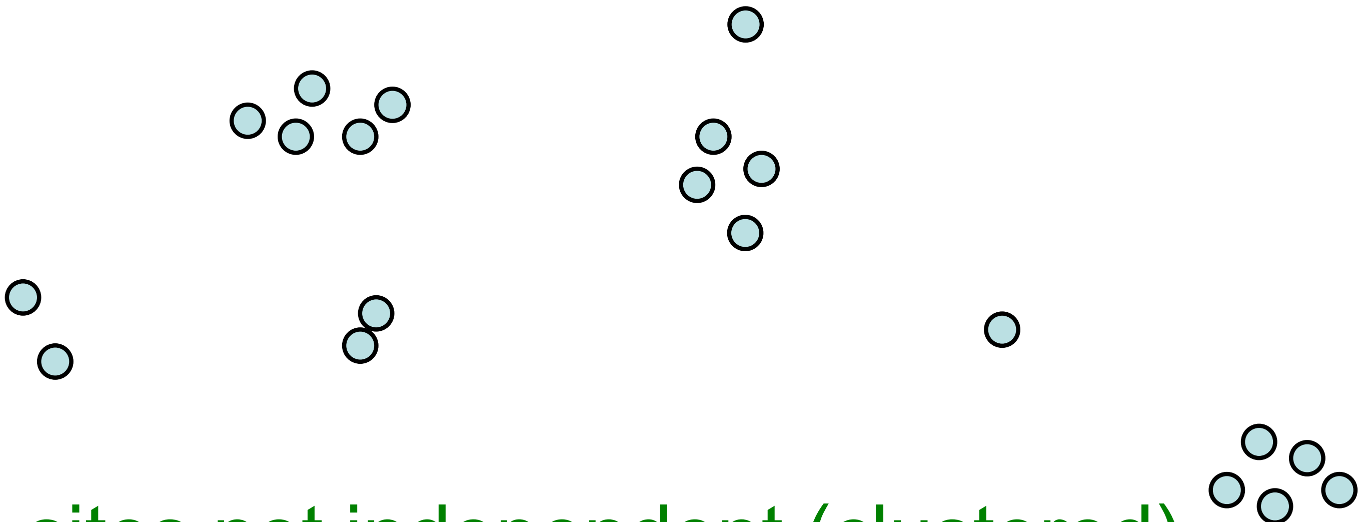

Latitude Effect (beta[1])
prior
likelihood
posterior
Probability Density
14
12
10
8
6
4
2
0
-0.6
-0.4
-0.2
0.0
Parameter Value



Elevation Effect (beta[2])
Probability Density
2000
1500
1000
500
0
-0.004
-0.002
0
Parameter Value
prior
likelihood
posterior

Poisson Regression

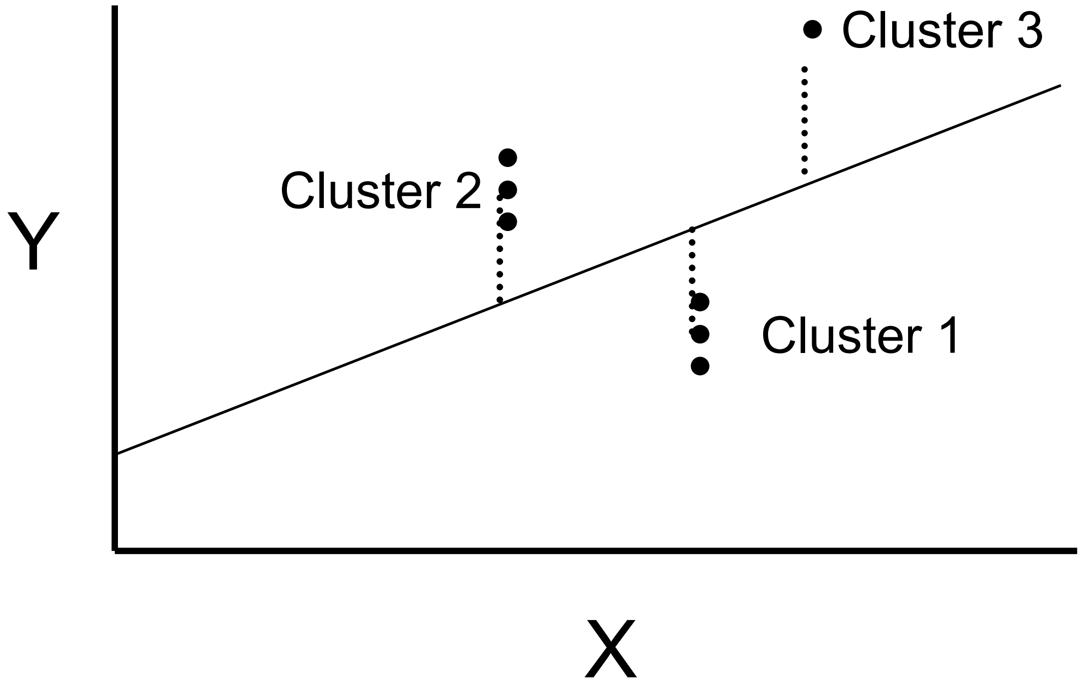
- Frog species richness in ponds
  - area
  - vegetation (emergent, submerged, fringing)
  - vertical wall

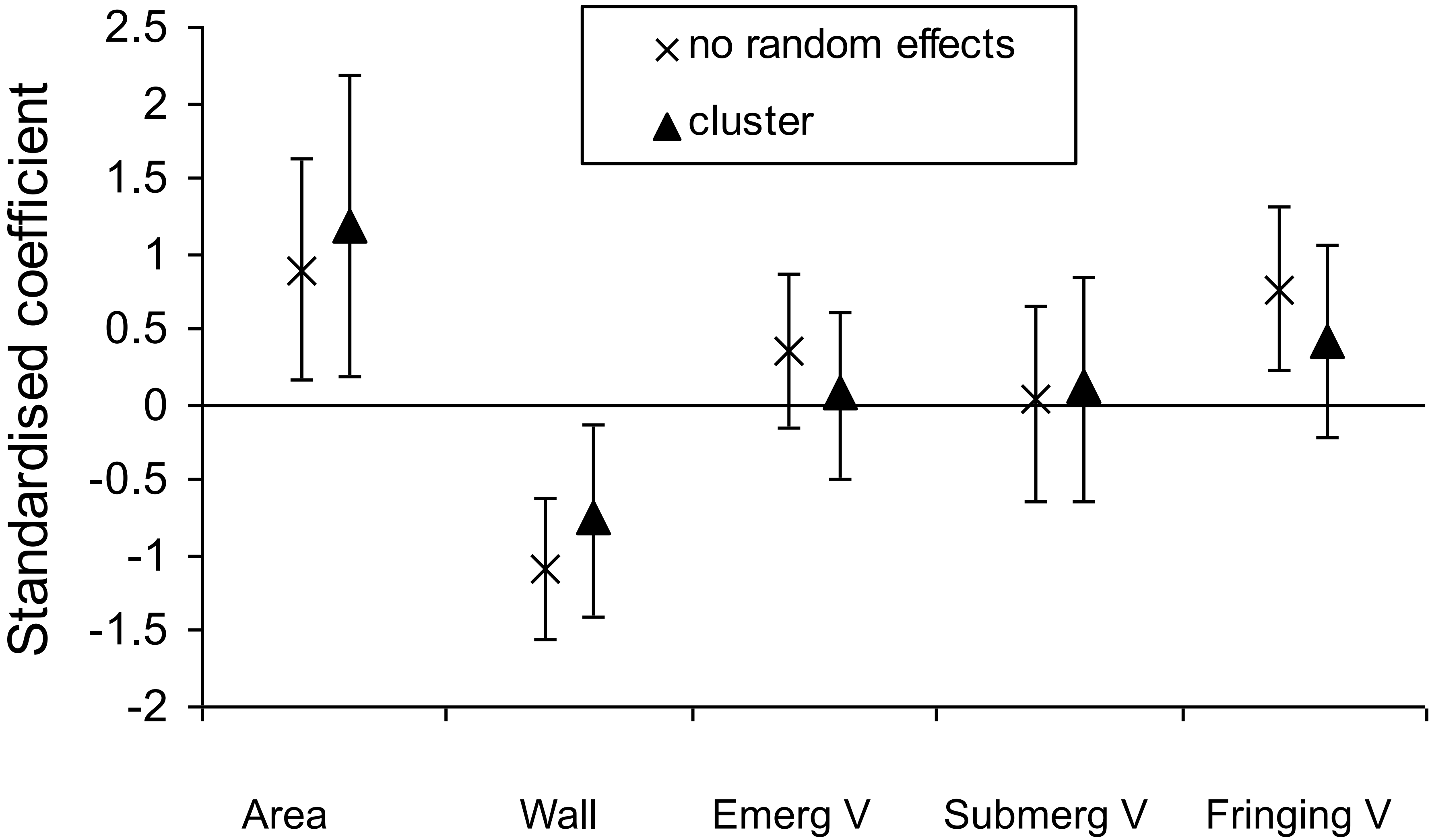


sites not independent (clustered)

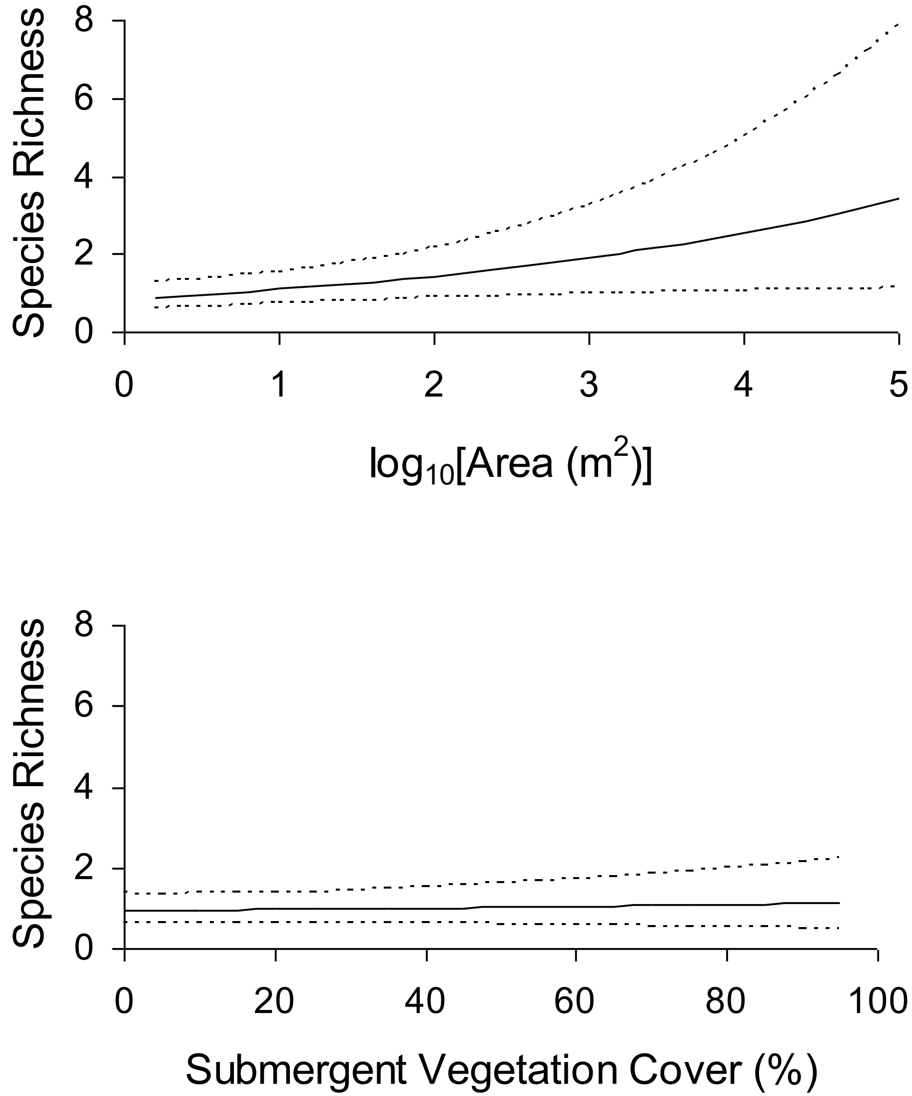
Random effects

- $\log(m) = a + b_1x_1 + b_2x_2 + \ldots + e_{cluster}$
- $e_{cluster}$ - random effect for cluster
- accounts for over-dispersion




no random effects
cluster
Standardised coefficient
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Area
Wall
Emerg V
Submerg V
Fringing V

Poisson regression



Interpreting DIC

ΔDIC	Degree of support
0-2	Substantial
4-7	Considerably less
>10	Essentially none

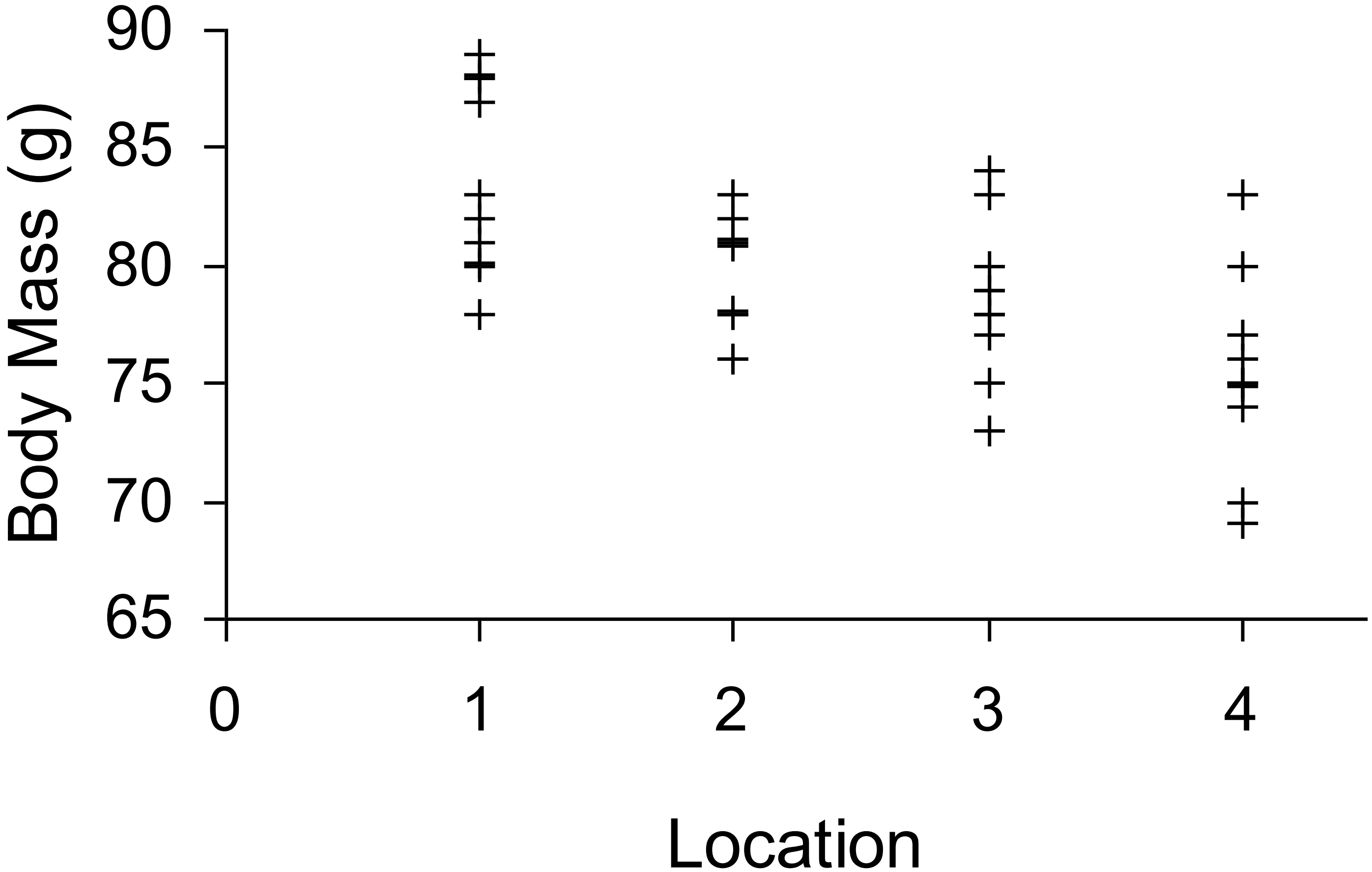
	Model	DIC	ΔDIC
	no random effects	330.5	34.4
Best	cluster	296.1	0.0

Choice

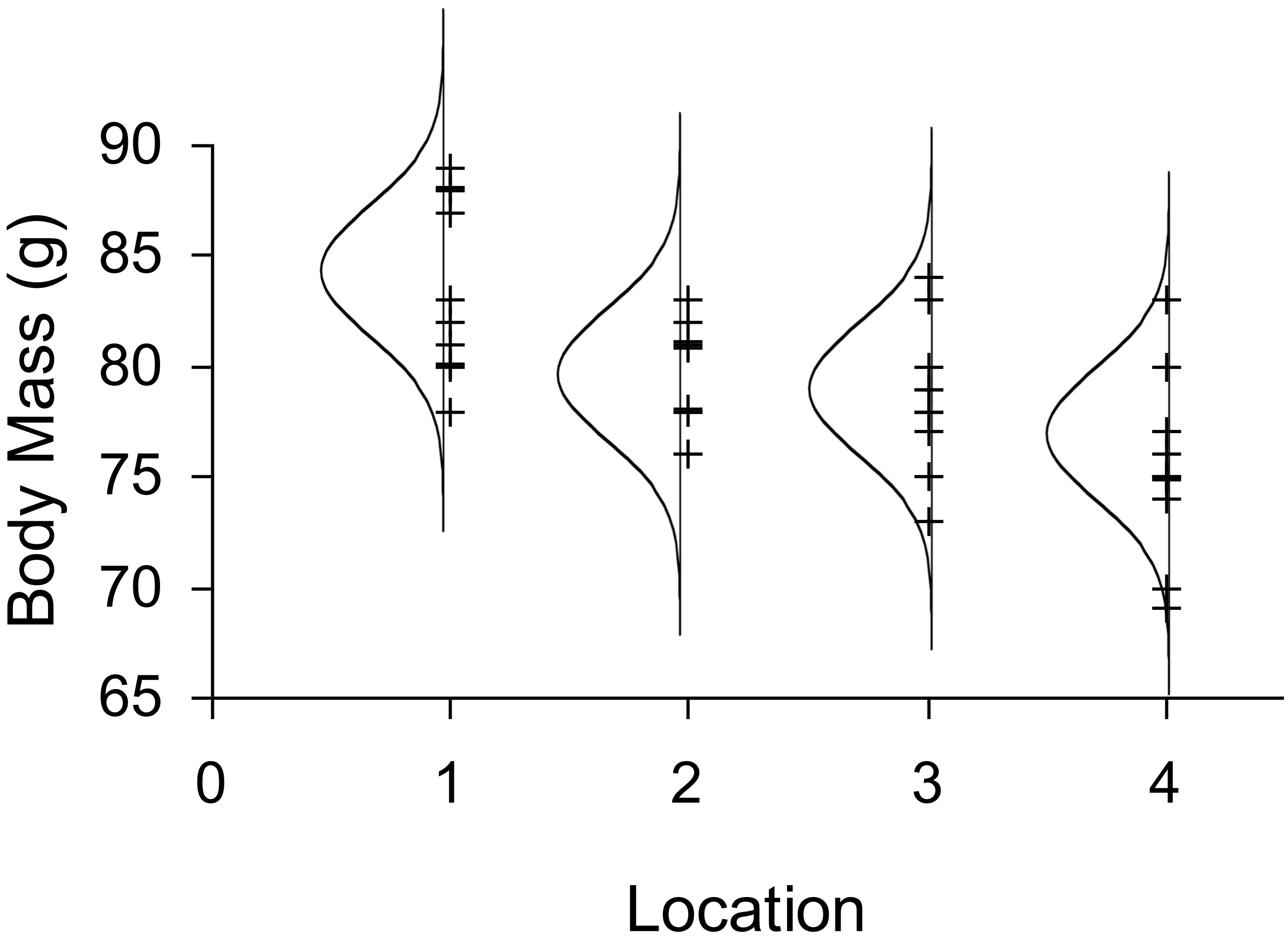
- ANOVA
- detectability

ANOVA

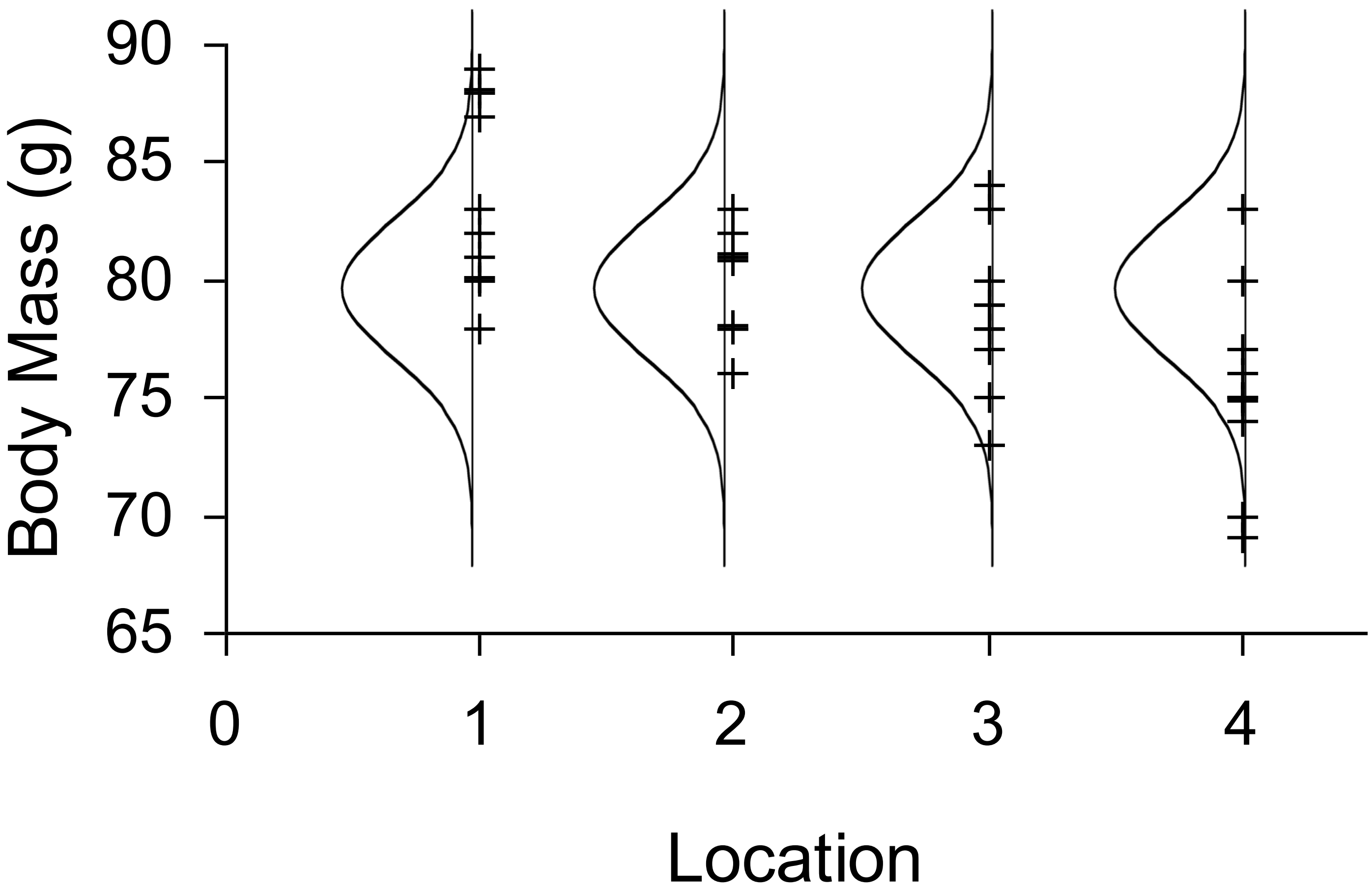
- Body mass of starlings



ANOVA



ANOVA



Detectability in ecology

- Is the species absent, or did I just fail to detect it?

- A problem that is pervasive in ecology

Is it really absent - detection of frogs

- survey for southern brown treefrogs at ponds in Melbourne
- 80% reliability of a survey
- one survey - frog is not observed
- probability of the frog being at the site?

Bayesian method

- two hypotheses
  - – frog is present (H_p)
  - – frog is absent (H_a)
- what is the probability of the hypothesis given the data?

Bayesian method

- two hypotheses
  - frog is present (H_p)
  - frog is absent (H_a)
- probability of data | H_p = 0.2
- probability of data | H_a = 1.0

Bayesian method

- prior probabilities for H_p & H_a?
- ignorance

  equal prior probabilities

  prior(H_p) = prior(H_a) = 0.5

- $\Pr(H_p|D) \propto 0.2 * 0.5 = 0.1$ (norm const = 0.6)
- $\Pr(H_a|D) \propto 1.0 * 0.5 = 0.5$,,

Bayesian method

- $\Pr(H_{\mathrm{p}}|D) = 0.1667$
- $\Pr(H_{\mathrm{a}}|D) = 0.8333$

(scaled so that probabilities sum to 1)

Bayesian method

- prior information based on poor habitat

  unequal prior probabilities

  prior(H_p) = 0.1 prior(H_a) = 0.9

- $\Pr(H_p|D) \propto 0.2 * 0.1 = 0.02$ (norm const = 0.92)

  $= 0.022$

- $\Pr(H_a|D) \propto 1.0 * 0.9 = 0.9$,,

  $= 0.978$

Bayesian method

- prior information based on good habitat

  unequal prior probabilities

  prior(H_p) = 0.75 prior(H_a) = 0.25

- $\Pr(H_p|D) \propto 0.2 * 0.75 = 0.15$ (norm const = 0.4)

  $= 0.375$

- $\Pr(H_a|D) \propto 1.0 * 0.25 = 0.25$,,

  $= 0.625$

How many surveys to be sure a species is absent?

- A question that is common in impact assessment (e.g. pre-harvest surveys for listed species)
- The simplest case:
- Probability of detection in a single visit, if present = p
- Probability of detection if present in v visits

$$= 1 - (1 - p)^n$$

But, what if you have prior information about the probability of occupancy?

- Invest more effort surveying good quality habitats?   It depends…

How many surveys?
Application of Bayes' theorem

$$\Pr(H_p \mid D) = \frac{\Pr(H_p) \times \Pr(D \mid H_p)}{\Pr(H_p) \times \Pr(D \mid H_p) + \Pr(H_a) \times \Pr(D \mid H_a)}$$

$$\Pr(H_p \mid D) = \frac{\Pr(H_p) \times (1 - p)^n}{\Pr(H_p) \times \Pr(1 - p)^n + (1 - \Pr(H_p)) \times 1}$$

$$\Pr(H_p \mid D) = \frac{0.75 \times (1 - 0.8)^2}{0.75 \times (1 - 0.8)^2 + 0.25 \times 1}$$

$$= 0.11$$

How many surveys?
Application of Bayes' theorem

$$\Pr(H_p \mid D) = \frac{\Pr(H_p) \times (1-p)^n}{\Pr(H_p) \times \Pr(1-p)^n + (1-\Pr(H_p)) \times 1}$$

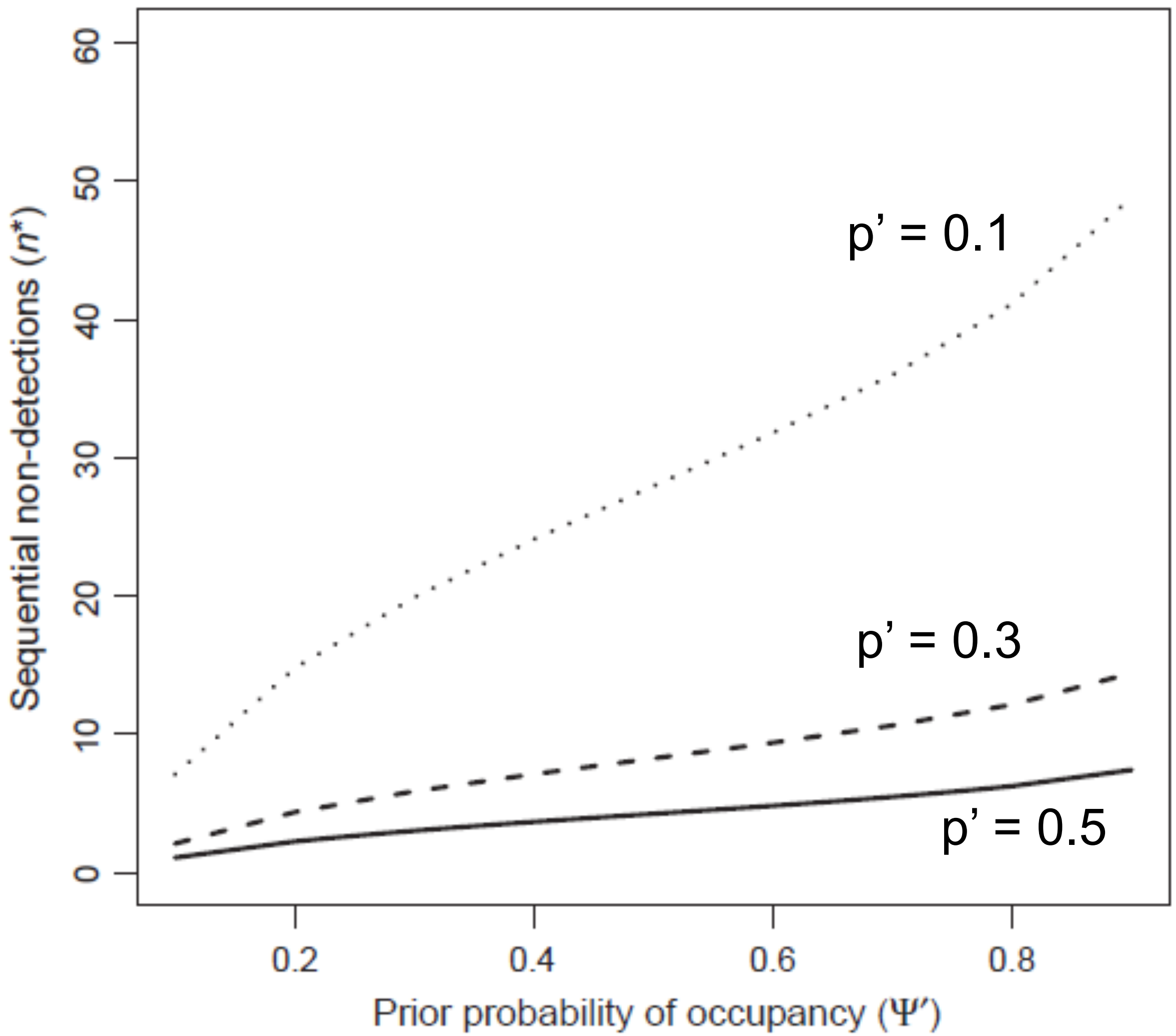
- n non-detections such that $\Pr(H_p|D) < 0.05$

$$n > \frac{\ln\left(\frac{\Pr(H_p|D)^*}{1-\Pr(H_p|D)^*}\right) - \ln\left(\frac{\Pr(H_p)}{1-\Pr(H_p)}\right)}{\ln(1-p)}$$

$$n > \frac{\ln\left(\frac{0.05}{1-0.05}\right) - \ln\left(\frac{0.75}{1-0.75}\right)}{\ln(1-0.8)} \quad = 2.5$$

Required non-detection

$Pr(H_p|D)^* < 0.05$



Detectability

- 8 visits to 50 sites
- 0 – 6 detections per site
- are some of the zeroes false absences?

Wintle B.A., McCarthy M.A., Kavanagh R.P., and Burgman M.A. (2005). The magnitude and management consequences of false negative observation error in surveys of arboreal marsupials and large forest owls. Journal of Wildlife Management 69: 905-917.

Detectability

- Data: y_i = 0,1,2,3… v.

- Observations are derived from a mixture of two distributions

1. Probability of observing a zero:

$$\Pr(Y=0)=1-q+q(1-p)^n$$

2. Probability of observing 1,2,3…v

$$\Pr(Y=y)=q\binom{n}{y}p^y(1-p)^{n-y}$$

Case study results

p = 0.45

ZIB occupancy = 0.50

naïve estimate = 25 occupied / 50 surveyed = 0.5

p = 0.12

ZIB occupancy = 0.98!

naïve estimate = 33 occupied / 50 surveyed = 0.66

Markov chain Monte Carlo (MCMC)

- Monte Carlo
  - random sampling from probability distributions
- Markov Chain
  - next sample depends on the previous one
  - removes constant of proportionality
    - “divides out” of the equation
- not resampling the data

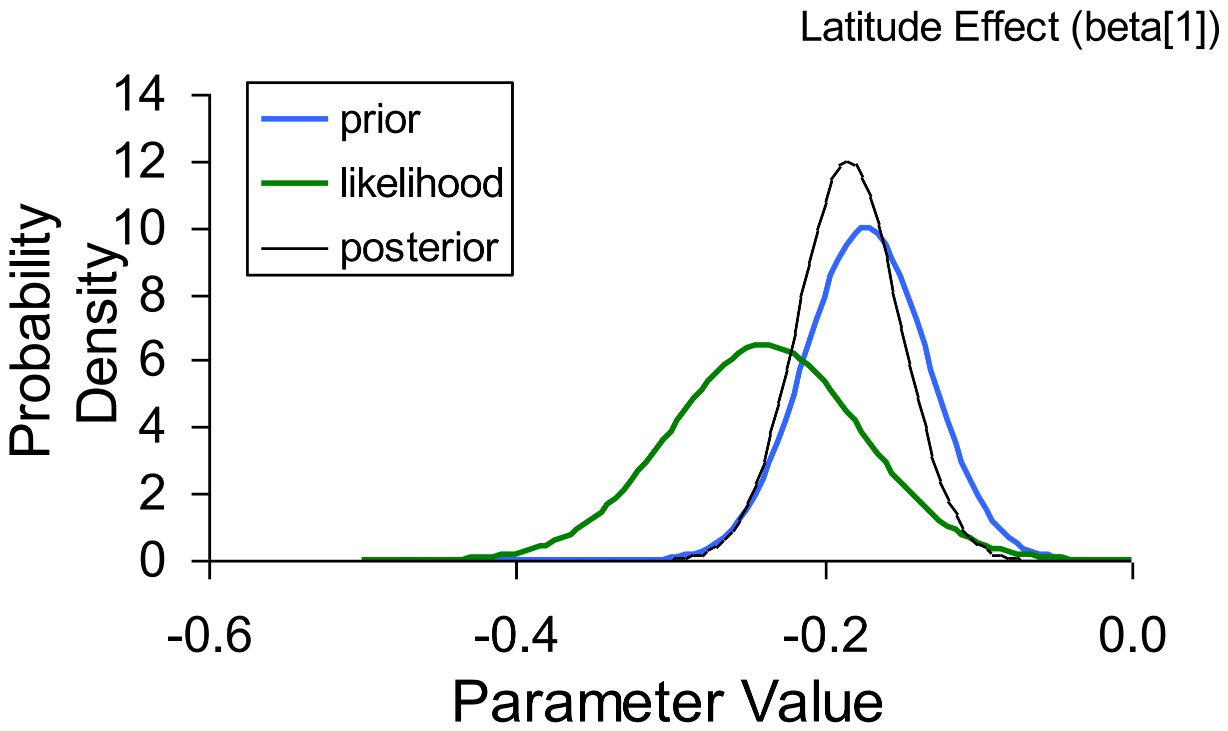
mmc…

Markov Chain

- high correlation means
  - need more samples for same precision
  - takes longer to reach stationary distribution
  - can get “trapped” in one part of posterior
- checks
  - check histories
  - take lots of samples
  - use more than one set of initial values
  - use parameterizations that reduce correlation

Distributions

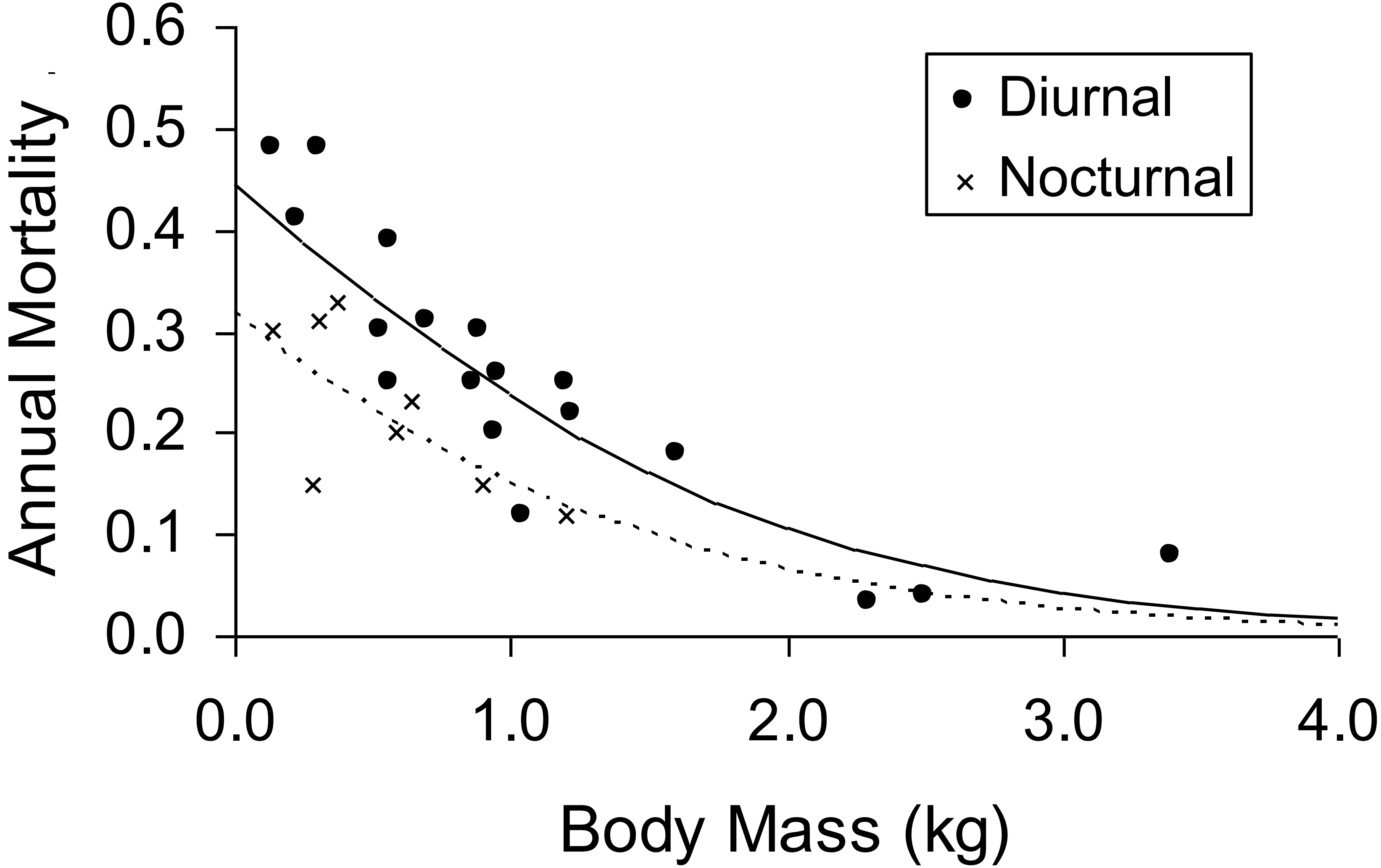
- Central to Bayesian inference
  - Prior distribution Pr(θ)
  - Data distribution Pr(D|θ) - likelihood
  - Posterior distribution: Pr(θ|D)



Distributions

- variation in the model
- uncertainty in the parameters
- different distributions
  - dbern()
  - dnorm()
  - dlnorm()
  - dpois()
  - dbin()
  - dgamma()

Non-linear Regression



Analysing Proportions

Owl Mortality

- 1 death in 35 bird years
- mean: 0.05
- 95% CI: 0.0068 – 0.145
- av. lifespan 6.9 – 147 years
- confidence interval consistent with data
- but inconsistent with common sense

Powerful Owl Mortality

- uniform (0, 1) prior implies 1 or ∞ years is reasonable
- uniform (0.02, 0.2) might be more reasonable
- this prior gives…
  - mean: 0.061
  - 95% CI: 0.022 - 0.146 (7 – 45 years)
- but arbitrary prior…

Powerful Owl Mortality

- uniform (5, 50) years for average lifespan
  - mean: 0.037
  - 95% CI: 0.020 - 0.090 (11 – 50 years)

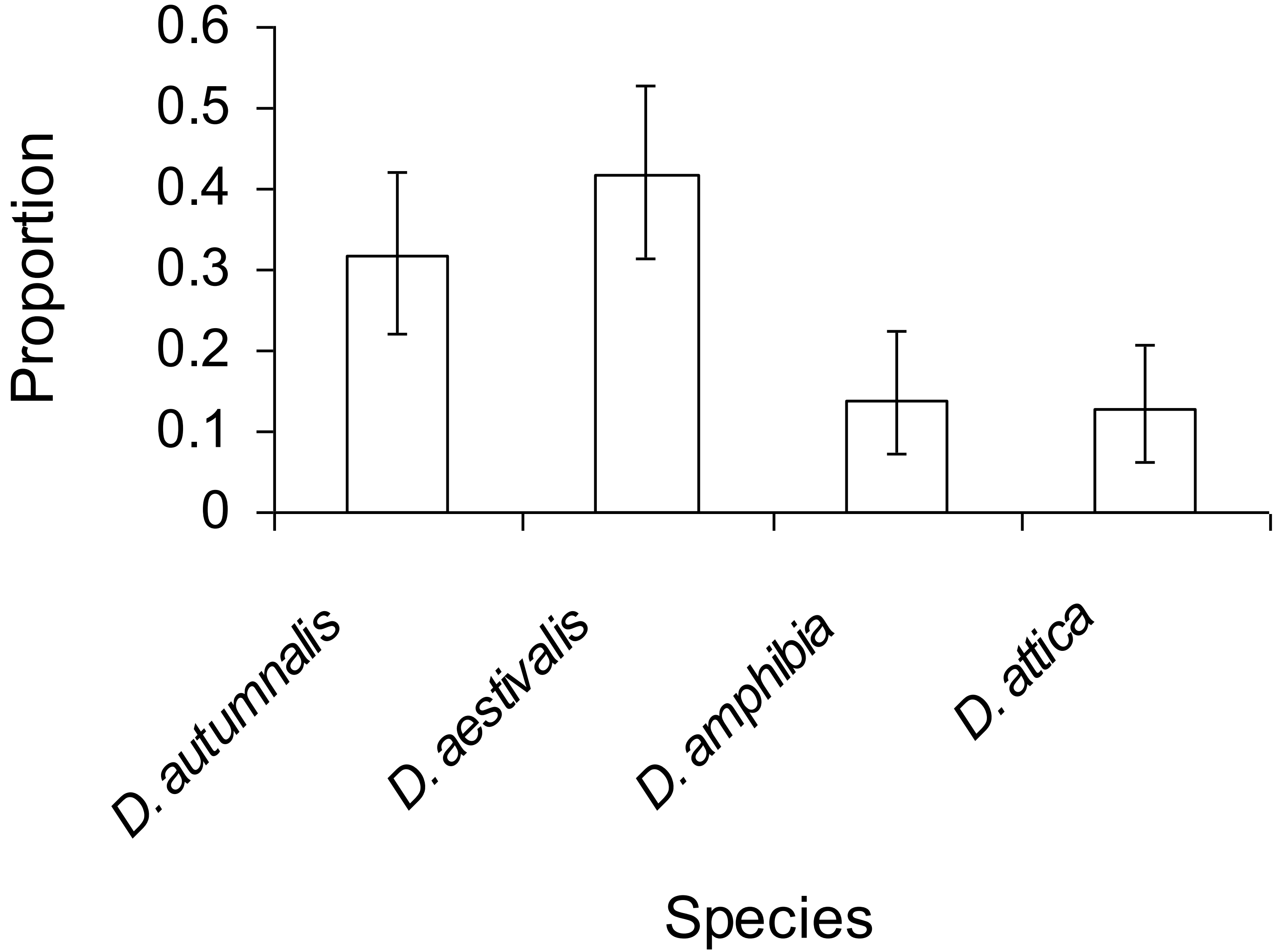
- uniform (0.02, 0.2) mortality
  - mean: 0.061
  - 95% CI: 0.022 - 0.146 (7 – 45 years)

Diversity of a pond community

75 pupae of *Dixella*

D. autumnalis	*D. aestivalis*	*D. amphibia*	*D. attica*
24	32	10	9

What is the proportion of each species in the community?



Proportion
0.6
0.5
0.4
0.3
0.2
0.1
0
D. autumnalis
D. aestivalis
D. amphibia
D. attica
Species

Data Transformation

- Shannon diversity index

$$H = -\sum_{i=1}^{S} P_i \ln(P_i)$$

- mean: 1.25

Data Transformation

- Shannon diversity index

$$H = -\sum_{i=1}^{S} P_i \ln(P_i)$$

- mean: 1.25
- 95% CI: 1.13 – 1.34

Data Transformation

- Shannon diversity index

$$H = -\sum_{i=1}^{S} P_i \ln(P_i)$$

- mean: 1.25
- 95% CI: 1.13 – 1.34

- what if we knew an un-sampled species was in the pond?

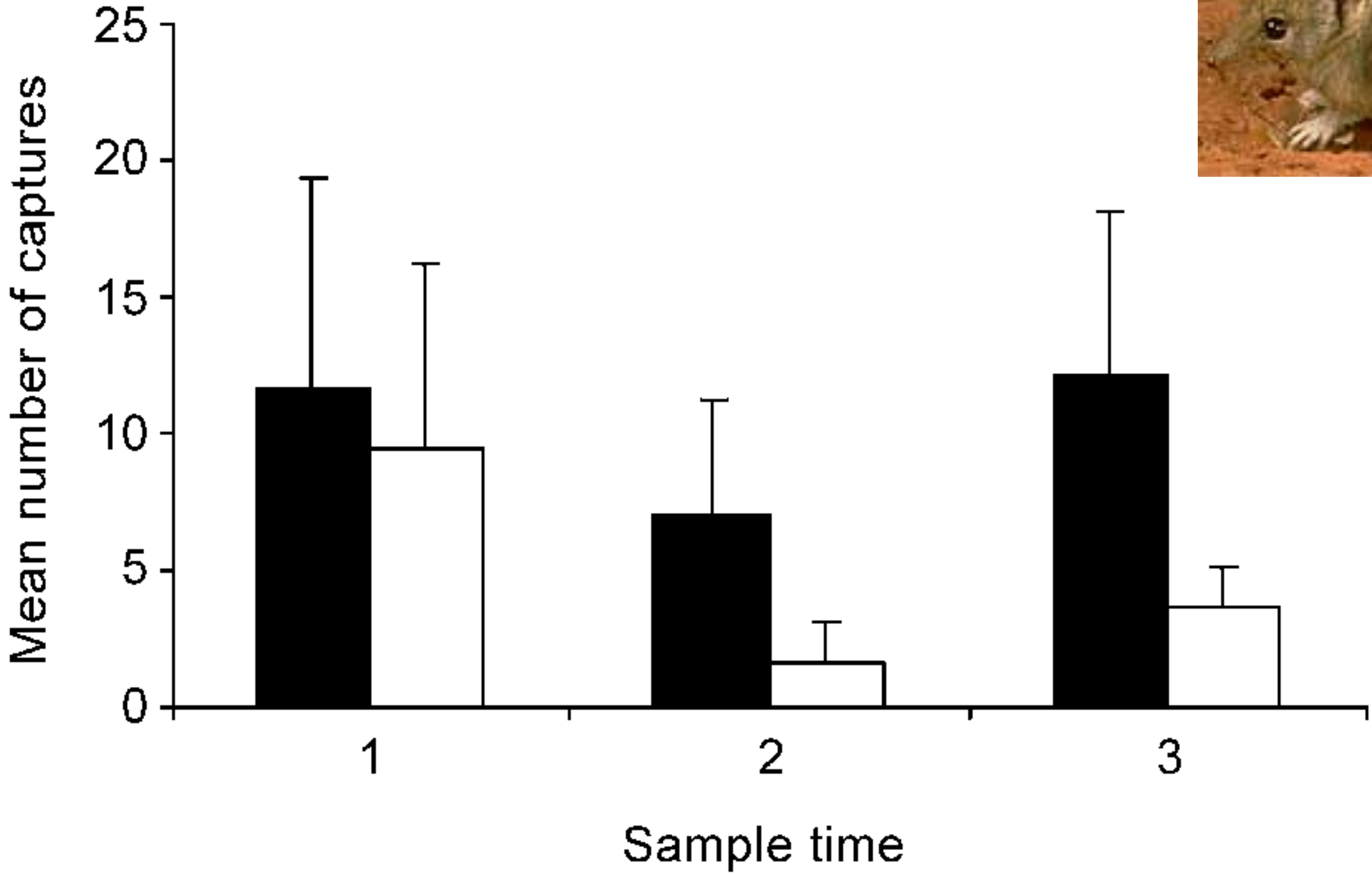


Fig. 3. Mean number of *Dasycercus cristicauda* captures on the (□) harvested and the (■) control plots. Sample before harvesting of spinifex (1), soon after (2), and a year after harvesting (3). Error bars represent SD.

from Masters *et al.* (2003)

Alternative models

- No effect of treatment
- Unknown effect of treatment

Alternative models

- No effect of treatment (“null”)
- Unknown effect of treatment (“alternative”)
- $P = 0.15$

Alternative models

- No effect of treatment
- Unknown effect of treatment
- Effect of treatment consistent with observational study

Alternative models

- No effect of treatment
- Unknown effect of treatment
- Effect of treatment consistent with observational study
- Three points of view

Mulgara results

Model	DIC
No effect	62.8
Uncertain effect	61.8
Effect consistent with observational study	58.0

Mulgara results

Model	DIC
No effect	62.8
Uncertain effect	61.8
Effect consistent with observational study	58.0

Best Model

Beta distribution

- mean = $m = a/(a+b)$
- variance = $v = ab/[(a+b)^2(a+b+1)]$

- $a = m[m(1-m)/v - 1]$
- $b = (1-m)[m(1-m)/v - 1]$

- exercise
  - prior (m=0.122, v=0.00281)
  - data (x=1, n=35)
  - posterior (a and b, m and v?)

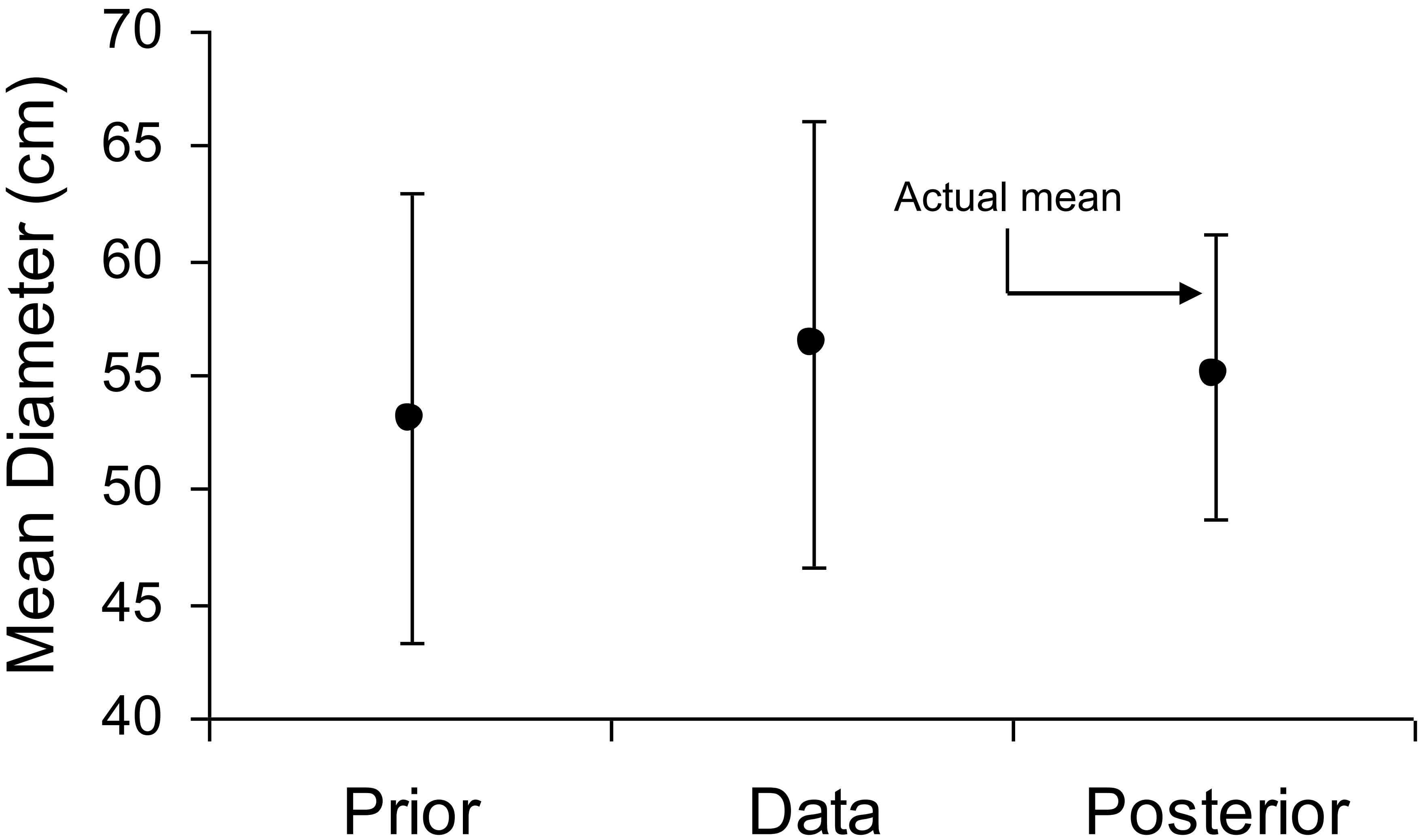
Beta distribution

- uniform prior, and…
- x observations in n independent trials lead to…
- beta distribution, with parameters x+1 and n+1

Beta distribution

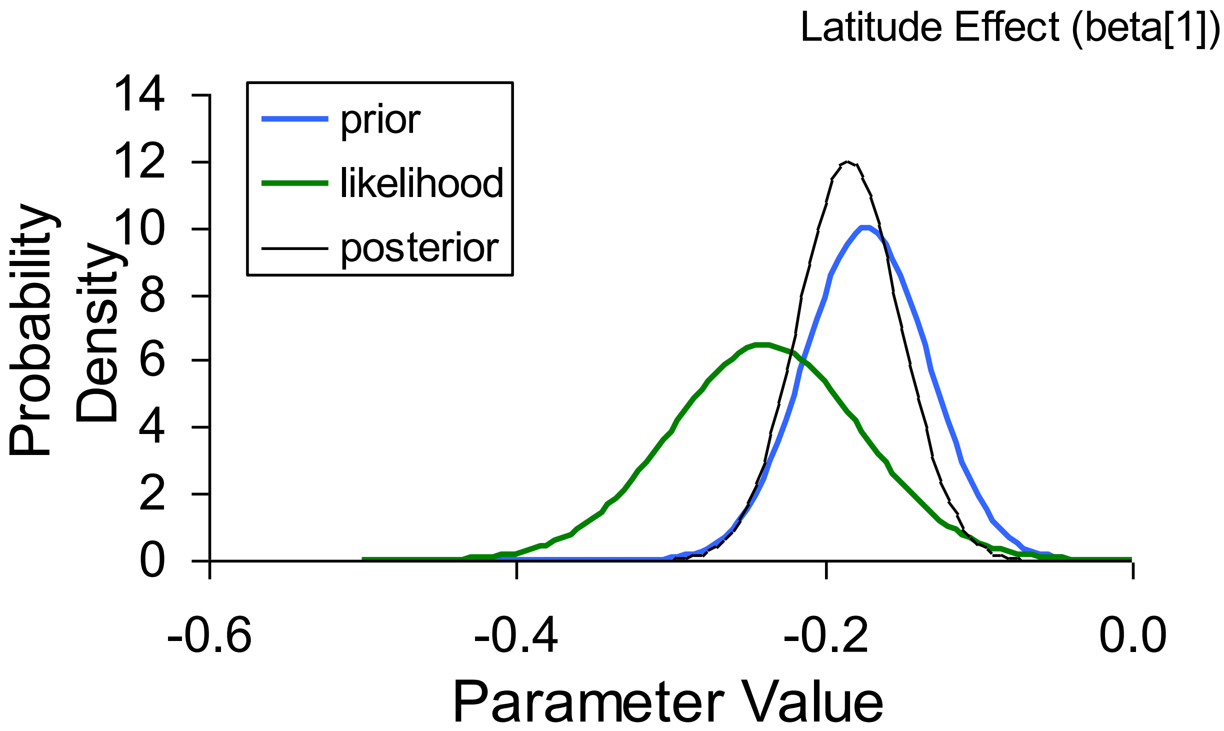
- prior
  - beta with parameters *a* and *b*
- data
  - *x* successes and *n*–*x* failures
- posterior
  - beta with parameters *a*+*x* and *b*+*n*–*x*
- successes and failures increment *a* and *b*

A more precise estimate



Distributions

- Central to Bayesian inference
  - Prior distribution Pr(θ)
  - Data distribution Pr(D|θ) - likelihood
  - Posterior distribution: Pr(θ|D)



Choosing the best models

- Information Criterion
  - fit penalised for complexity (parsimony)
- Deviance
  - how poorly does the model fit the data
- Number of estimated parameters
  - how complex is the model
- The best model is a trade-off
  - fewer parameters -> more precision per parameter
  - more parameters -> better fit to data

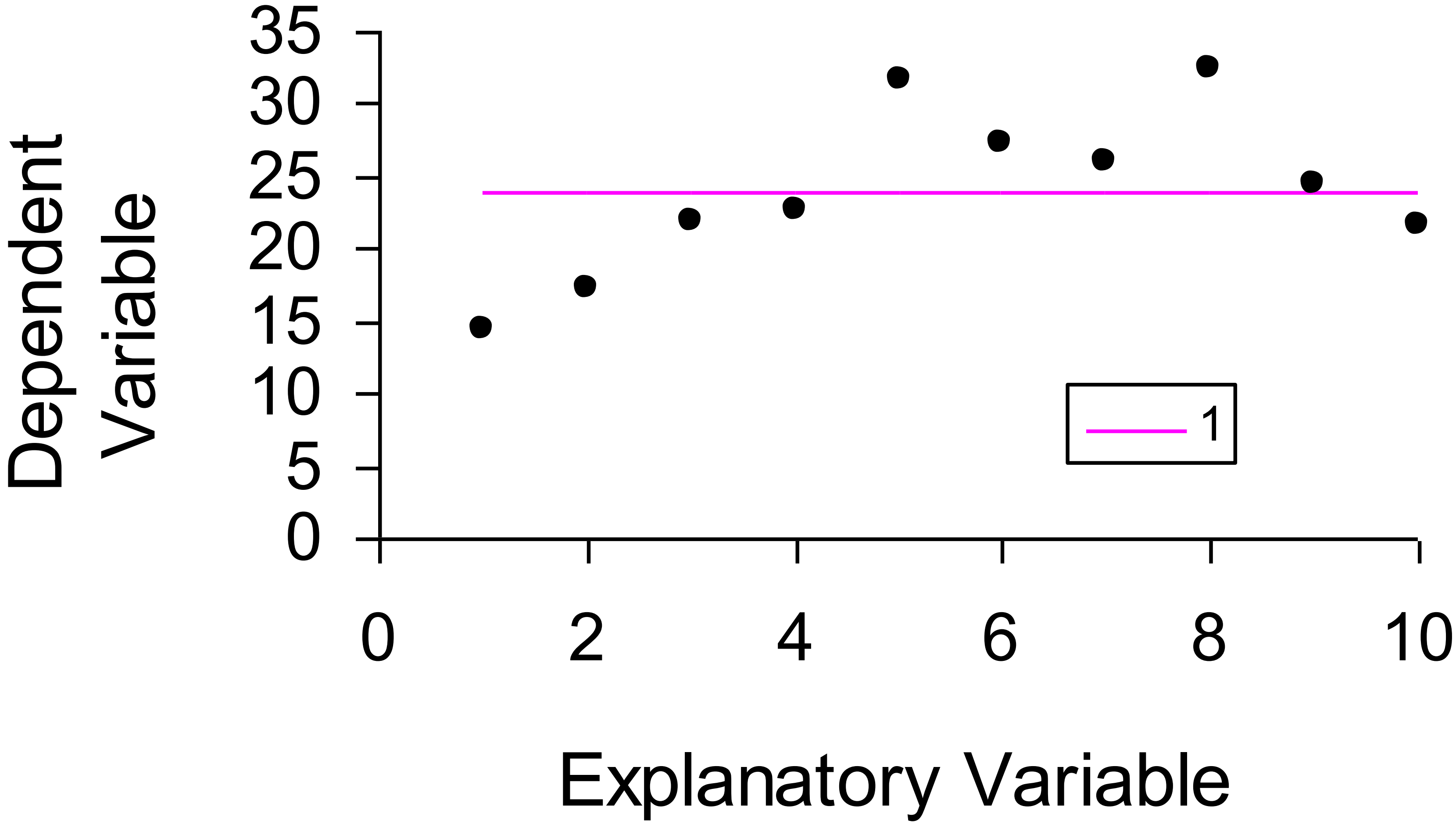

9 kB

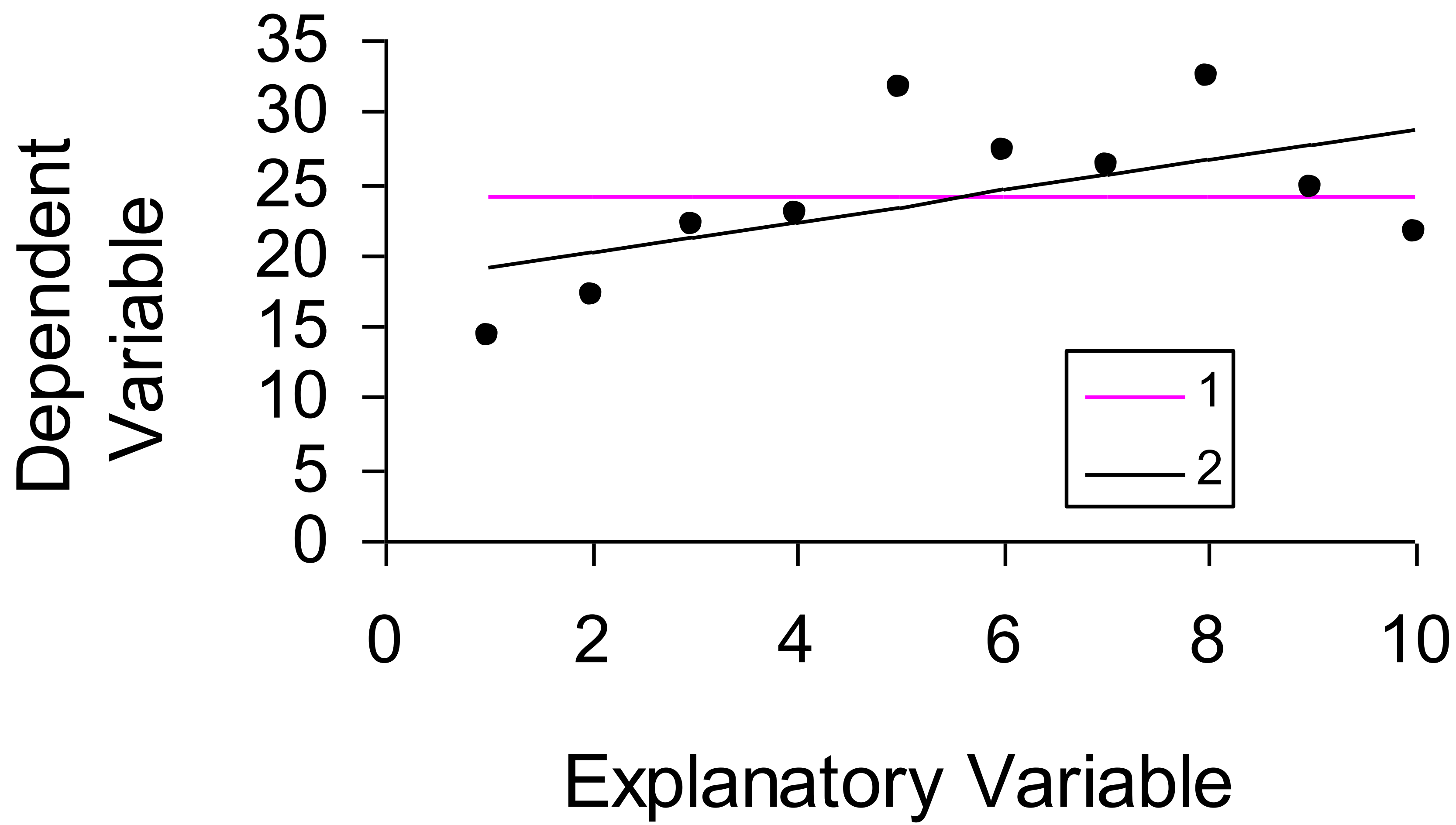

25 kB

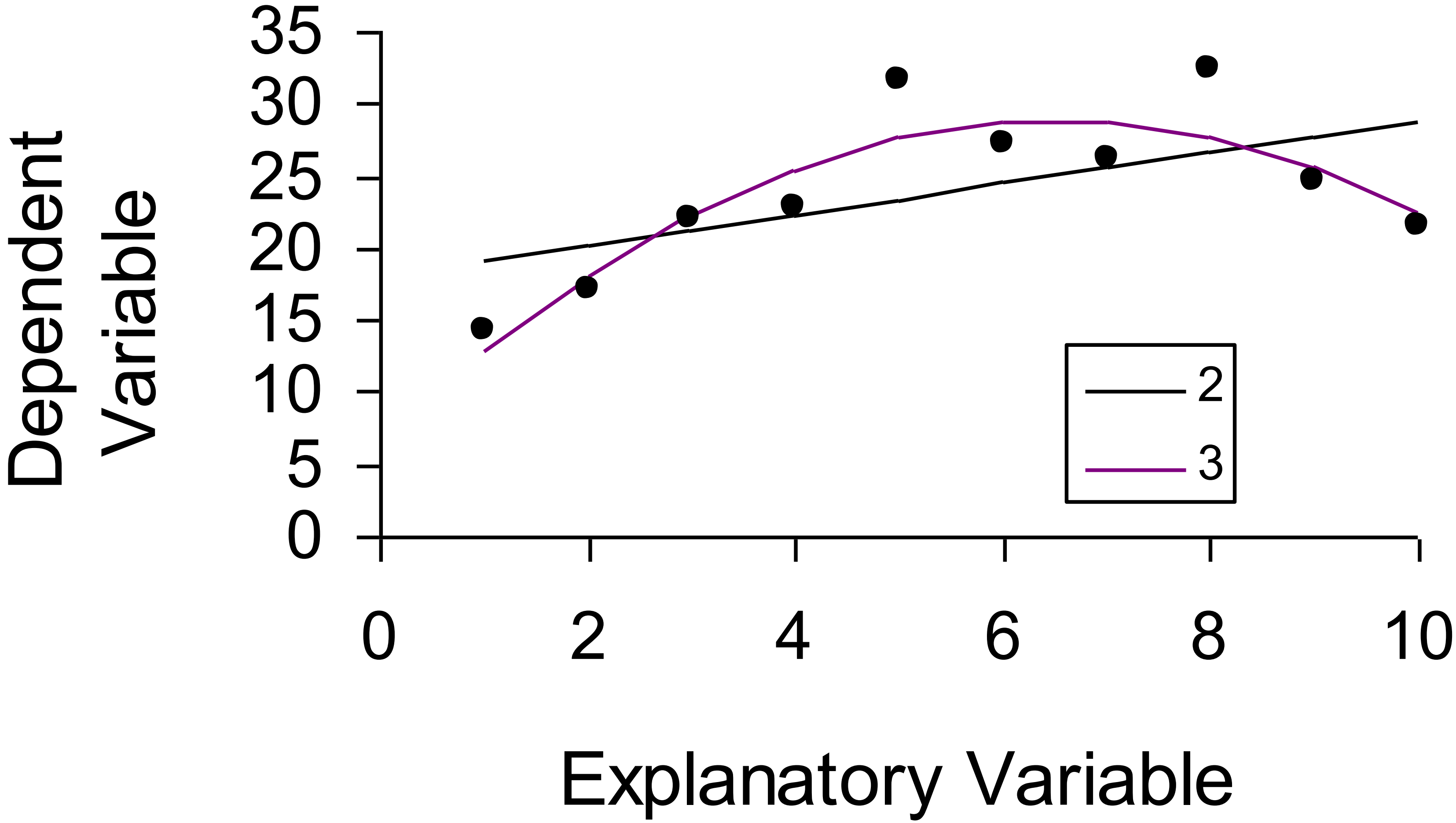

67 kB

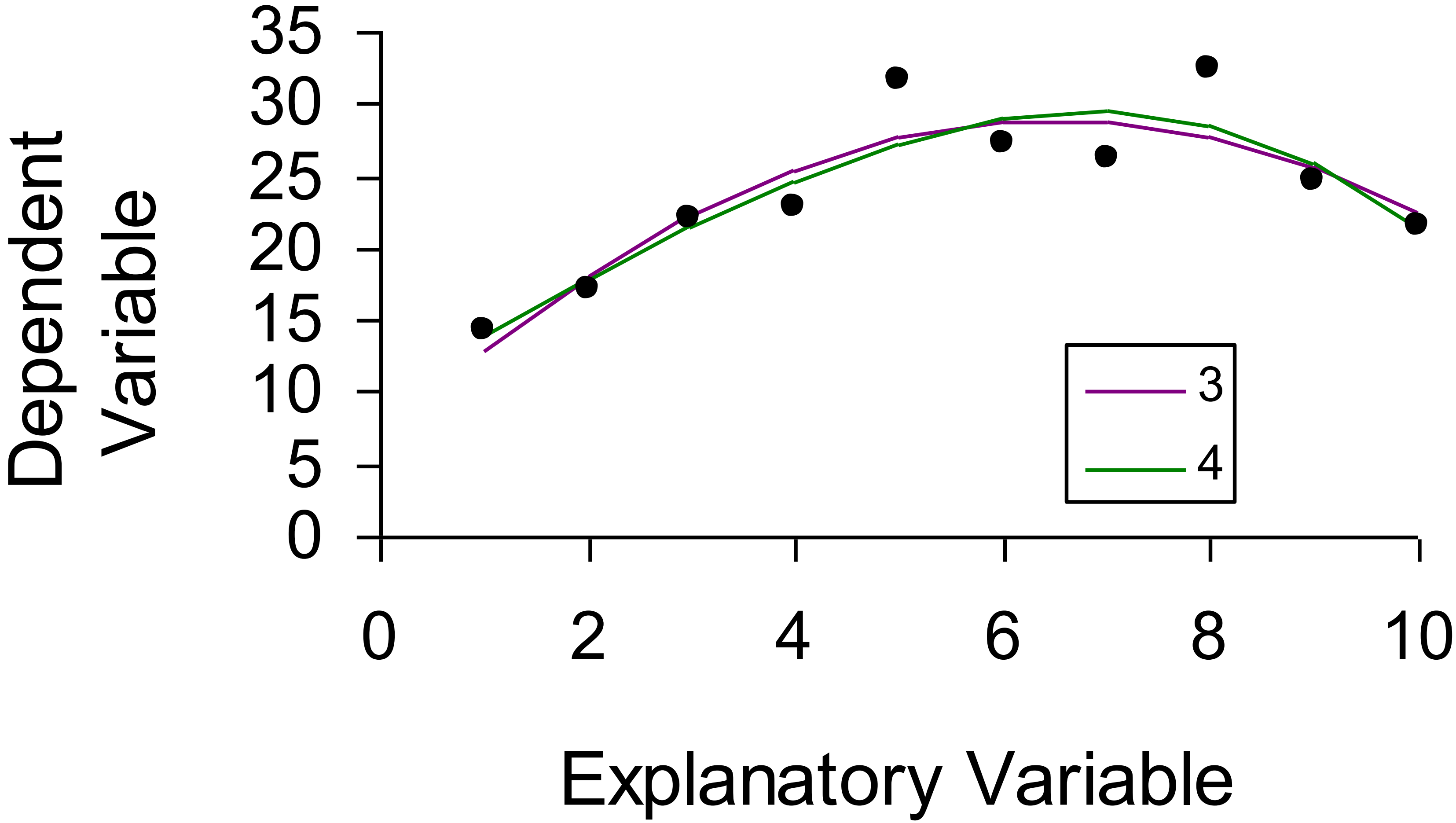

255 kB


758 kB

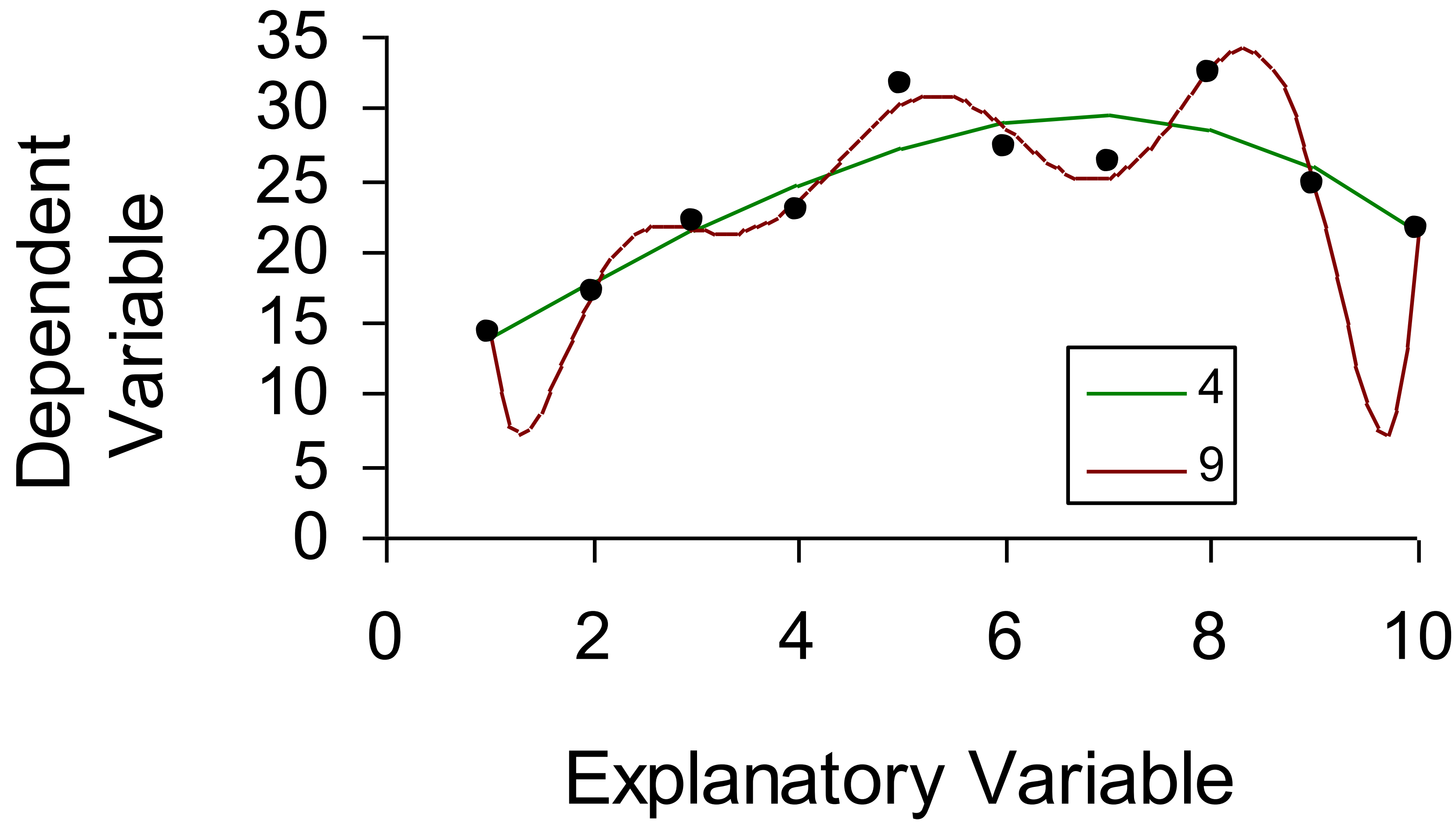

35
30
25
20
15
10
5
0
Dependent Variable
1
0
2
4
6
8
10
Explanatory Variable

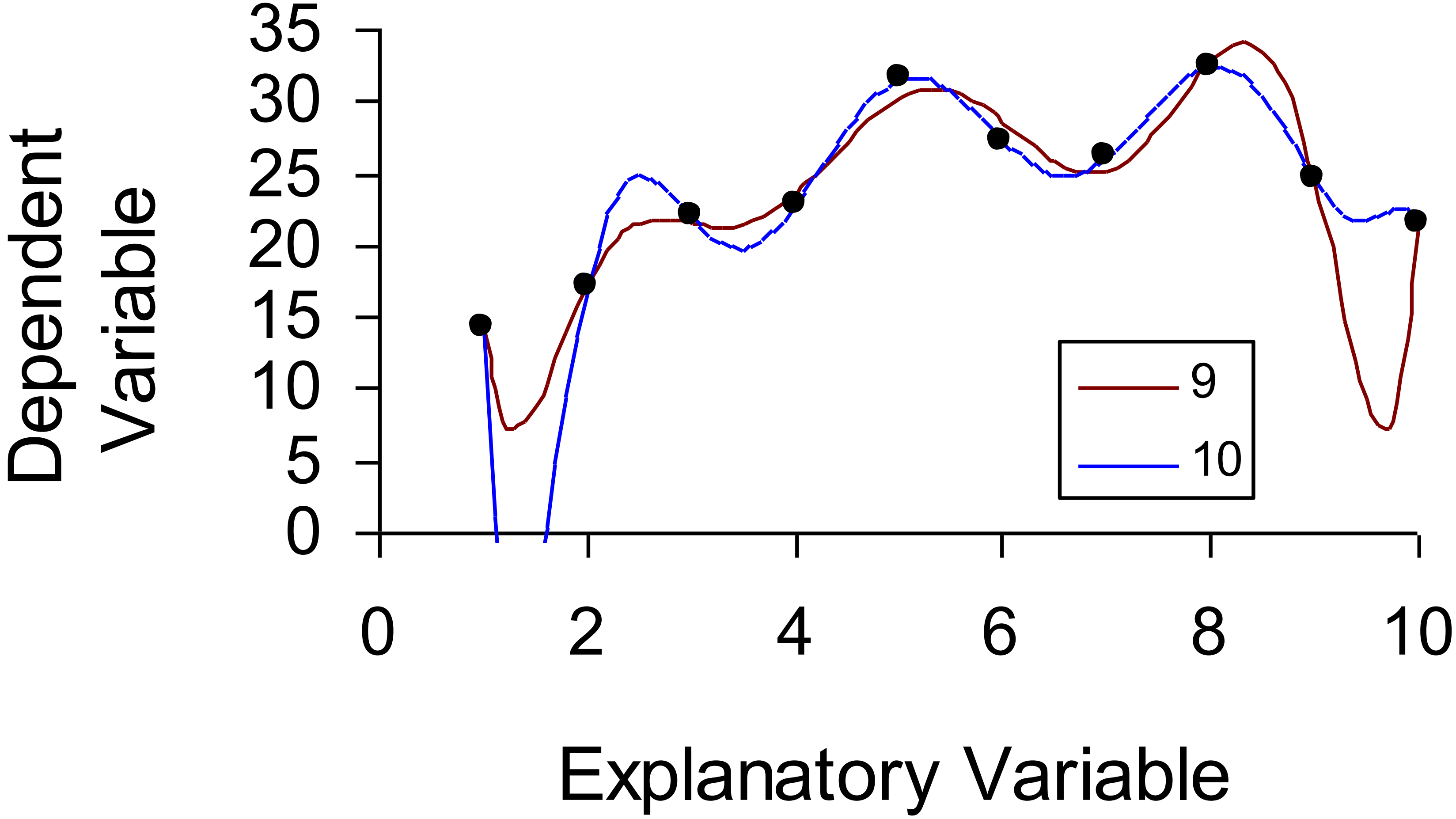

35
30
25
20
15
10
5
0
0
2
4
6
8
10
Dependent Variable
Explanatory Variable
1
2


Dependent Variable
35
30
25
20
15
10
5
0
0
2
4
6
8
10
Explanatory Variable
2
3



Dependent Variable
35
30
25
20
15
10
5
0
0
2
4
6
8
10
3
4
Explanatory Variable


Dependent Variable
35
30
25
20
15
10
5
0
0
2
4
6
8
10
Explanatory Variable
4
9



Dependent
Variable
35
30
25
20
15
10
5
0
0
2
4
6
8
10
Explanatory Variable
9
10

The principle of parsimony

- "… the smallest number of parameters for adequate representation of the data". (Box & Jenkins 1970)
- The bias versus variance trade-off.

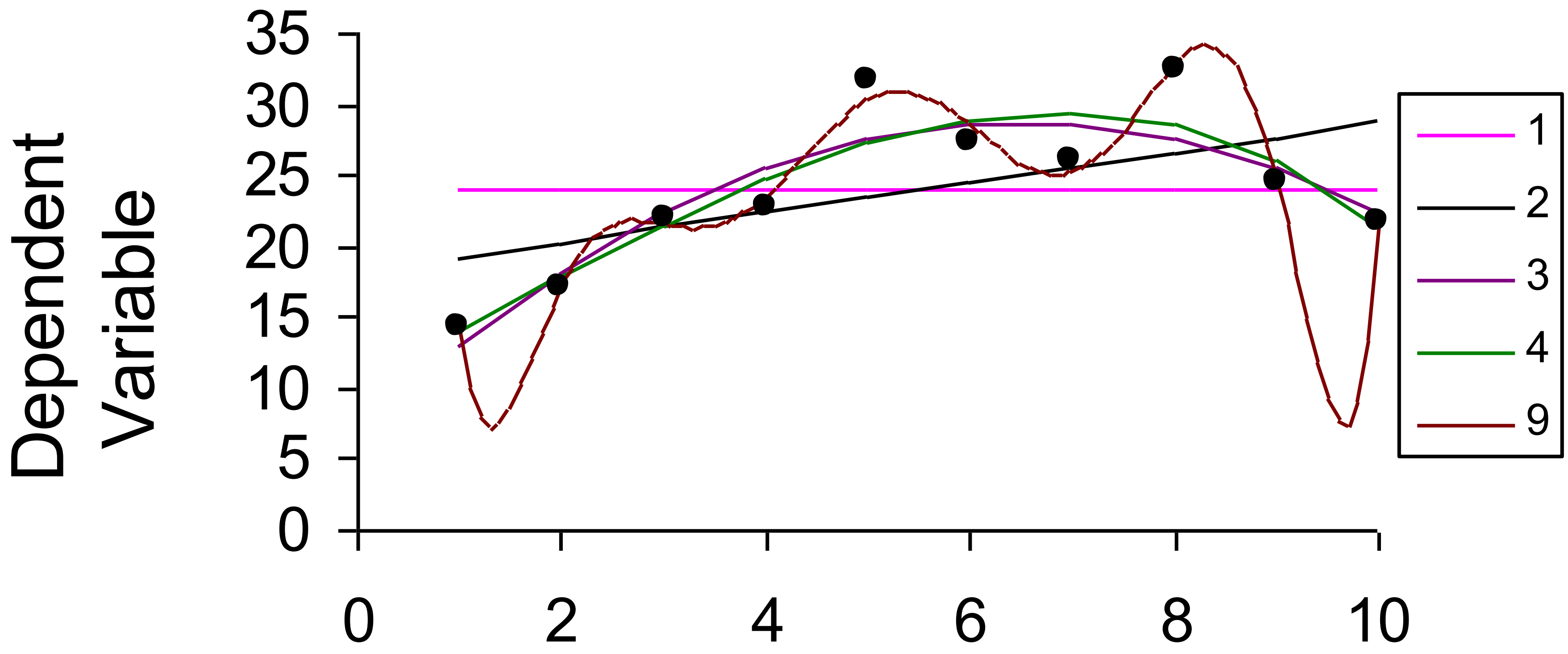
Choosing the best models

- Akaike Information Criterion
  - deviance + 2*k*
  - *k* = number of parameters being estimated
  - deviance = n*ln(MSE) + C (least squares & normal)
  - deviance = -2ln(Likelihood)

Interpreting AIC

ΔAIC	Degree of support
0-2	Substantial
4-7	Considerably less
>10	Essentially none

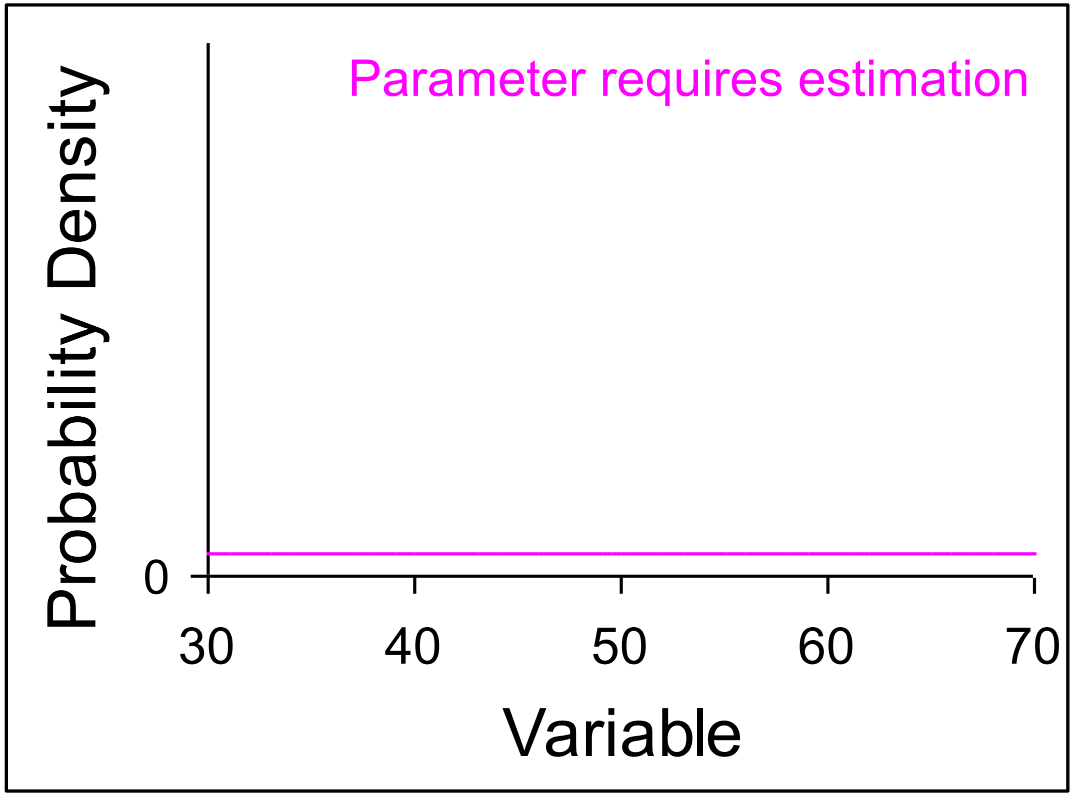
Approximation to a posterior model probability ~ $\exp[-\Delta AIC/2]$



	k	AIC
	1	58.9
	2	57.1
Best →	3	47.0
	4	48.0
	9	53.4

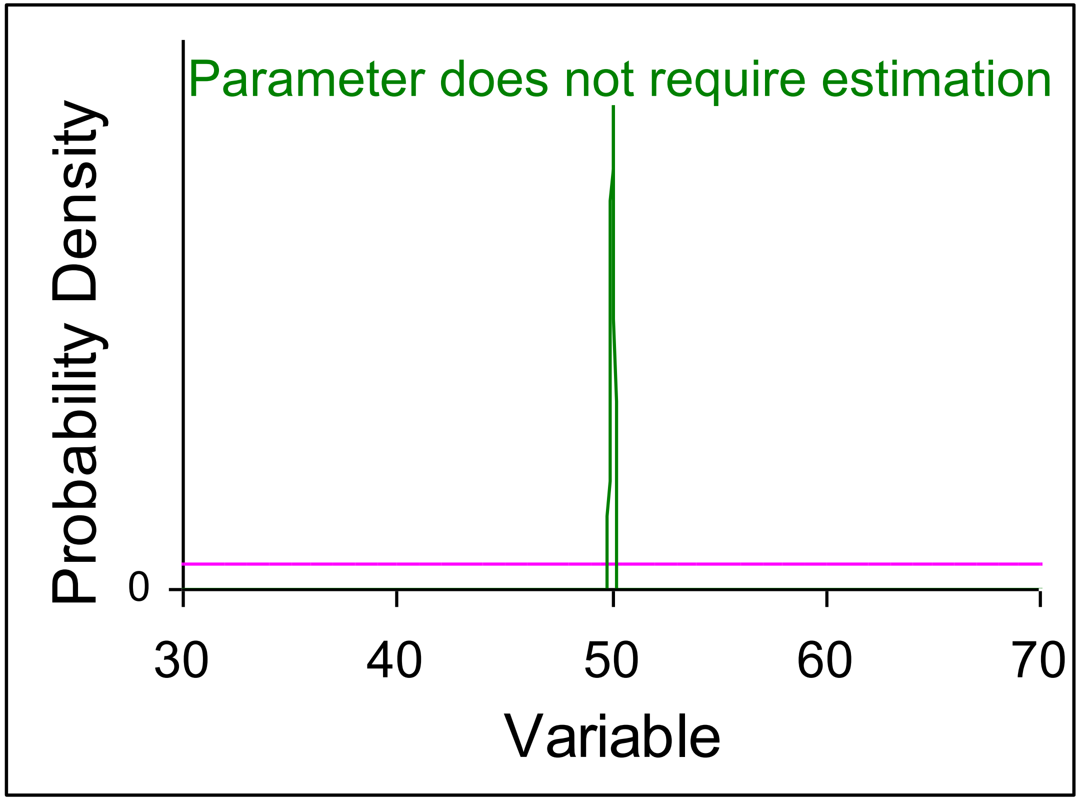
Not the full parameter…

uninformative prior



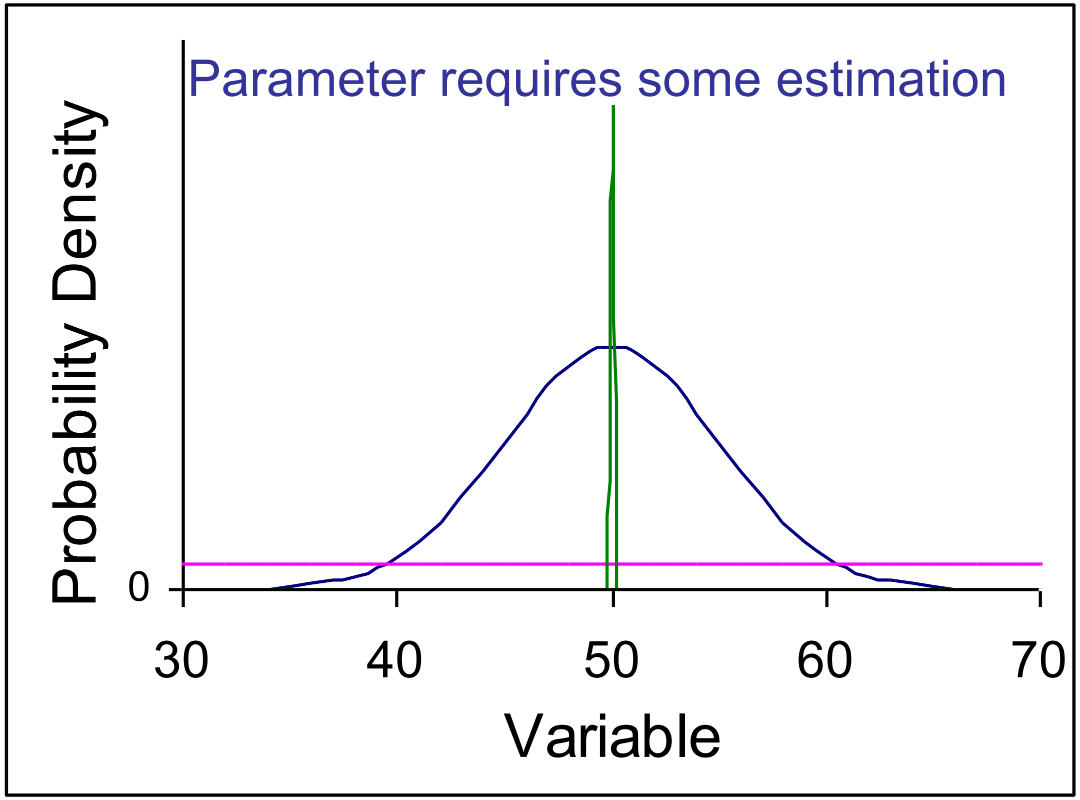
Not the full parameter…

extremely informative prior



Not the full parameter…

moderately informative prior



Choosing the best models

- Deviance Information Criterion
  - AIC but with *k* estimated
  - useful for models with random effects and priors
- k ~ mean(Dev) – Dev(at posterior mean)

Combine prior and posterior

Mean of posterior

$= (prec_{prior}*\mu_{prior} + prec_{data}*\mu_{data})/(prec_{data} + prec_{prior})$

$= 55.15$

Precision of posterior

$= prec_{prior} + prec_{data} = 1/var_{prior} + 1/var_{data}$

$= 0.0067$

$variance_{post} = 1/0.0067 = 149.3$