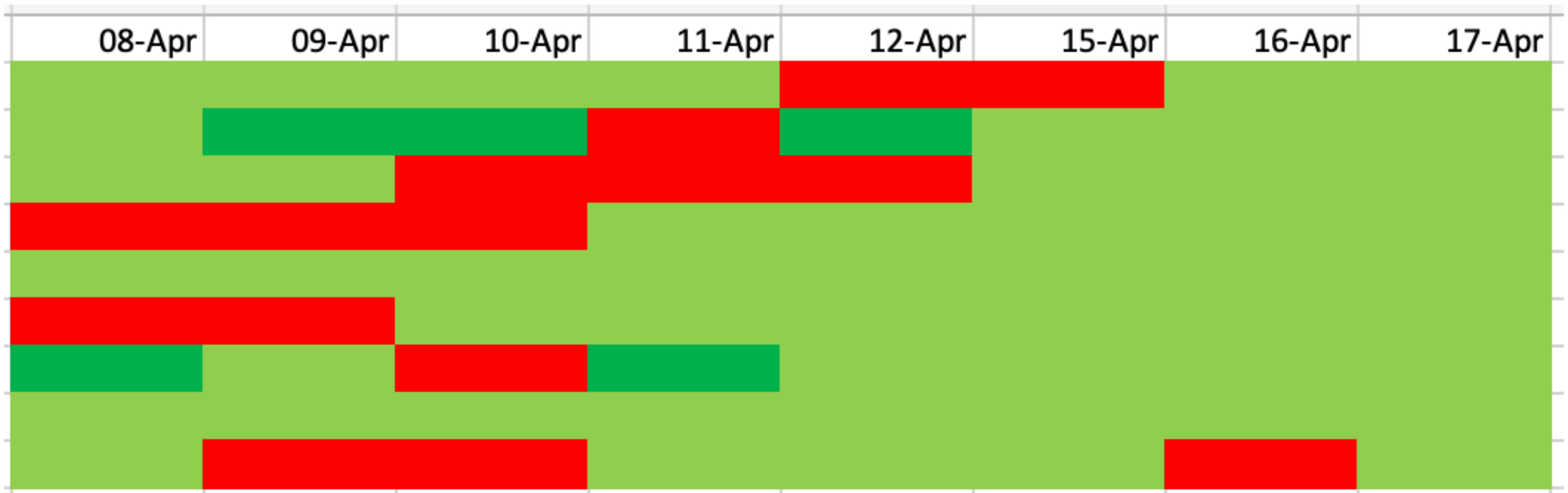
Last Day!

Final Presentations


08-Apr
09-Apr
10-Apr
11-Apr
12-Apr
15-Apr
16-Apr
17-Apr

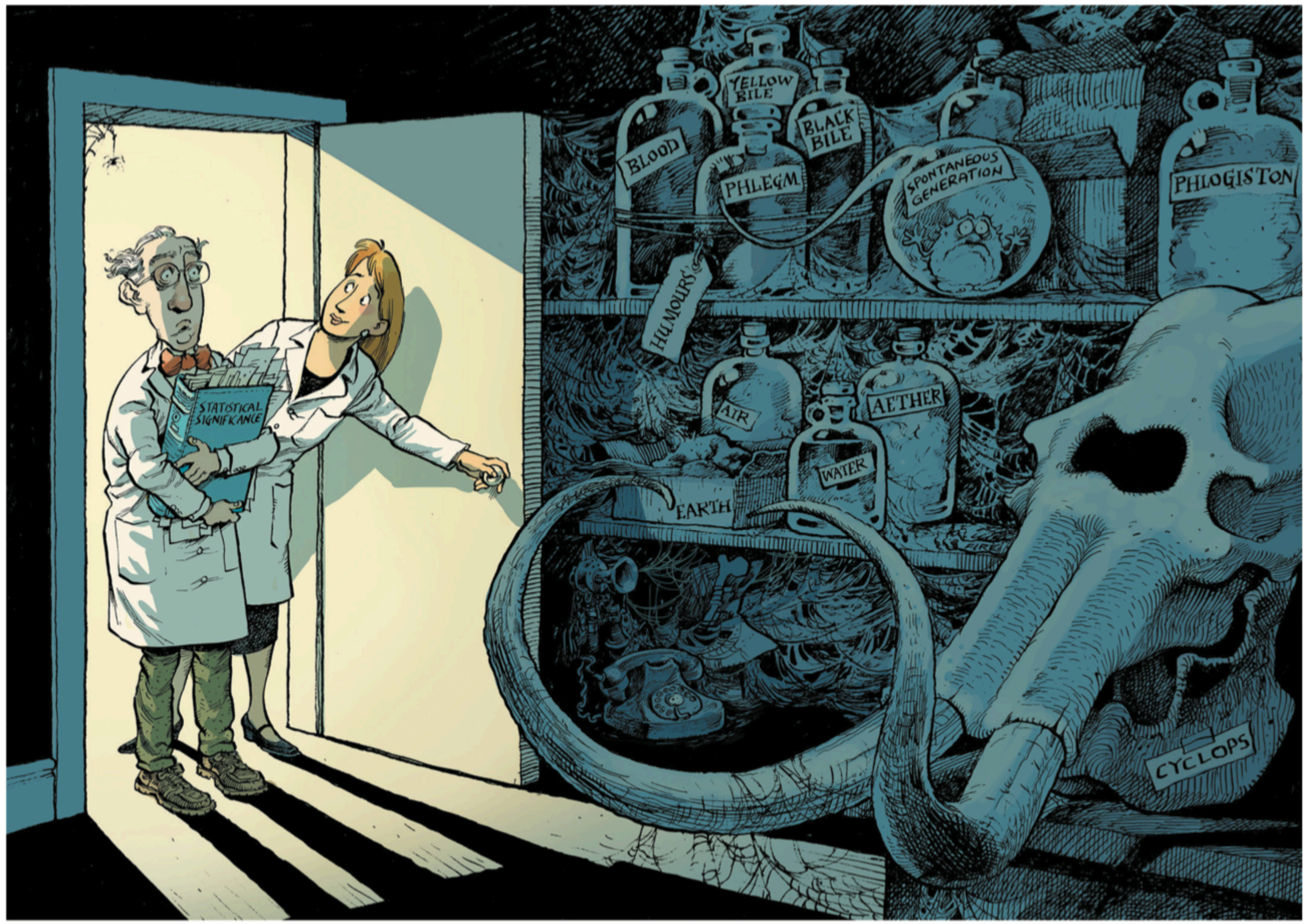
Final Presentations: Analysis

- Appropriate data visualization and interpretation prior to analyses
- Appropriate model development and choice of priors
- Appropriate evaluation of MCMC performance
- Appropriate visualization and interpretation of posteriors
- Appropriate check of model
- Appropriate evaluation of parameters of interest
- Appropriate conclusions

Review
& Recap

P-values and NHST

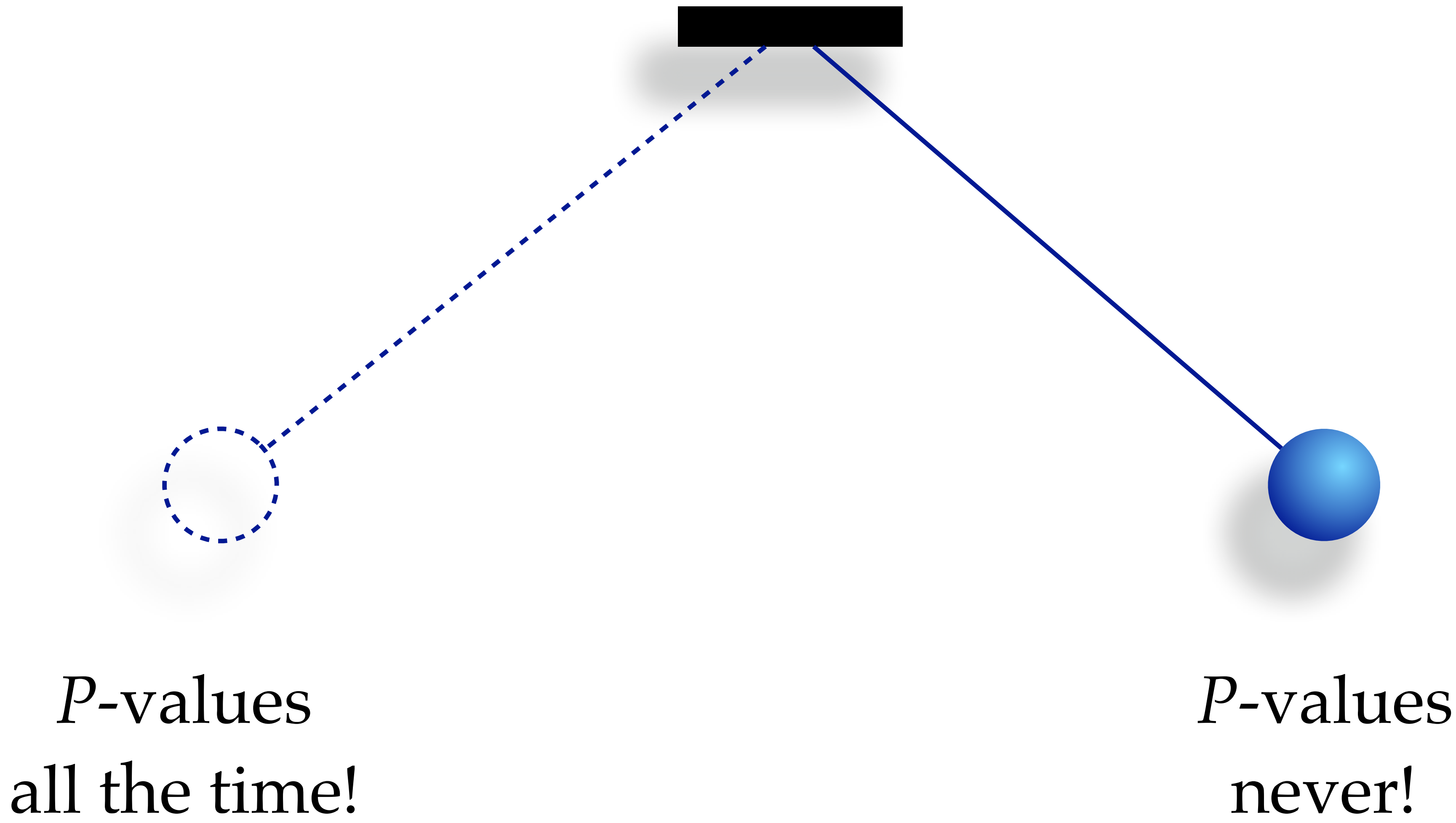
ILLUSTRATION BY DAVID PARKINS



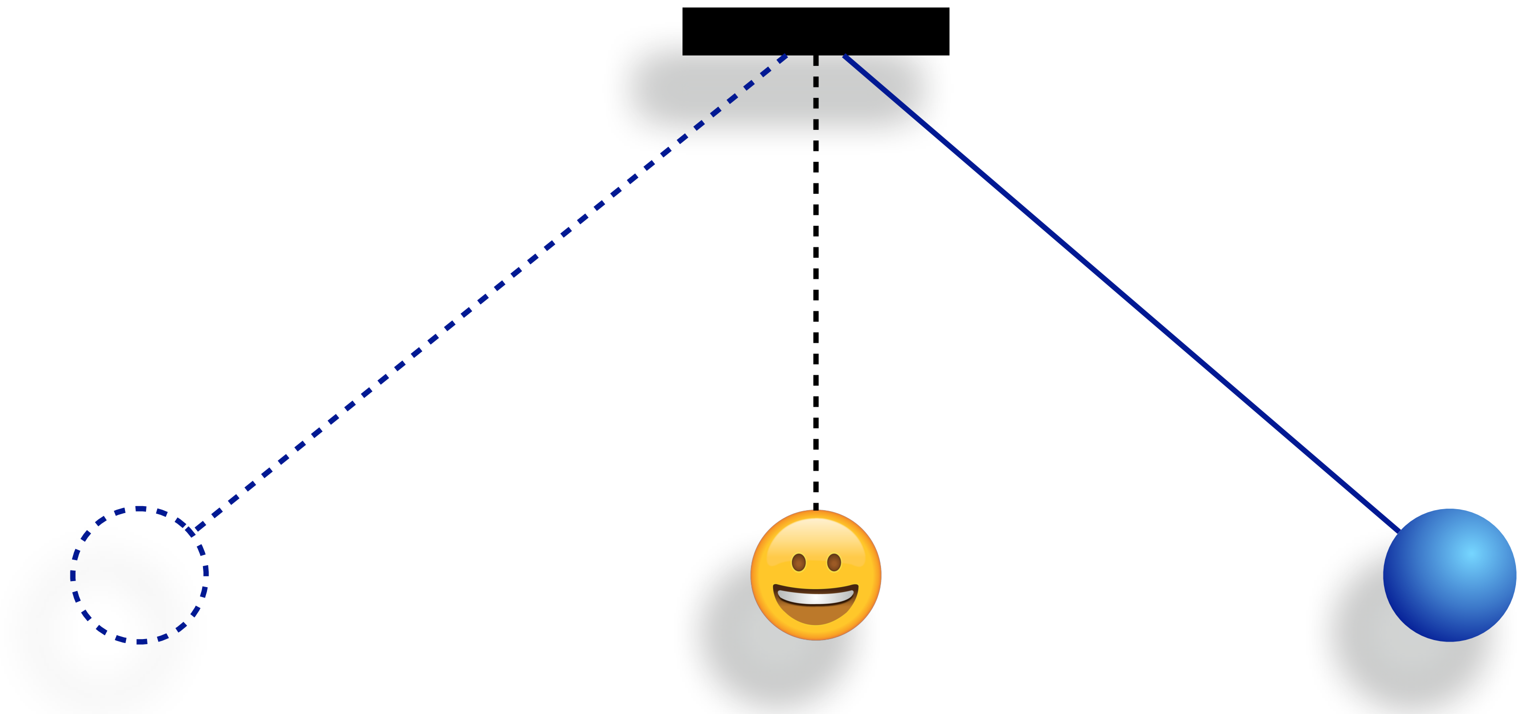
Retire statistical significance

Valentin Amrhein, **Sander Greenland**, **Blake McShane** and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

P-values and NHST



P-values and NHST



P-values
all the time!

P-values
when appropriate,
and with appropriate
interpretation

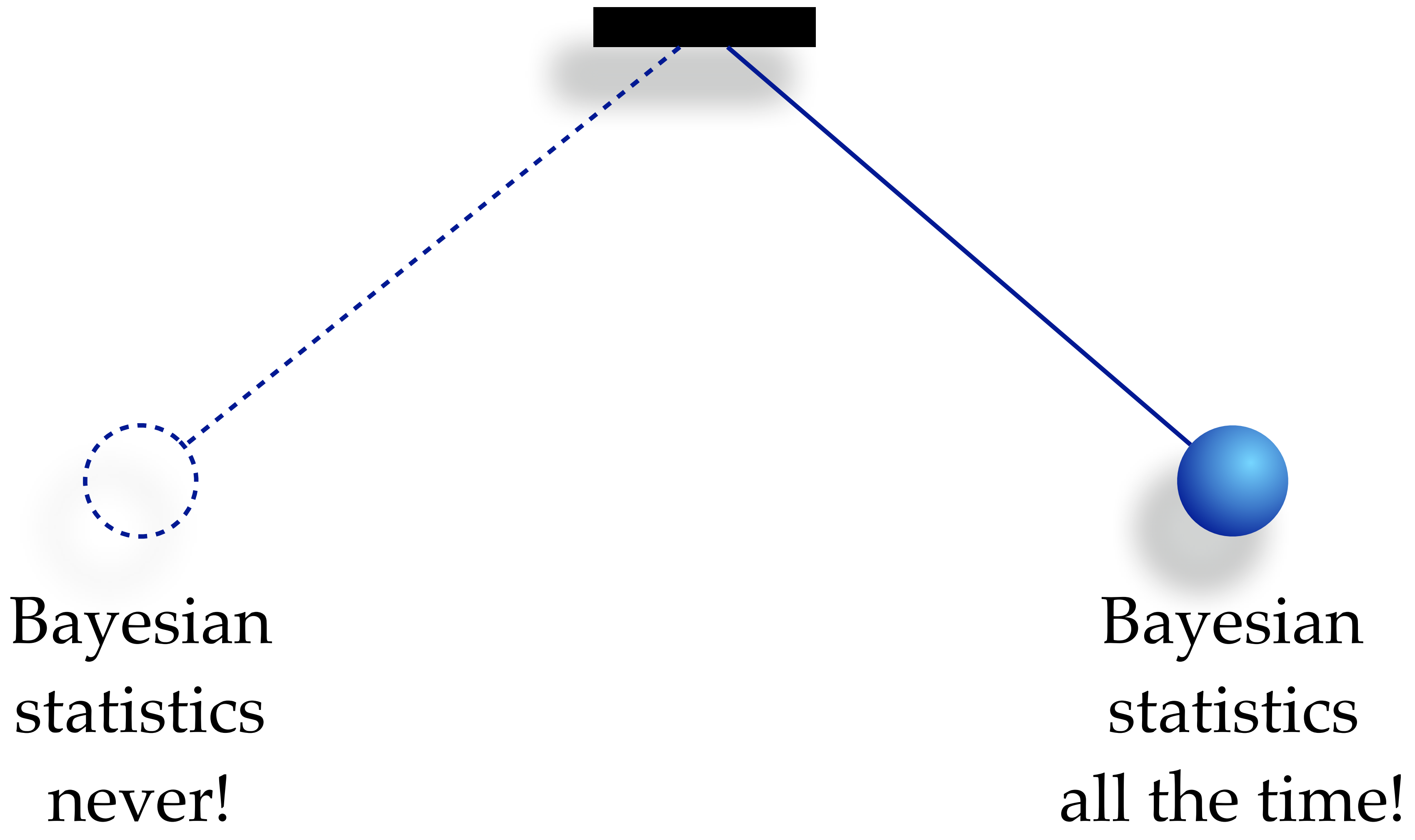
P-values
never!

Bayesian Statistics

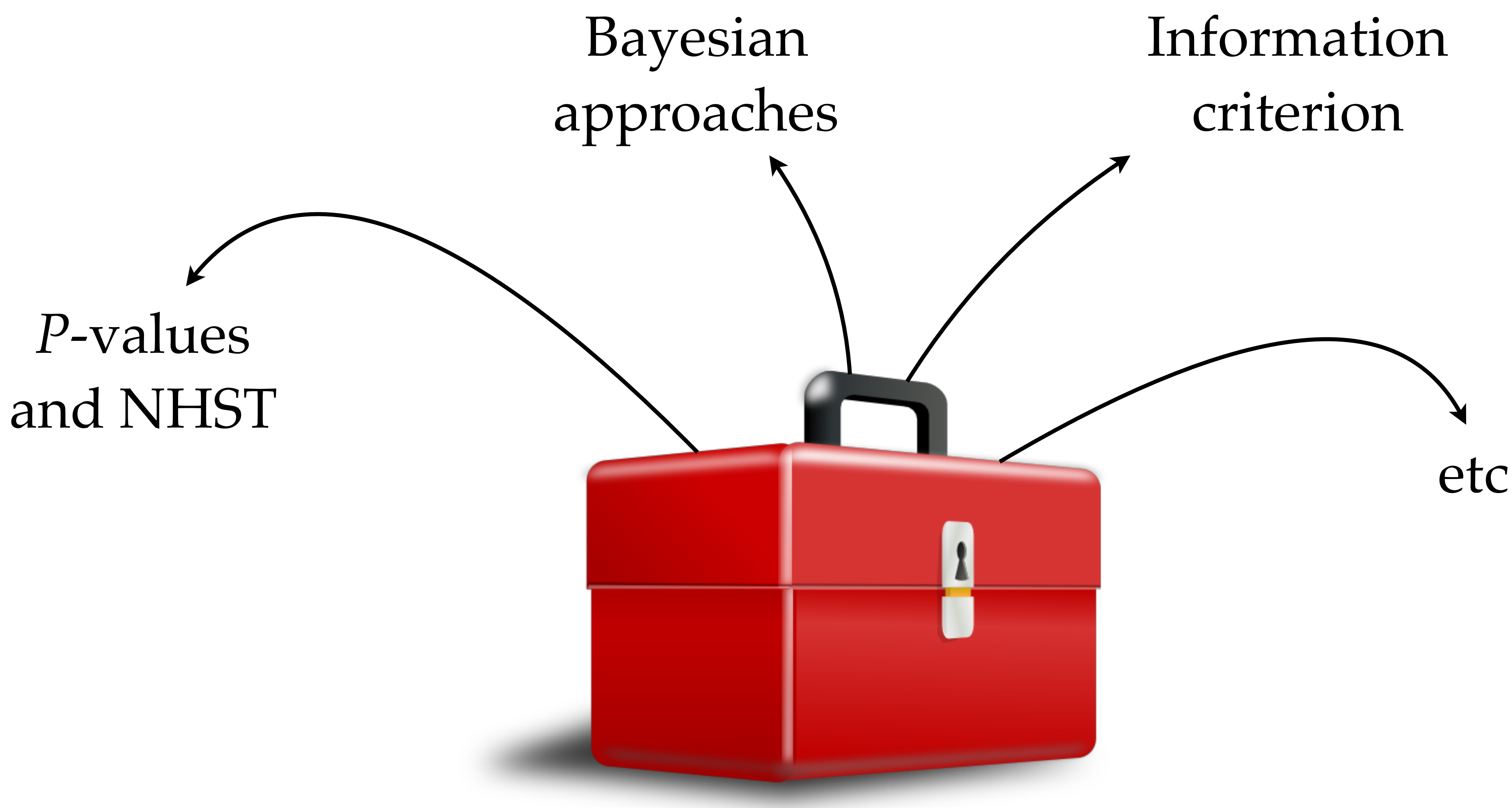
- An explicit way to allow data to change our beliefs about the world - the whole point doing science!!!
  - Priors represent prior beliefs about hypotheses
  - Posteriors represent the revision of the credibility of those hypotheses based on the data

Bayesian Statistics

- Not a panacea either!!

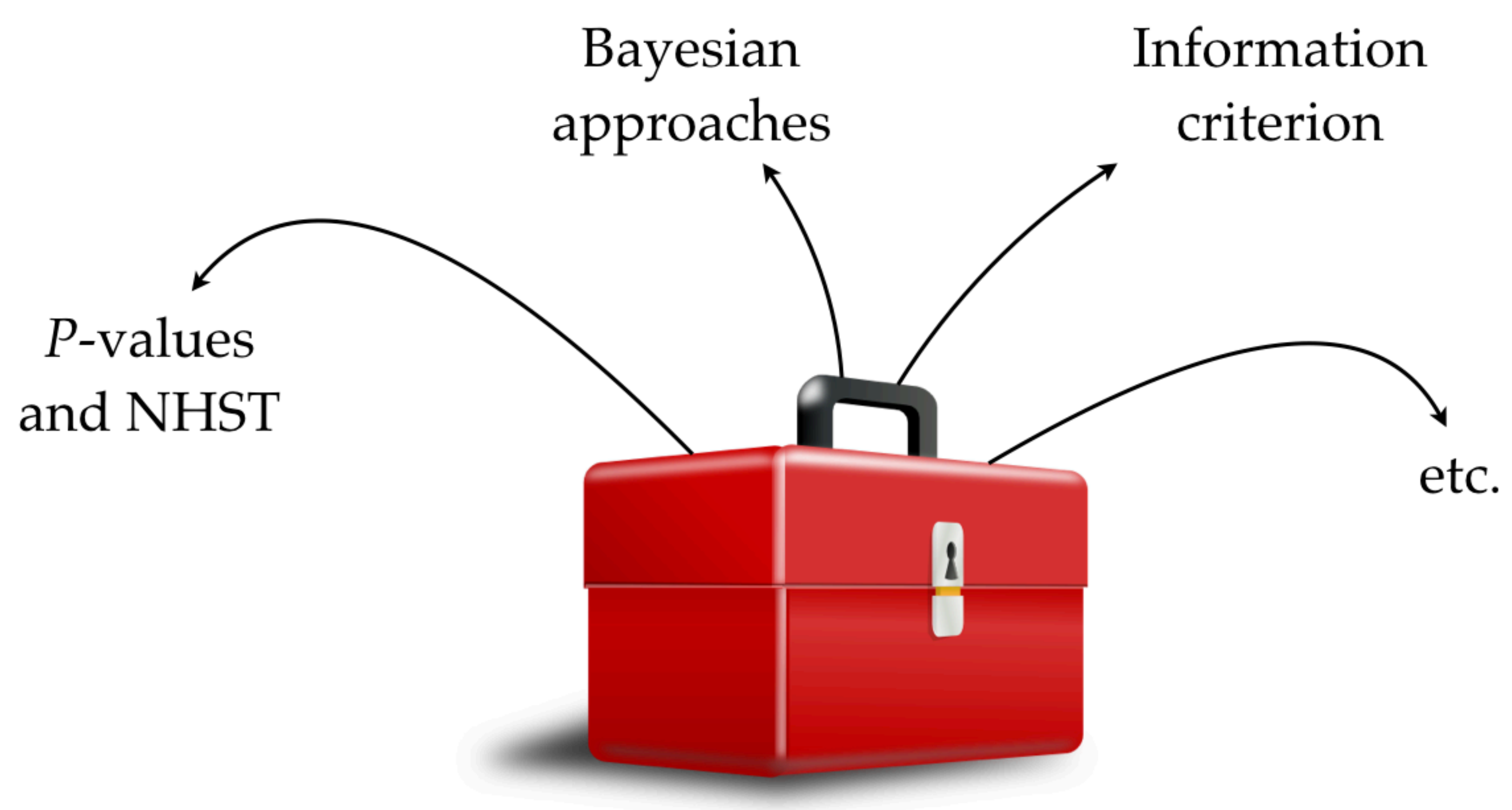


Statistical Toolbox

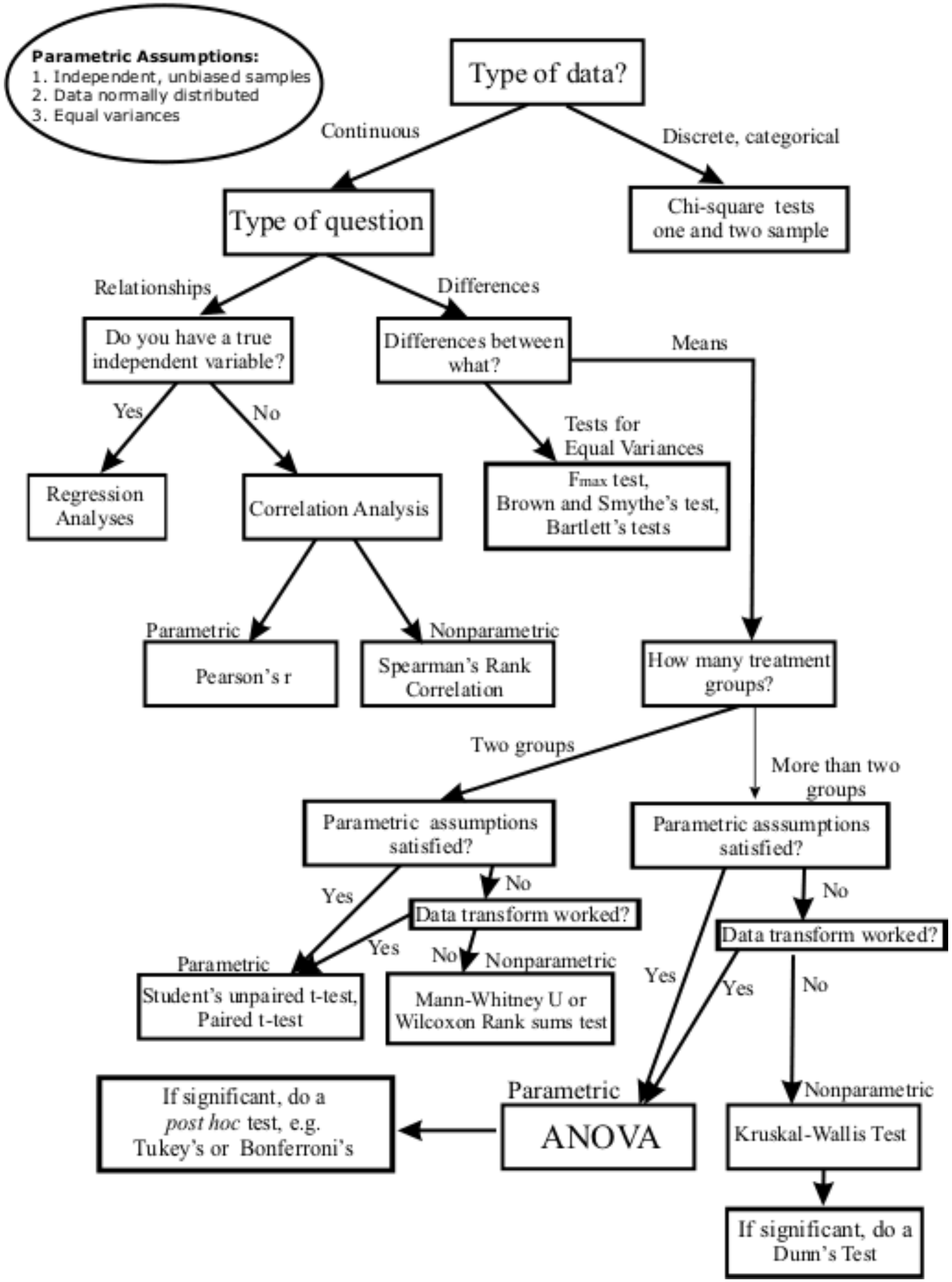


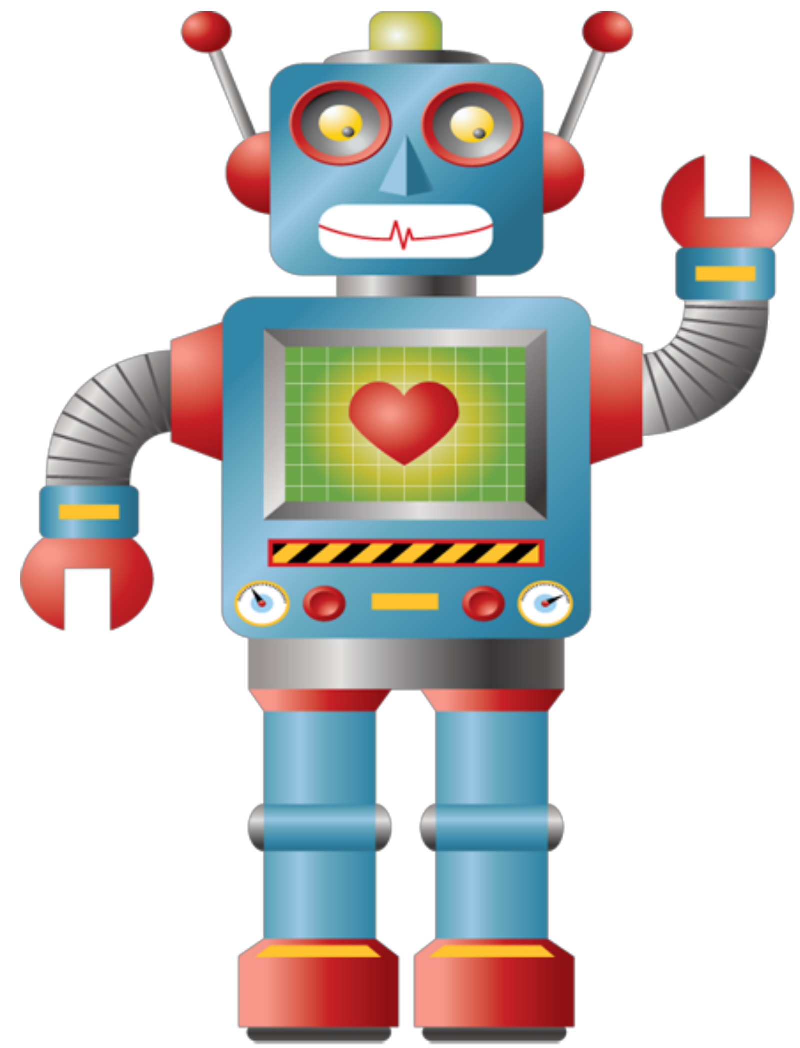
Statistical Toolbox

- Statistics should not be the endpoint itself
  - Is a tool to help you interpret what your data tell you about the world



Flow Chart for Selecting Commonly Used Statistical Tests





Bayesian Statistics

- Build your own robot!
  - Will often be more appropriate than a canned one

Bayesian Statistics

- Build your own robot!
  - Will often be more appropriate than a canned one
- But…always be sceptical of your robot!
  - Remember, it is an oracle

Bayesian Statistics

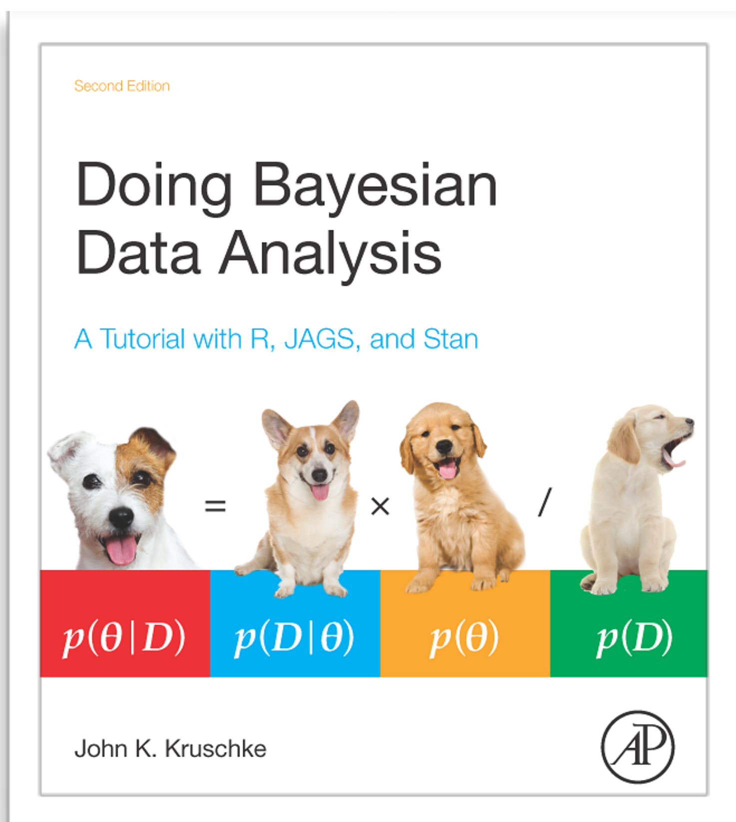
- When in doubt (or actually all the time) simulate data with known effects, and make sure model recovers appropriate estimates

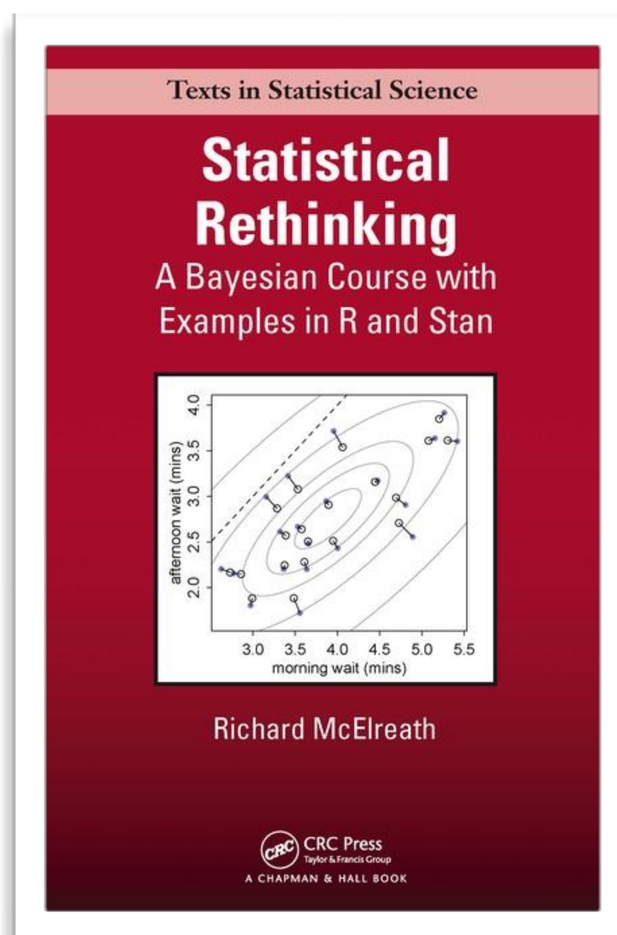
Bayesian Statistics

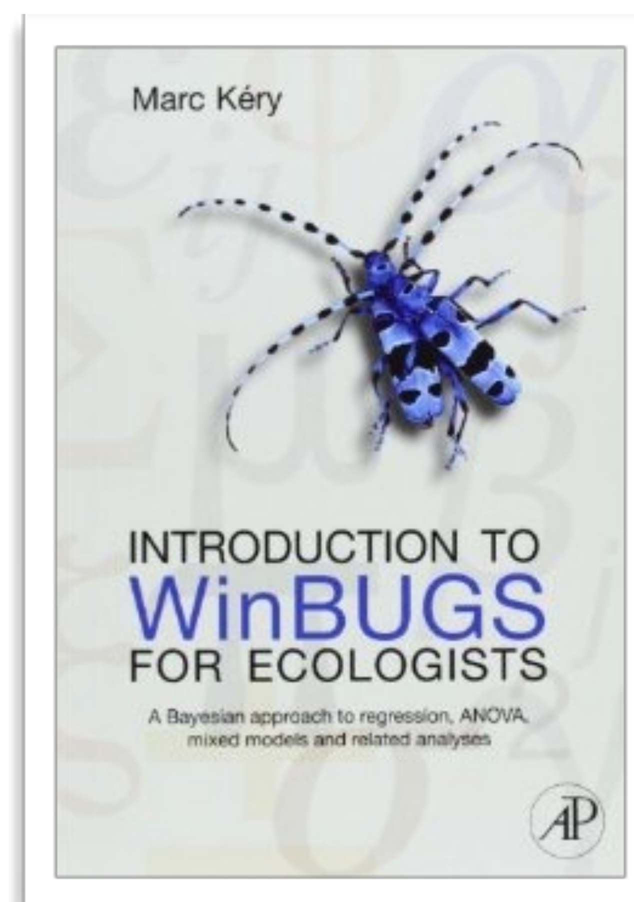
- Just the beginning
- Must continue your training…

Bayesian Statistics

- Just the beginning
- Must continue your training…

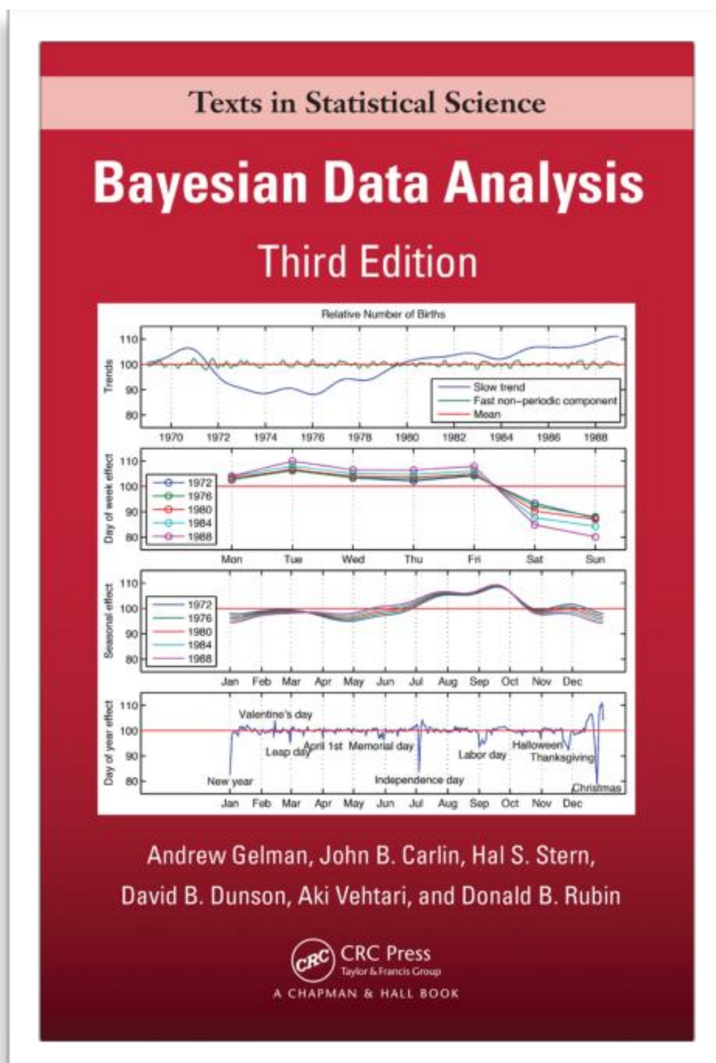


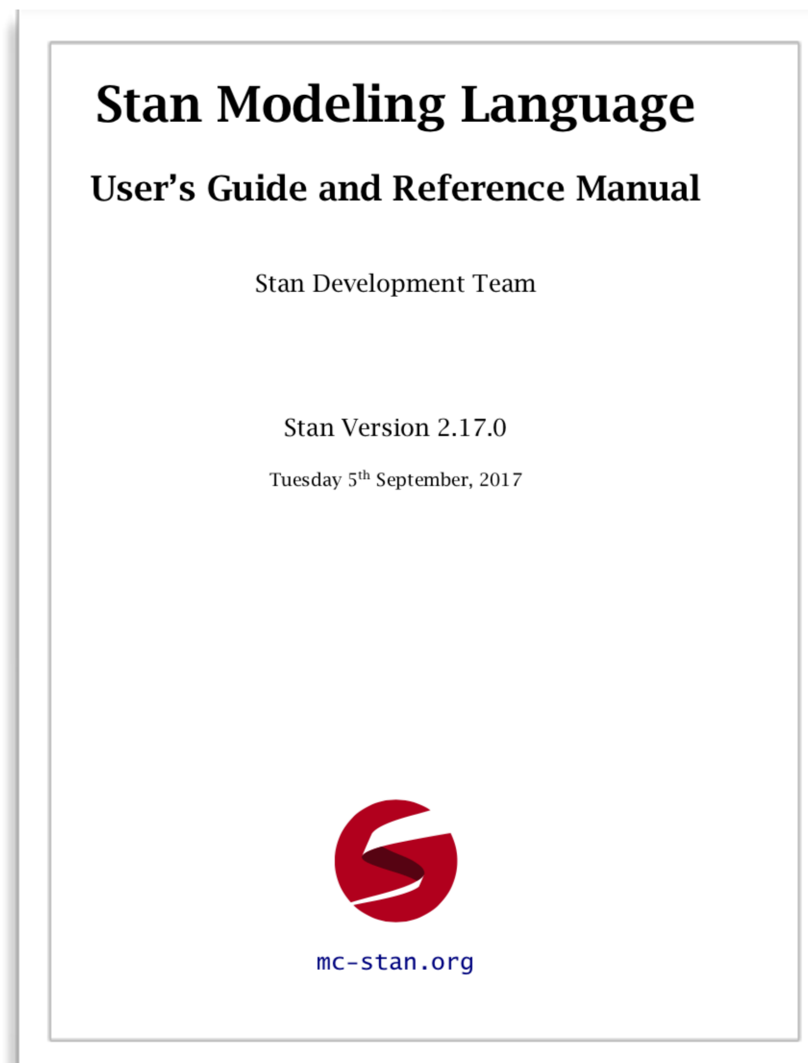




Bayesian Statistics

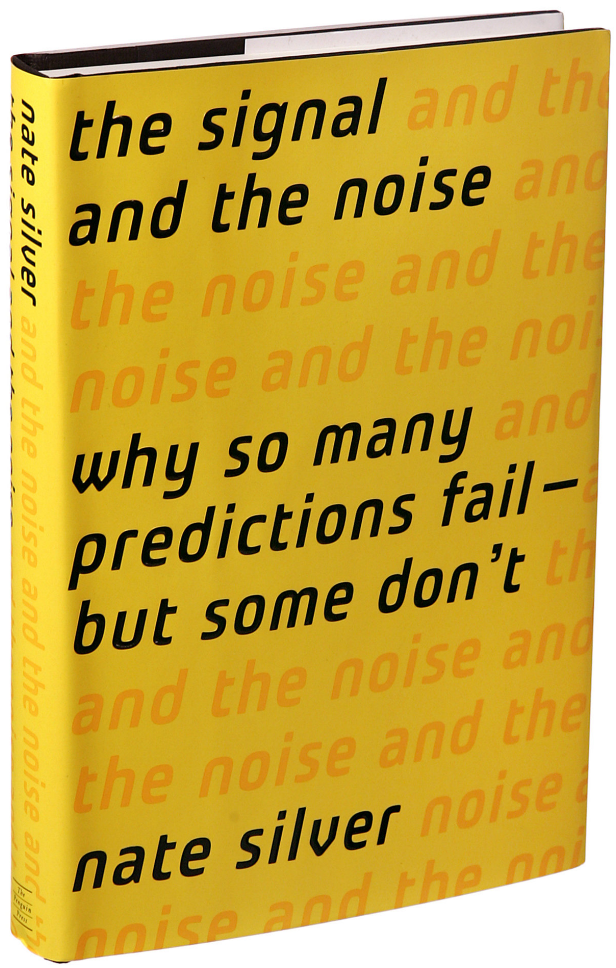
- Just the beginning
- Must continue your training…

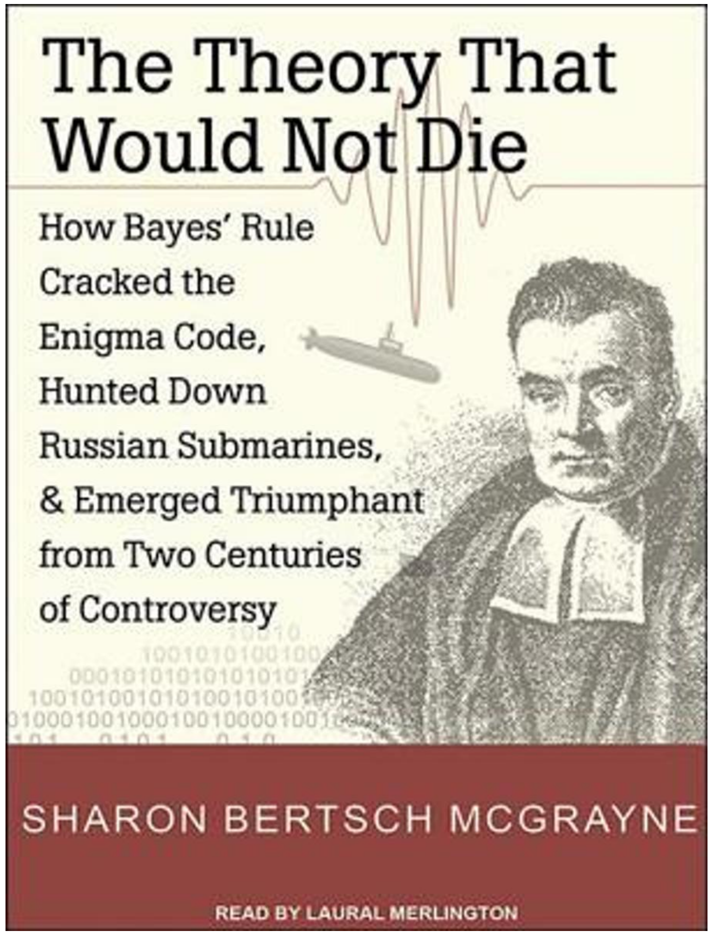




Bayesian Statistics

- Just the beginning
- Must continue your training…





Last Few Models (briefly):

Full presentations, data, and code are online

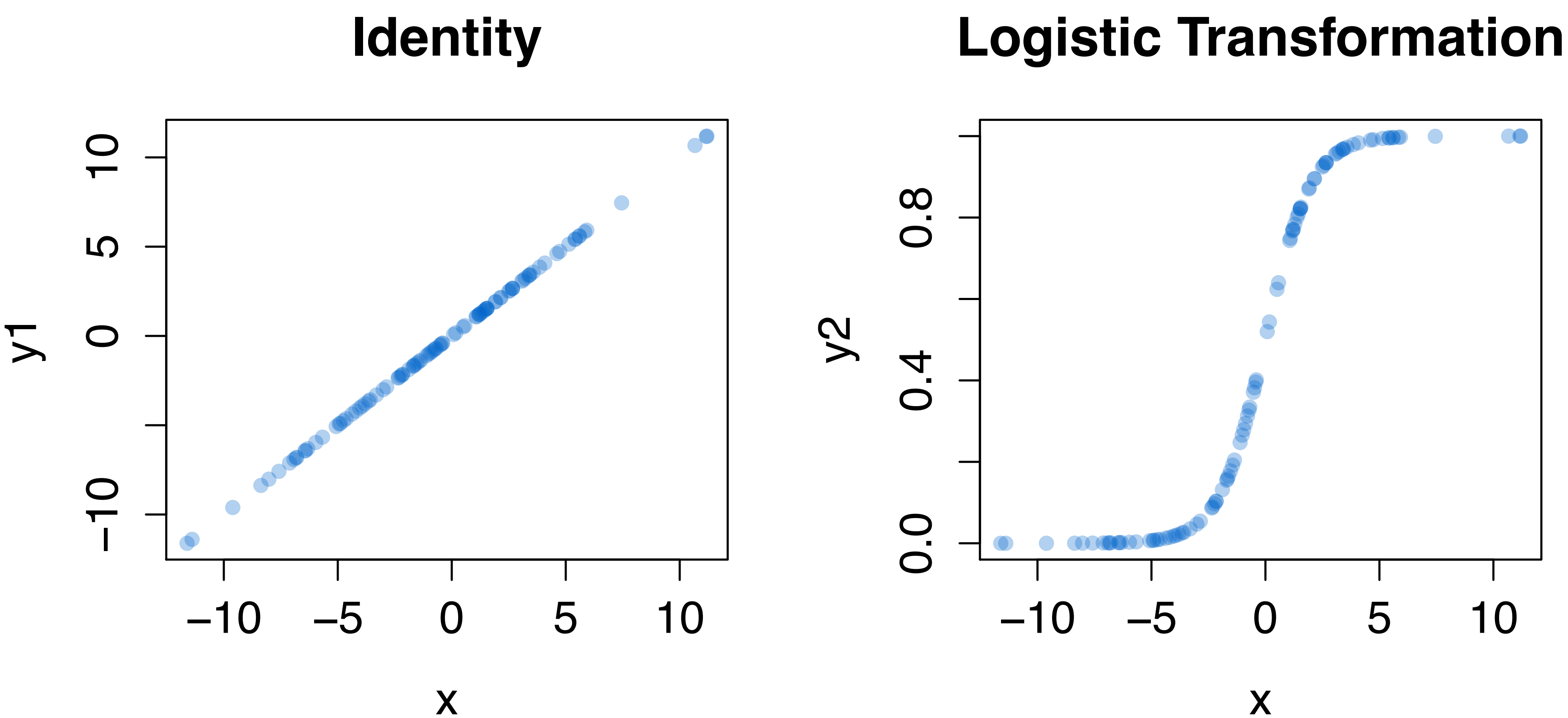
Logistic Regression

Logistic Predictor Variable

- Outcome is 0 or 1
- Have to squish (transform) predictors to this scale
  - Use the *logit* function

$$\text{logit}(mu) = 1 \;/\; (1 + \exp(-mu))$$

Logistic Predictor Variable



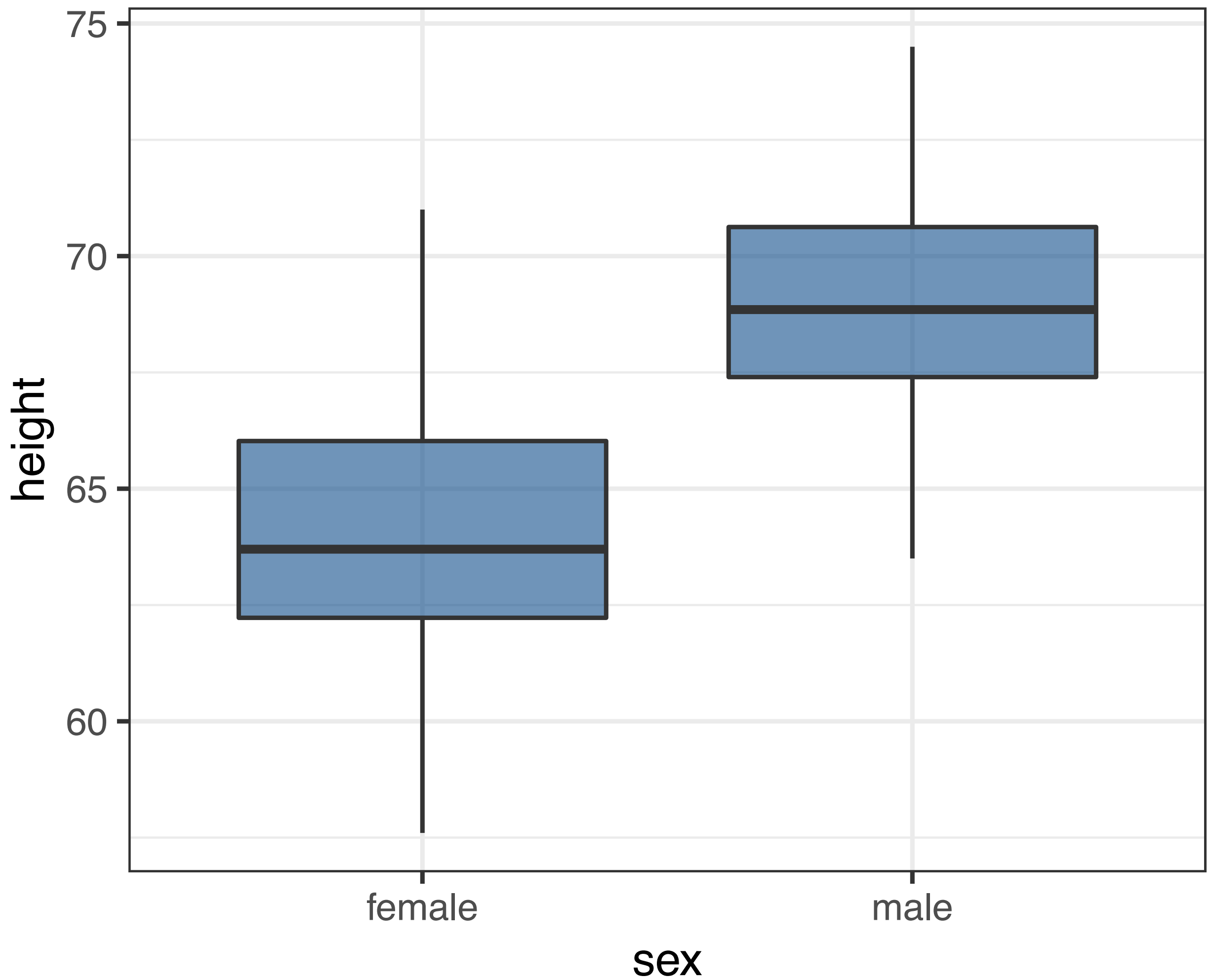
Data

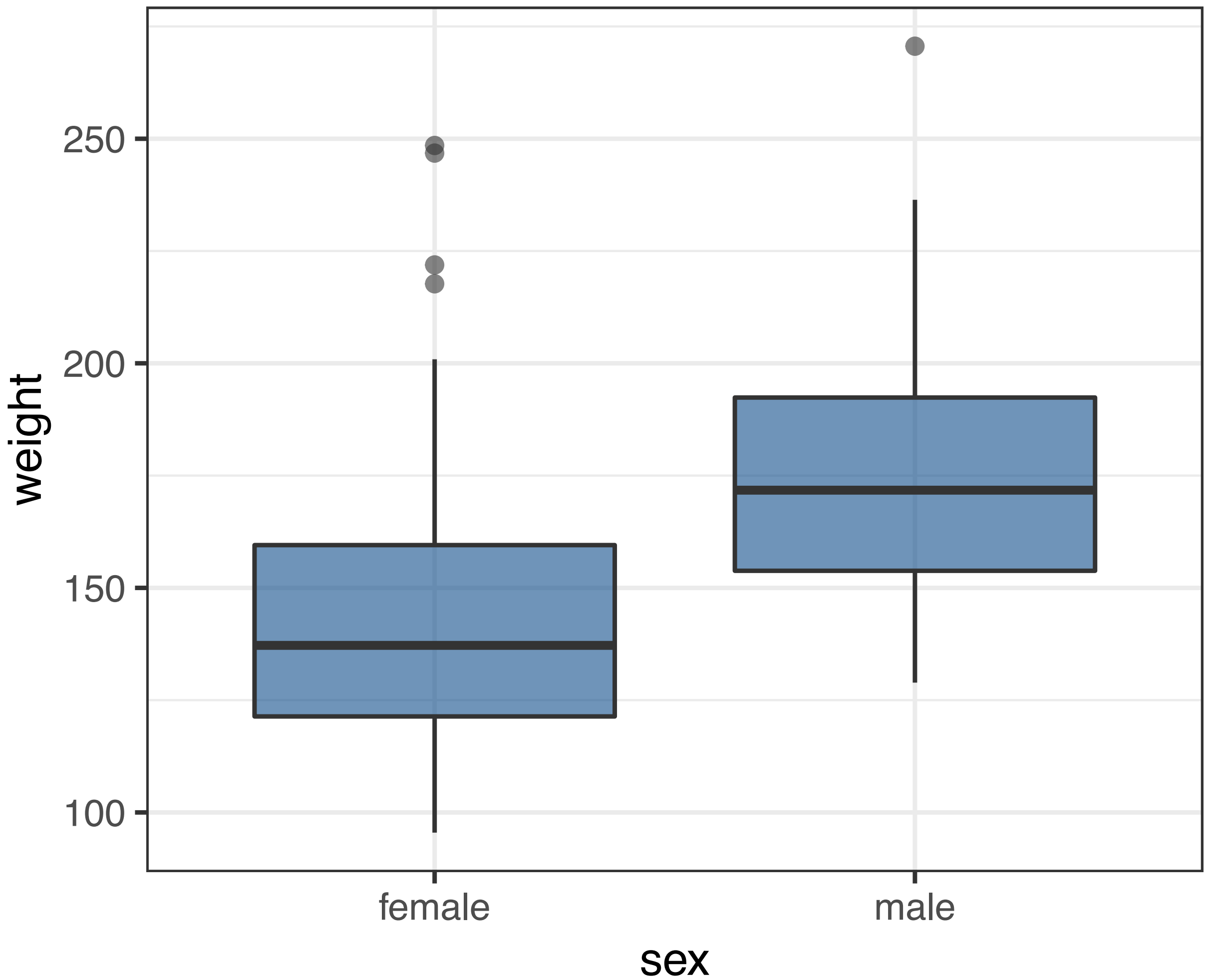
- Modified height and weight data from before (but also including sex now)

htwt

	male	height	weight
1	0	63.2	168.7
2	0	68.7	169.8
3	0	64.8	176.6
4	0	67.9	246.8
5	1	68.9	151.6
6	1	67.8	158.0
7	1	68.2	168.6
8	0	64.8	137.2
9	1	64.3	177.0
10	0	64.7	128.0
11	1	66.9	168.4
12	1	66.9	136.2
13	1	67.1	160.3
14	1	70.2	233.9
15	1	67.4	171.7

Data


75
70
65
60
height
female
male
sex


250
200
150
100
weight
female
male
sex

Model

$$y = \text{bernoulli\_logit}(mu)$$

$$mu = \beta_0 + \beta_1 height + \beta_2 weight$$

Model

$$y = \text{bernoulli\_logit}(mu)$$

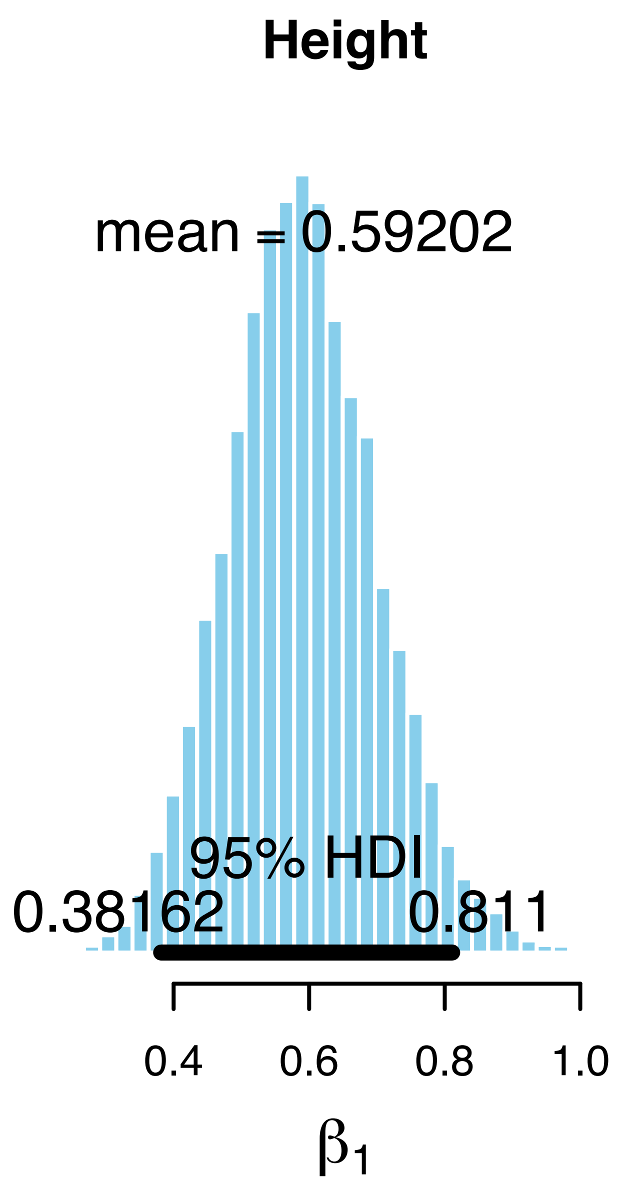
$$mu = \underline{\beta_0 + \beta_1 height + \beta_2 weight}$$

The “black box” into which we can put any equations that we have dealt with before (or more)

Model

```
model {
    // Definitions
    vector[N] mu;

    // Likelihood
    for (i in 1:N) {
      mu[i] = b0 + (b1 * height[i]) + (b2 * weight[i]);
      y[i] ~ bernoulli_logit(mu[i]);
    }

    // Priors
    b0 ~ normal(0, 1);
    b1 ~ normal(0, 1);
    b2 ~ normal(0, 1);
  }
```

Interpreting Posteriors



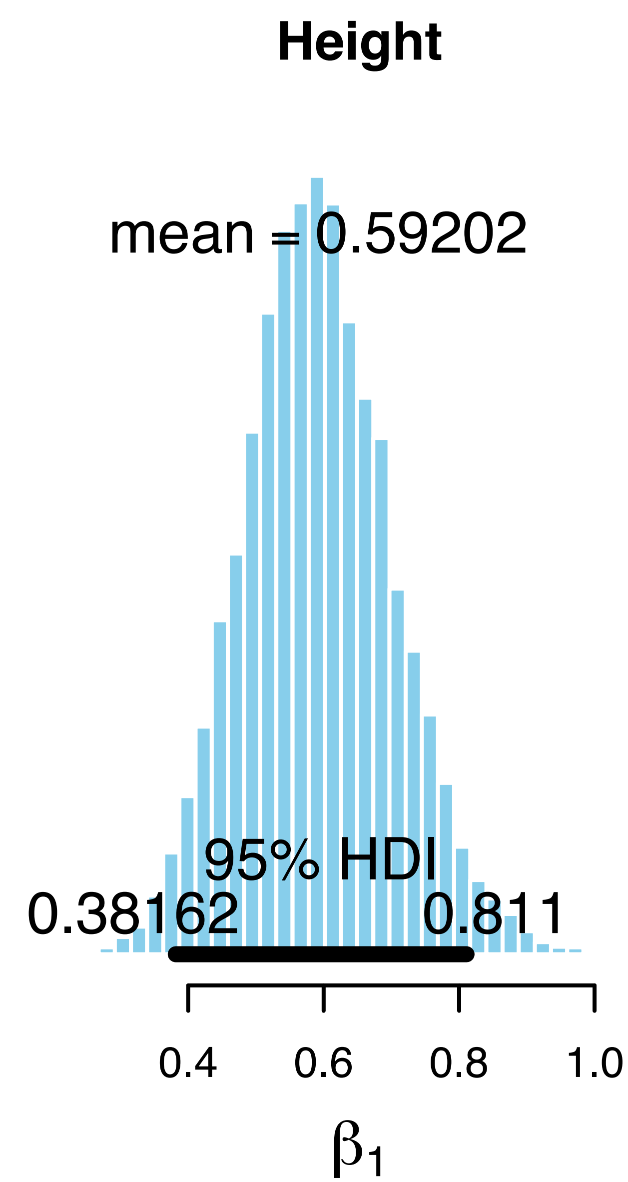
what does this mean?

Interpreting Posteriors

- Coefficients of logistic regression tell us about the log odds

$$\text{logit}(mu) = \log \frac{p(y=1)}{p(y=0)}$$

- When x_i increases by 1 unit, the log odds increase β_i units



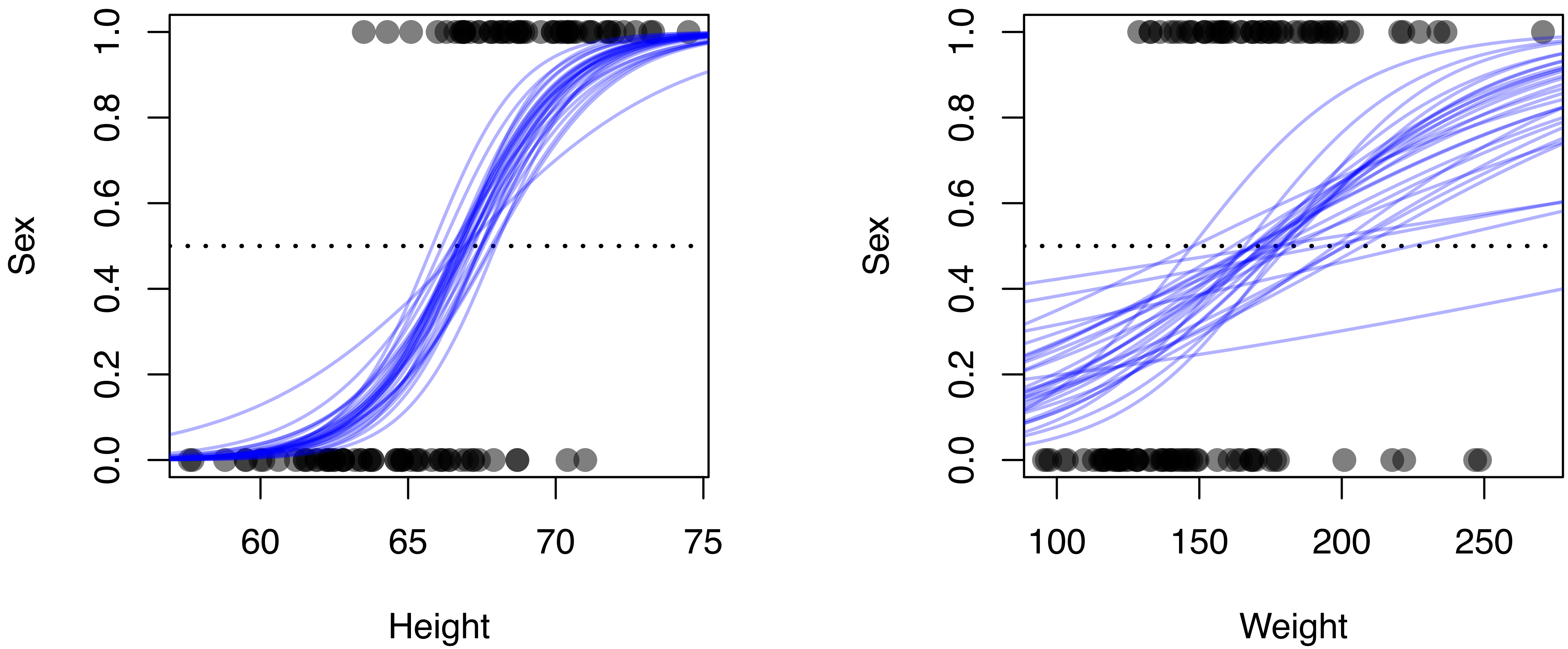
- When height increases by 1 inch, the log odds of being a male increases by ~0.592

Assessing Model Fit
Single parameters (height)

- Can ask, "How does probability of being male change as a function of height, for someone of average weight*"?
  - Evaluate effects of one parameter, while holding the others constant

*Graph will differ if you use different weights, but this will give you a general idea of fit

Assessing Model Fit



Ordinal Regression

Goals and General Idea

Goals

When would we use this type of analysis?

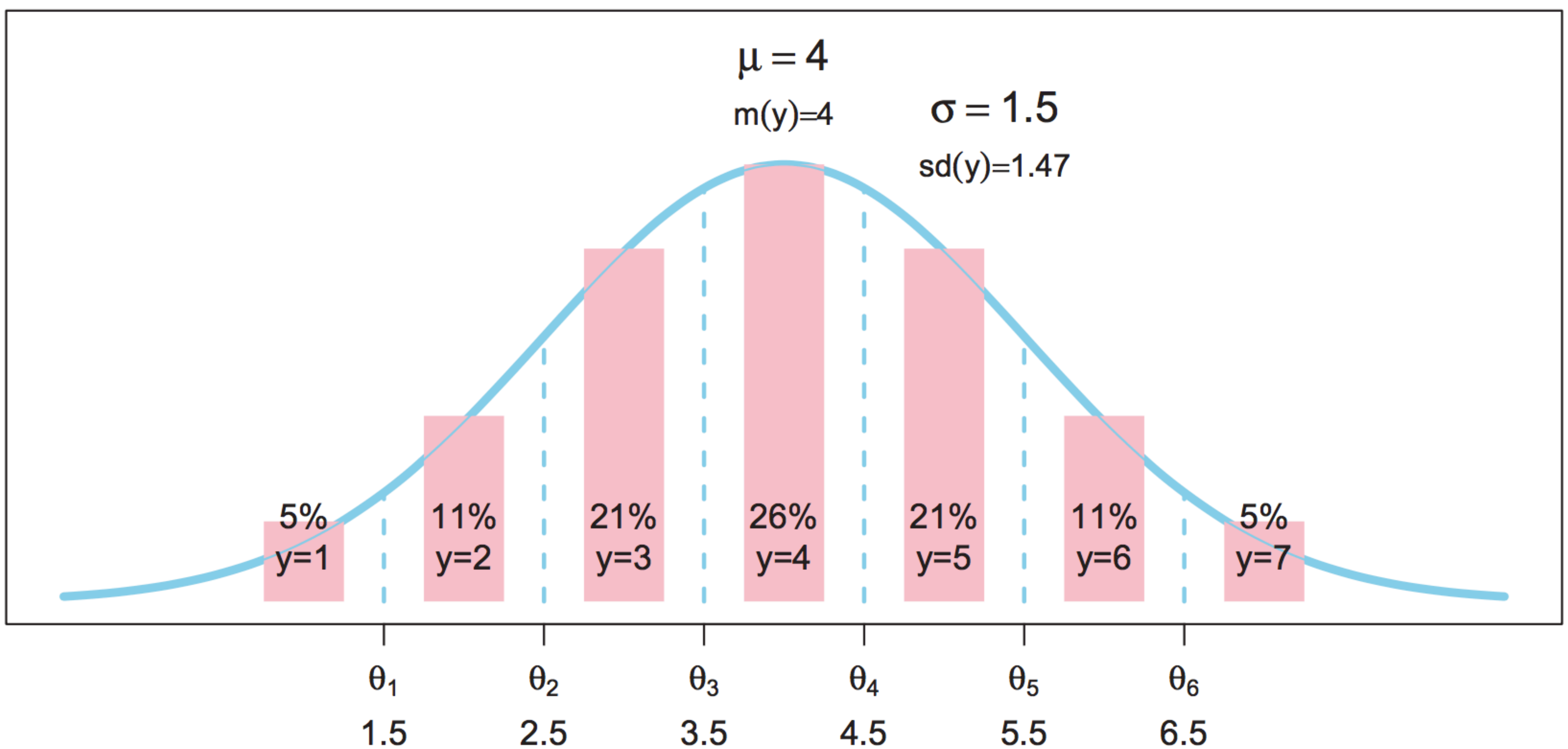
- When the predicted variable is ordinal!
  - Places in a race (1st, 2nd, 3rd, etc.)
  - Surveys on a Likert scale (5 = strongly agree, 4 = agree, 3 = neutral 2 = disagree, 1 = strongly disagree)
  - Scaled responses (good, mediocre, bad)
  - etc.

Characteristics

- Know order, but not necessarily equally spaced
  - How much do you like fish (1-hate to 5-love)?
    - May be harder to go from 1→2 than 4→5
- As predictor variables “increase”, should sequentially step through predicted values
  - How can we ensure this happens?

Characteristics

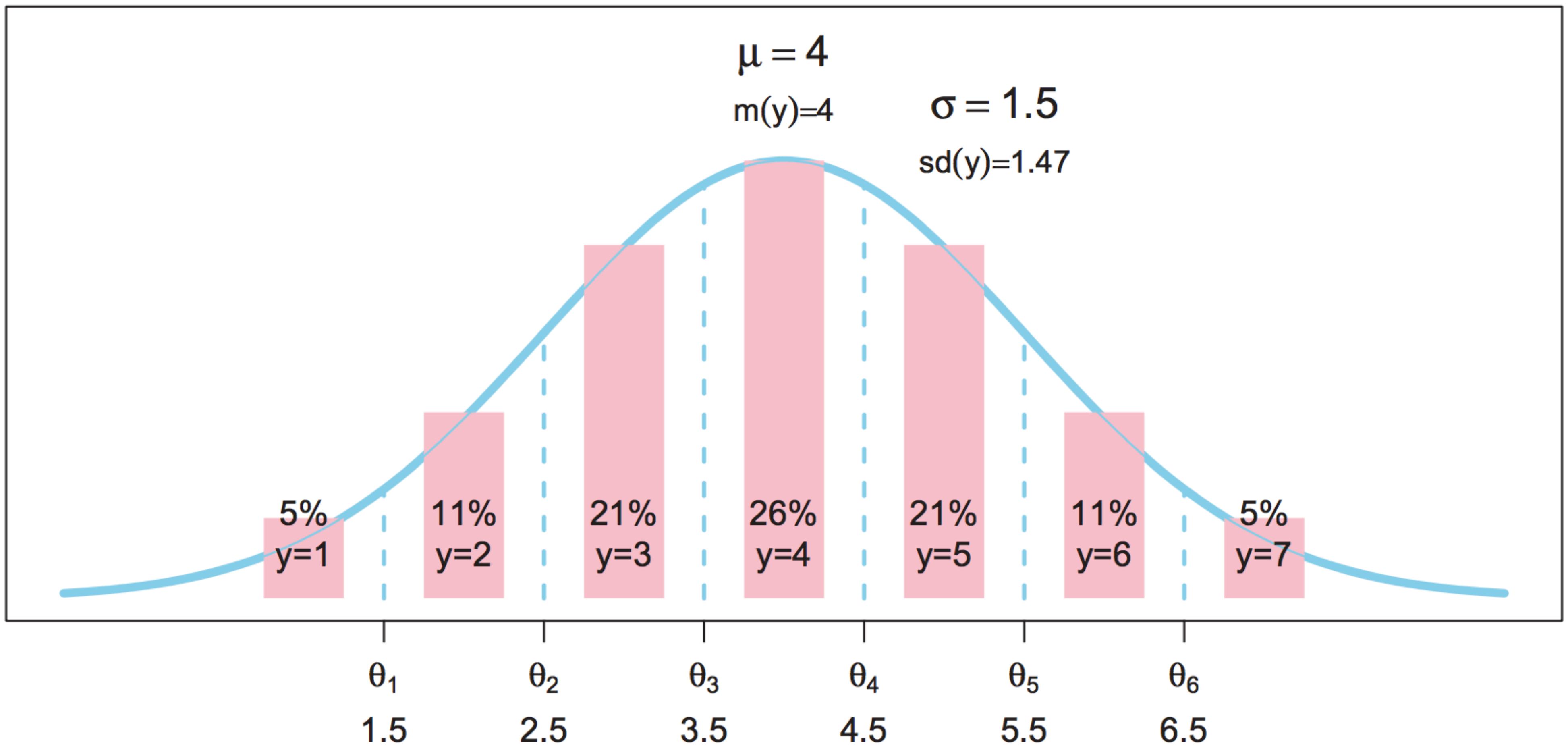
- Suppose ordinal data with 7 "levels"
- There will be cut-off points (thresholds) between levels, indicating where it switches from one to another (indicated here as θs)
- If there are *k* levels, there will be *k*-1 of these thresholds



From Kruschke (2015) p. 673

Characteristics

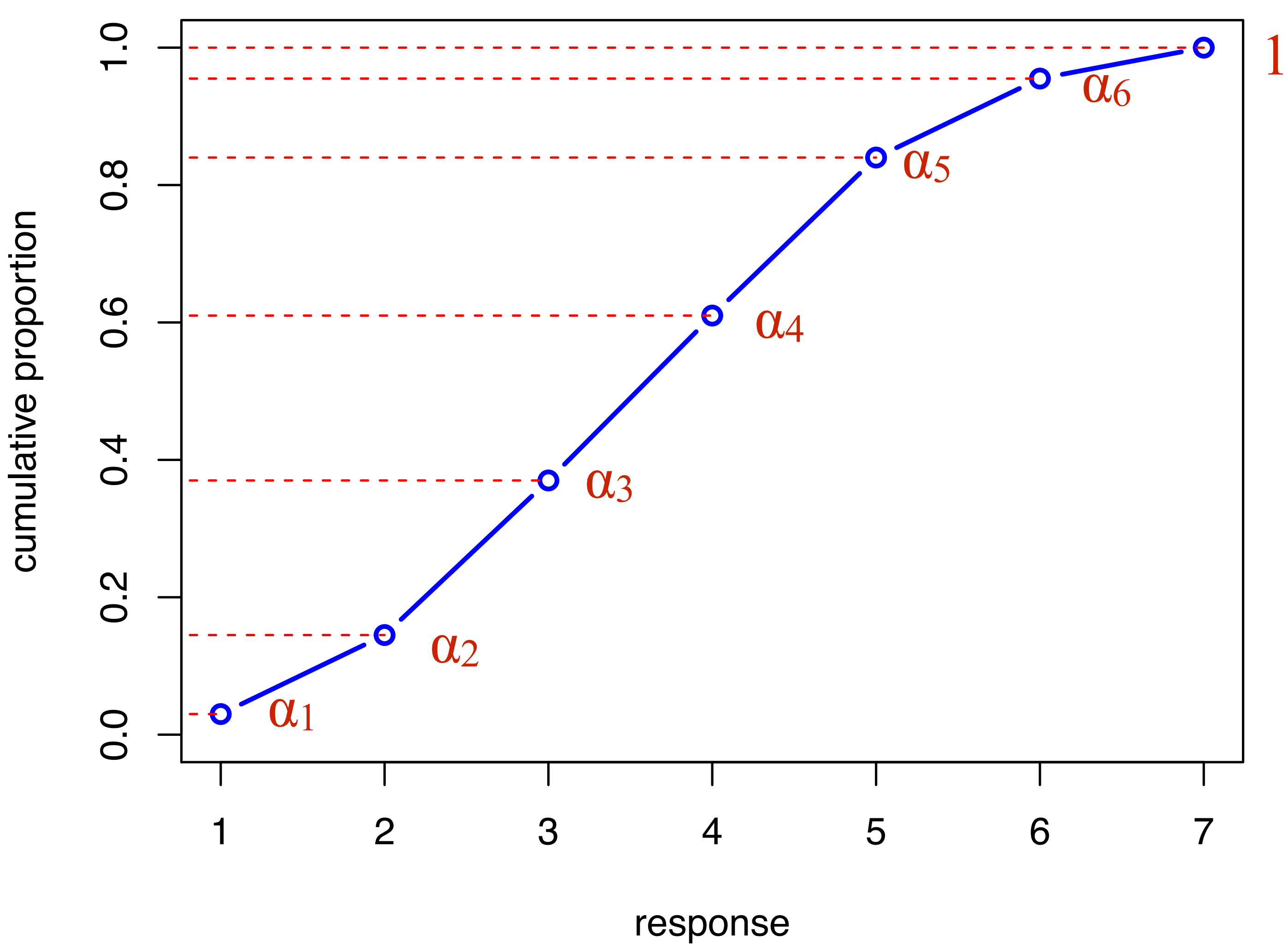
- How do we get probabilities for each level?
  - Cumulative distribution



From Kruschke (2015) p. 673

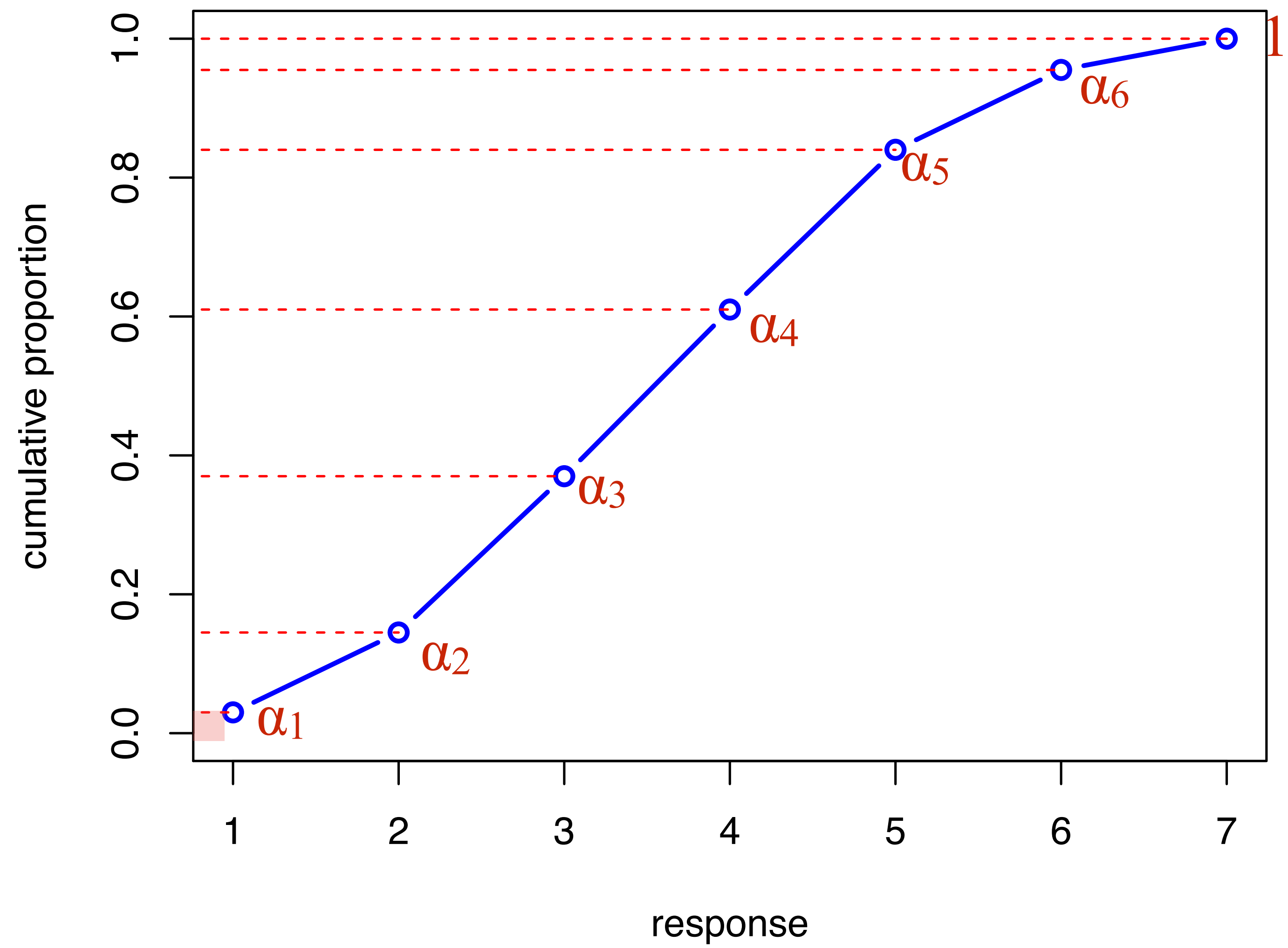
Characteristics

- Now values range from 0 to 1
- Probability for each level is the cumulative area up to the threshold just above that level minus the cumulative area up to the threshold just below that level
- Call each threshold point an α value



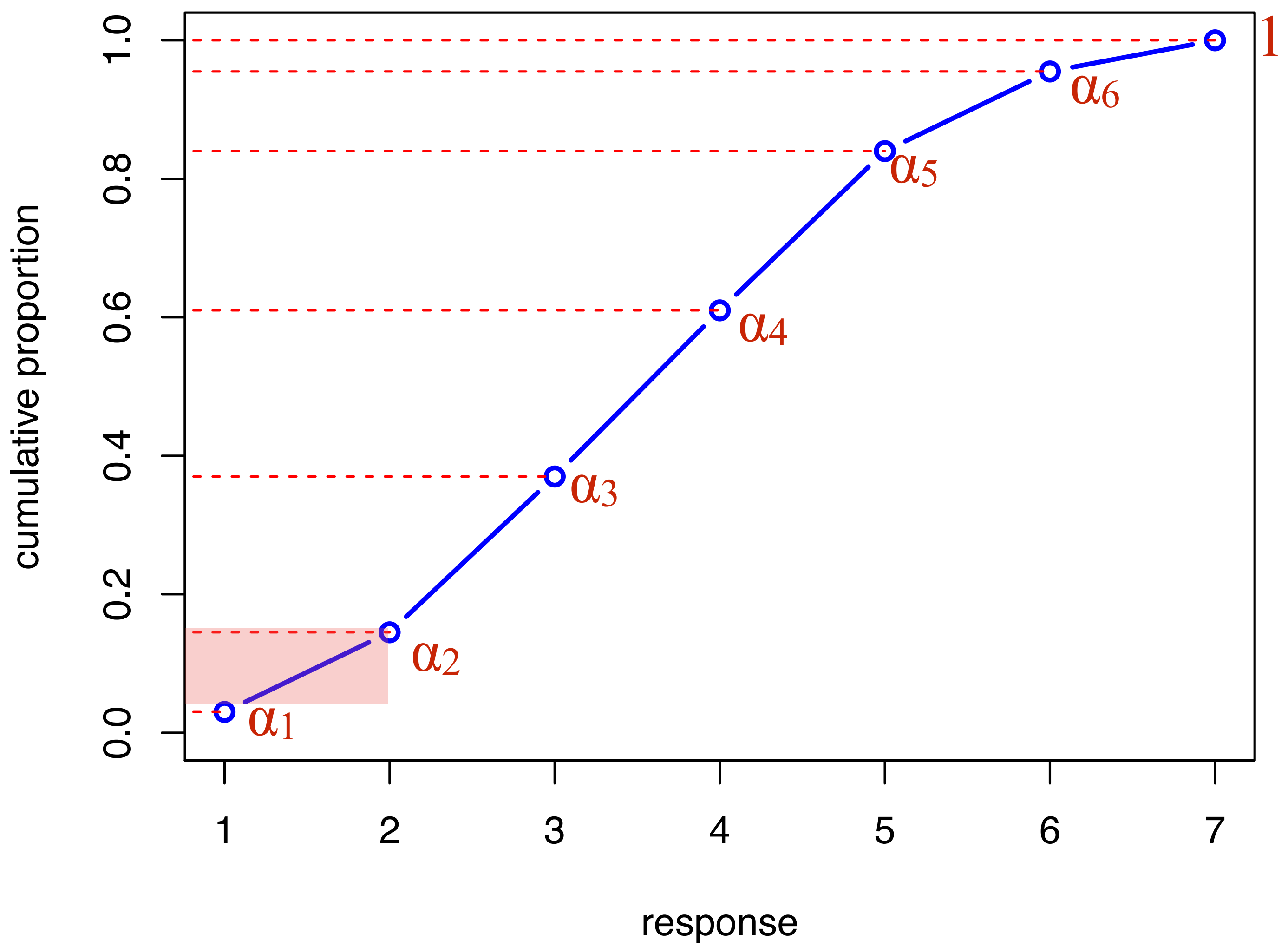
Characteristics

- For first category, probability is cumulative probability for that value, minus zero
  - Considering the mean and sd of the underlying distribution



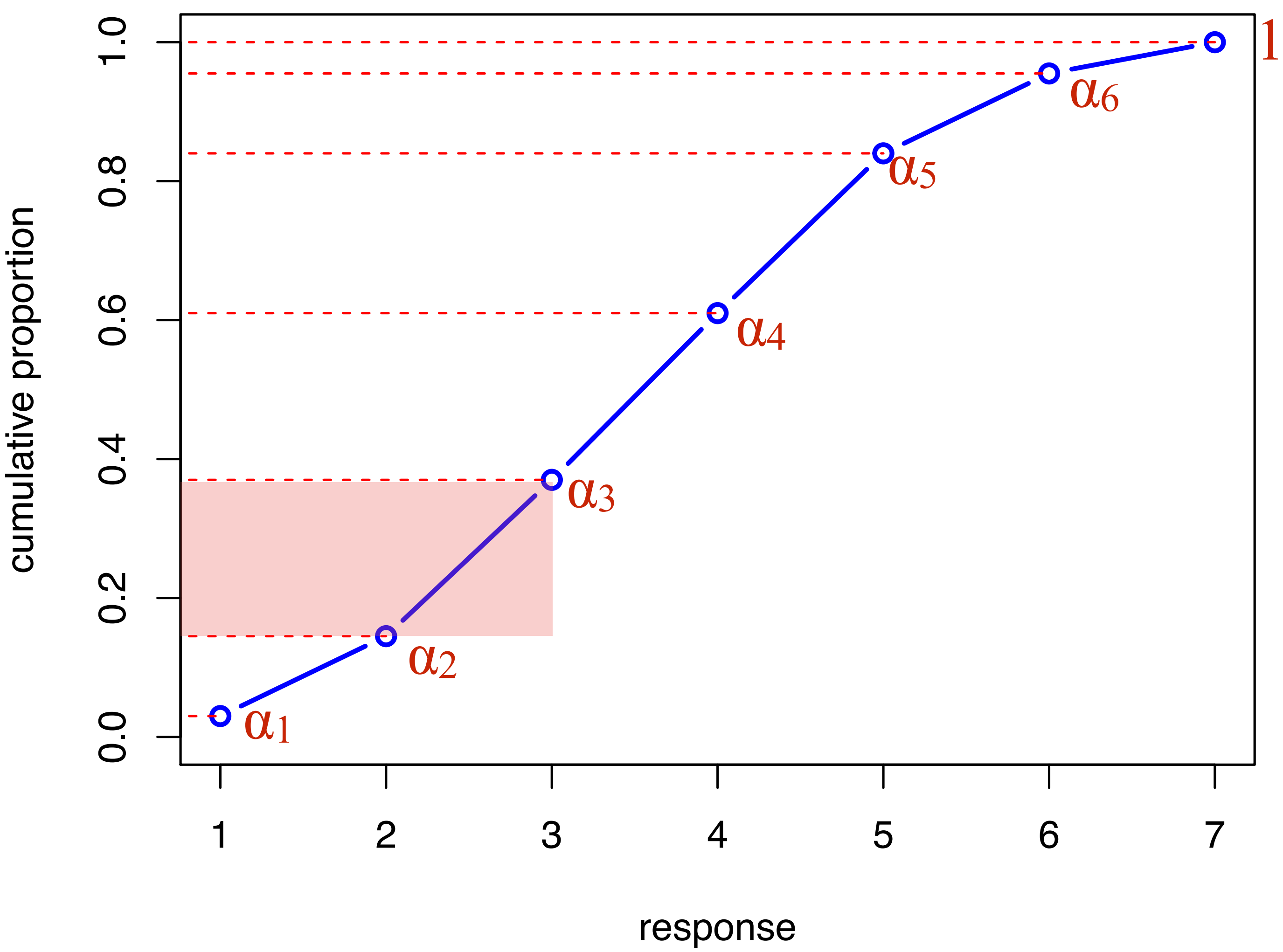
Characteristics

- For second category, probability is cumulative probability for that value, minus that for the first category



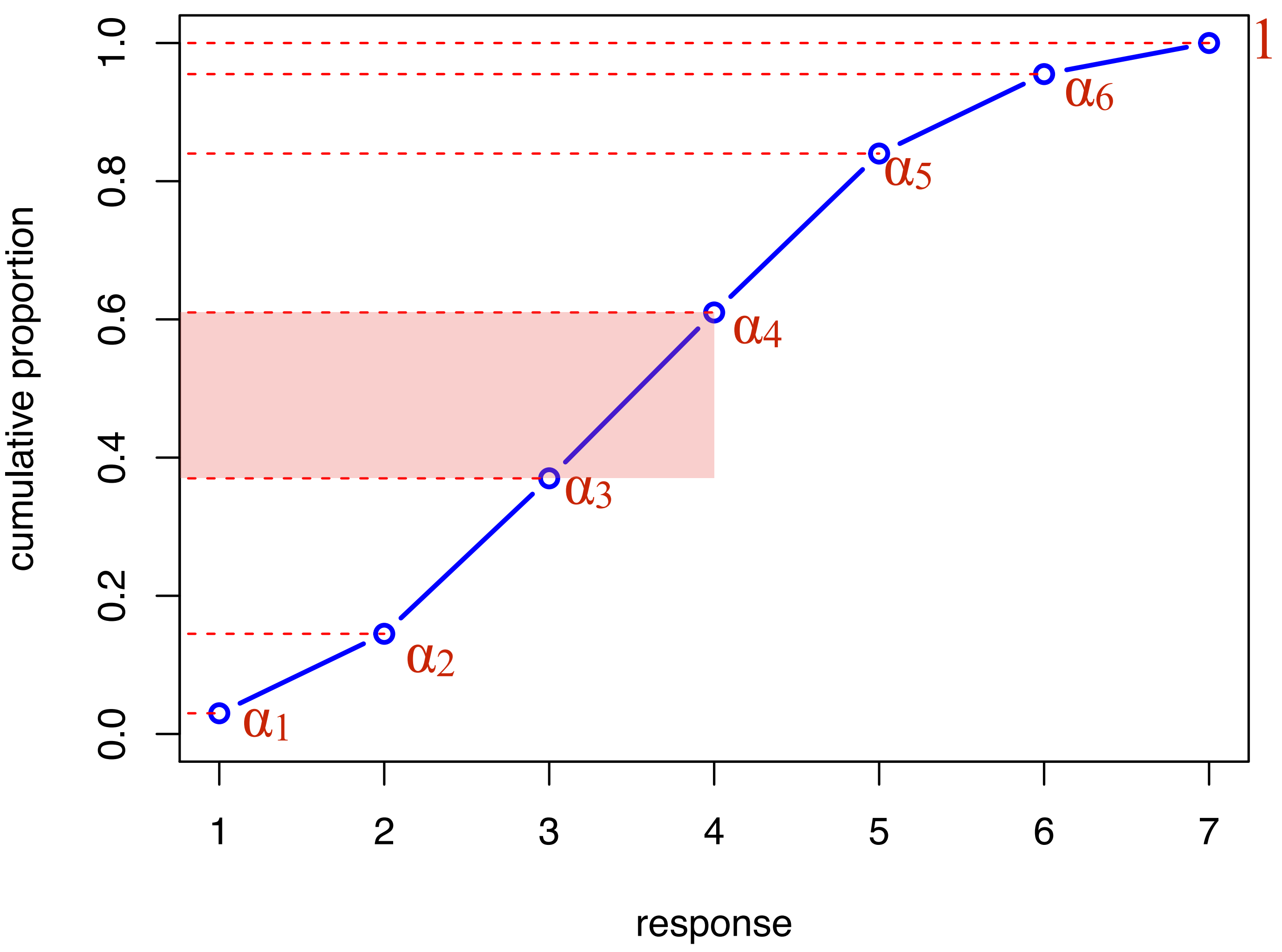
Characteristics

- For third category, probability is cumulative probability for that value, minus that for the second category



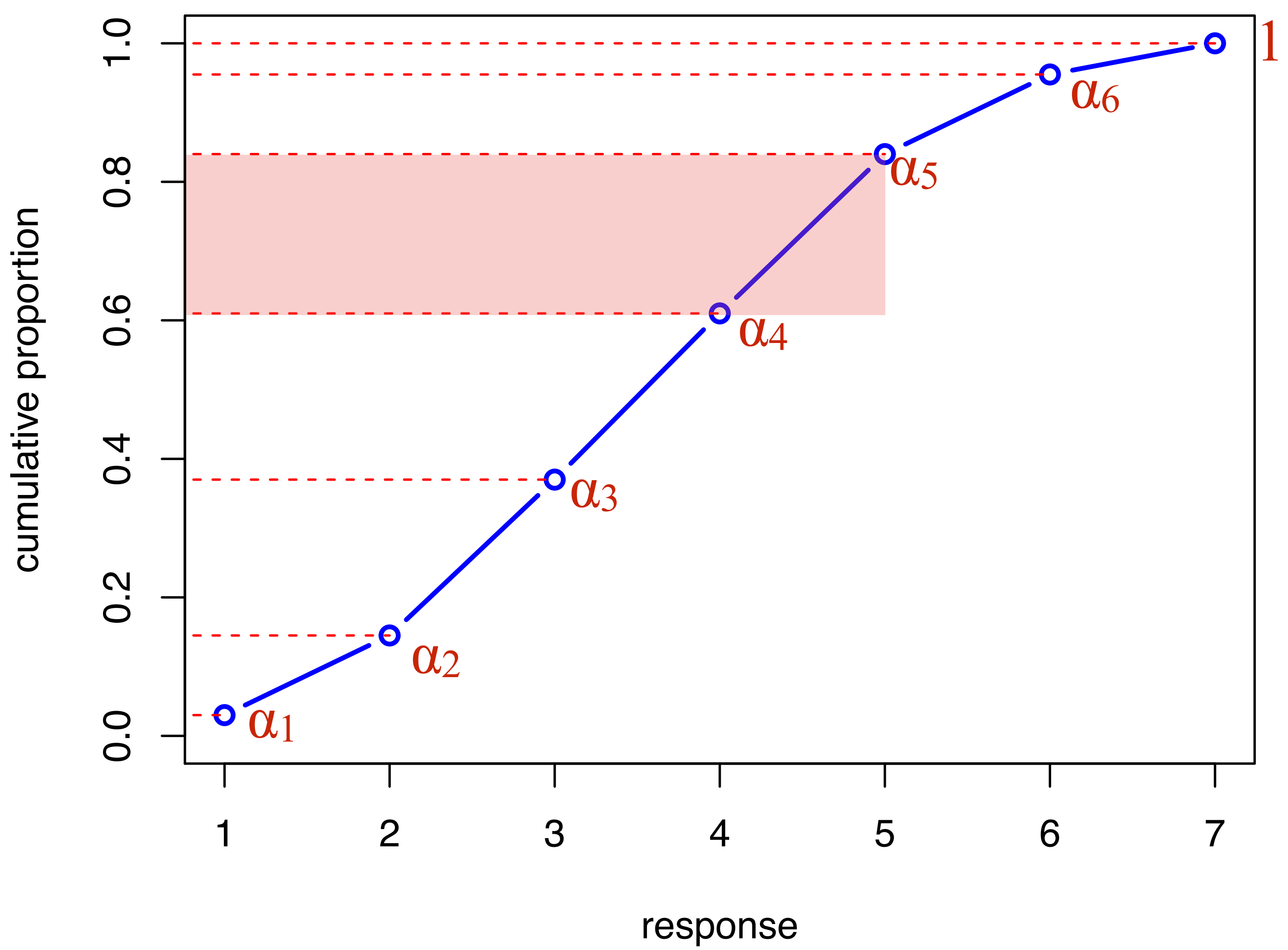
Characteristics

- For fourth category, probability is cumulative probability for that value, minus that for the third category



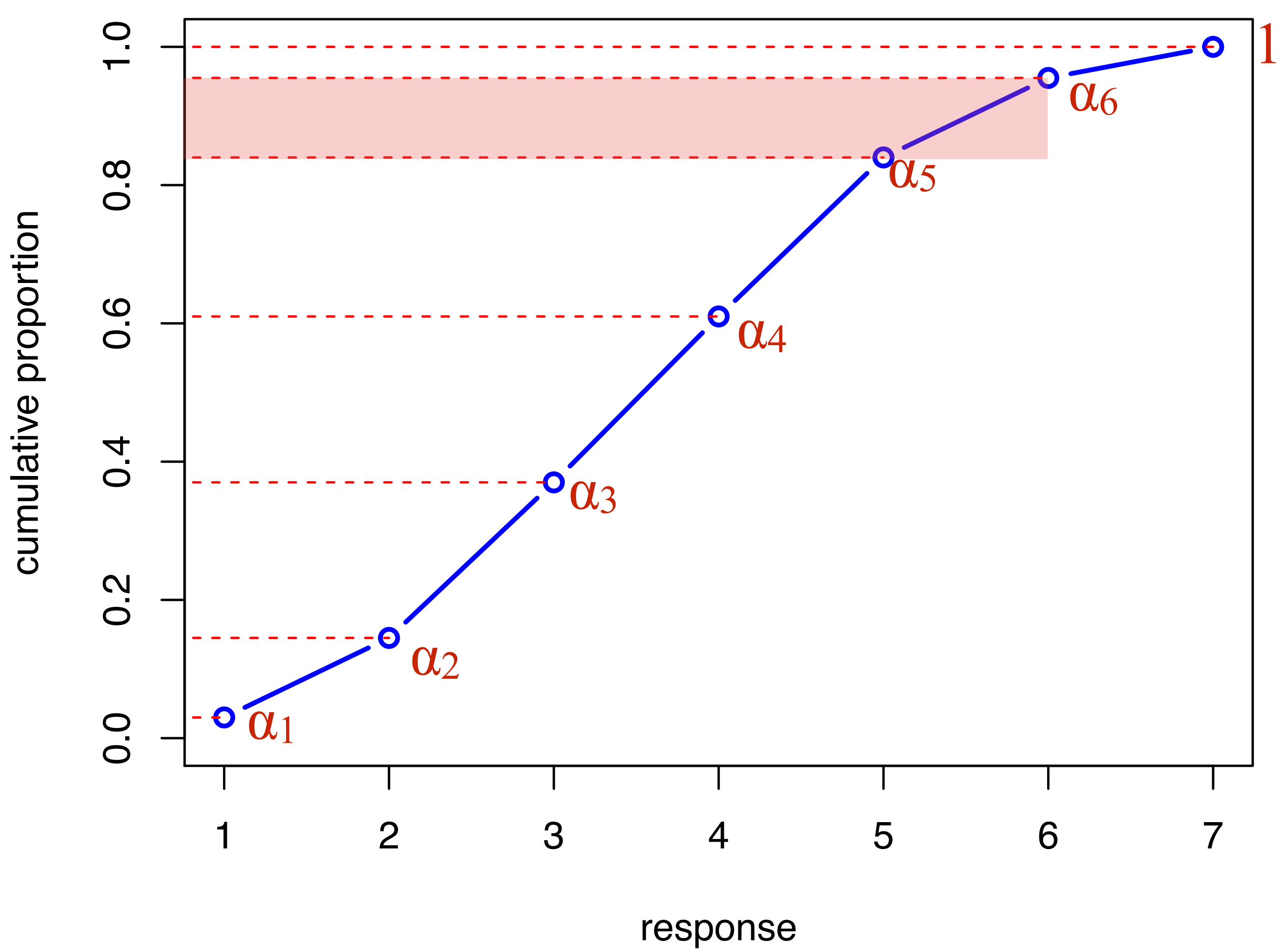
Characteristics

- For fifth category, probability is cumulative probability for that value, minus that for the fourth category



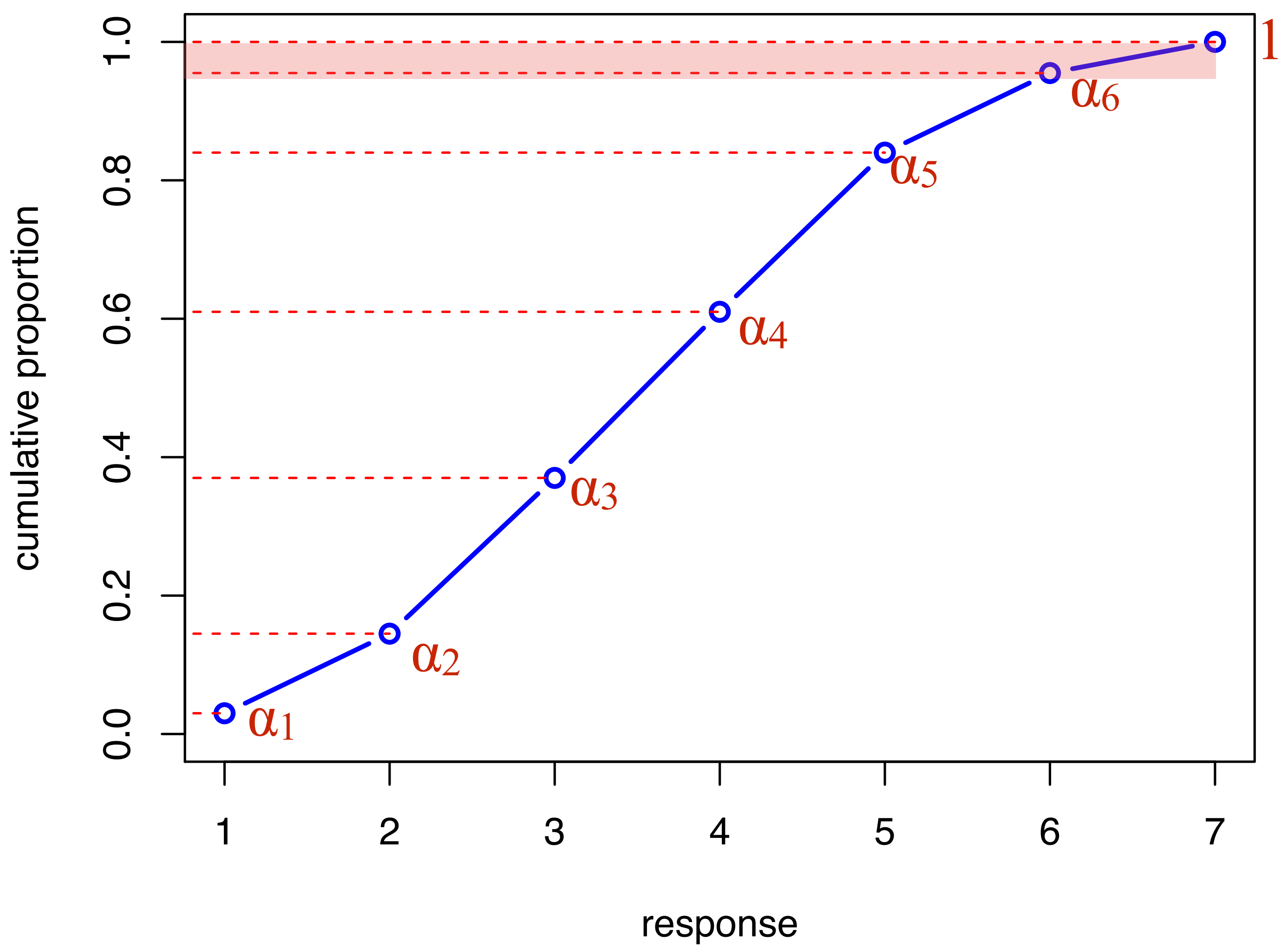
Characteristics

- For sixth category, probability is cumulative probability for that value, minus that for the fifth category



Characteristics

- For seventh category, probability is one, minus the cumulative probability for the 6th category

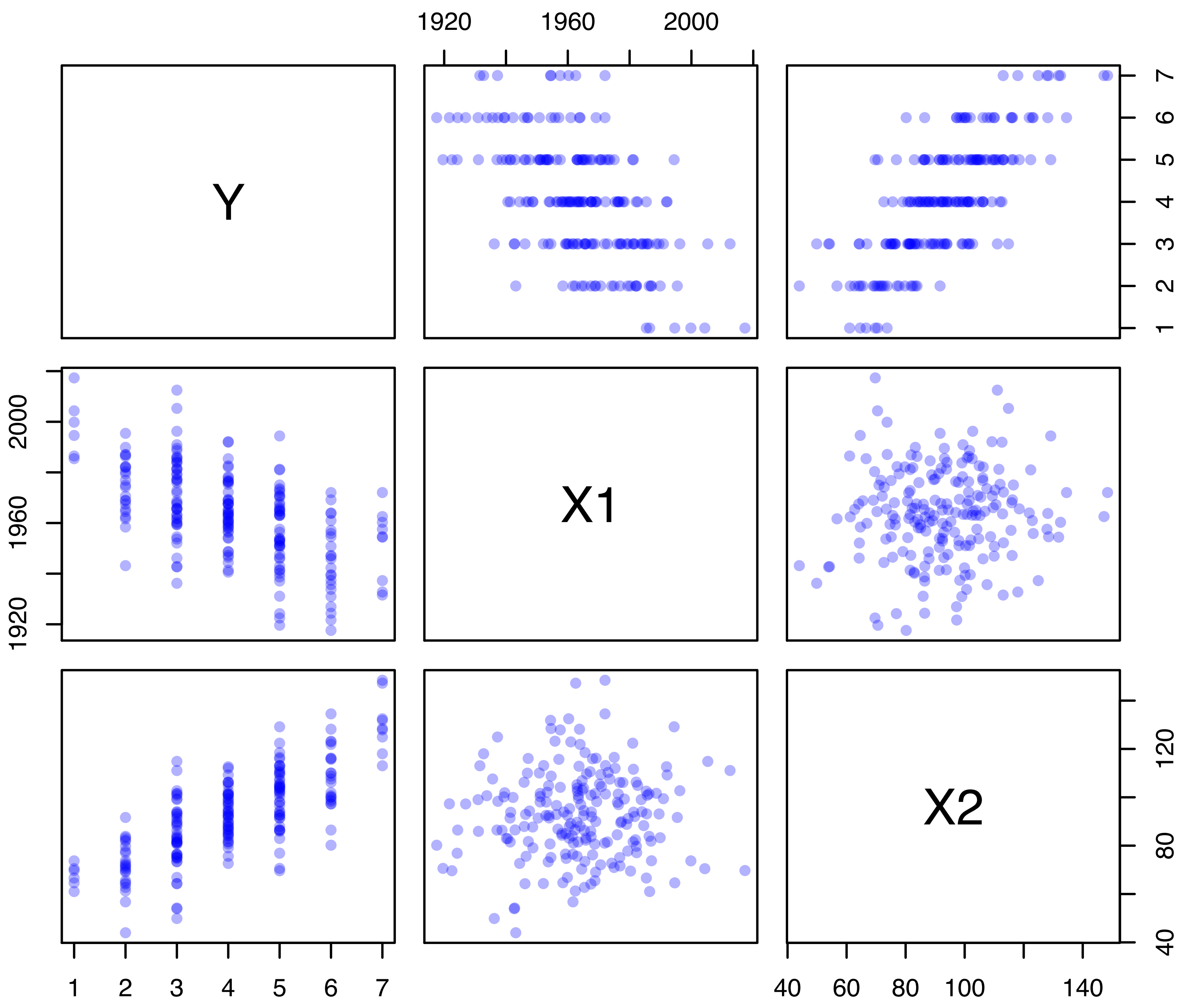


Data

- Fake data generated from code in Kruschke (2011)

Y	X1	X2
4	1962.355	86.57301
5	1966.074	109.66692
5	1953.017	98.04524
4	1960.928	89.13965
3	1968.705	80.63023
7	1932.752	118.00796
3	1976.787	75.07696
5	1965.596	118.49754
3	1959.245	76.51348
3	1962.996	75.61261
3	1952.087	64.32620

Data



Define the Model

- The **parameters** block

```
parameters {
    real b0;
    real b1;                    // Effect of first predictor variable
    real b2;                    // Effect of second predictor variable
    ordered[K-1] c;             // Vector of cutpoints
  }
```

Stan is going to estimate our cut points for us (with JAGS you had to do this yourself!!!)

Define the Model

- The **model** block

```
model {
    // Definitions
    vector[N] mu;

    // Likelihood
    for (i in 1:N) {
      mu[i] = b0 + (b1 * x1[i]) + (b2 * x2[i]);
      y[i] ~ ordered_logistic(mu[i], c);
    }

    // Priors
    b0 ~ normal(0, 1);
    b1 ~ normal(0, 1);
    b2 ~ normal(0, 1);
  }
```

Define the Model

- The **model** block

The “black box” into which we can put any equations that we have dealt with before (or more)

```
model {
    // Definitions
    vector[N] mu;

    // Likelihood
    for (i in 1:N) {
      mu[i] = b0 + (b1 * x1[i]) + (b2 * x2[i]);
      y[i] ~ ordered_logistic(mu[i], c);
    }

    // Priors
    b0 ~ normal(0, 1);
    b1 ~ normal(0, 1);
    b2 ~ normal(0, 1);
  }
```

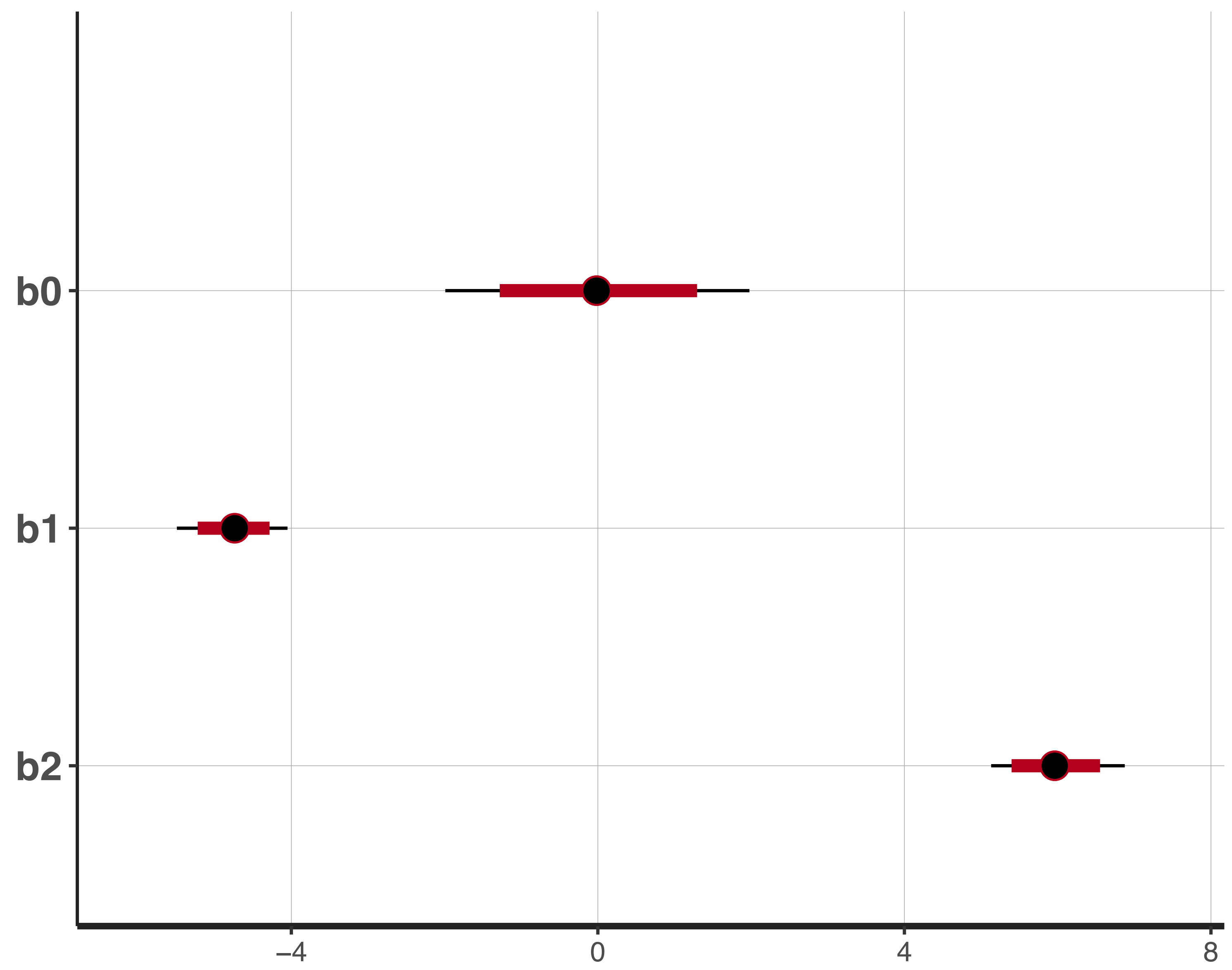
Define the Model

- The **model** block

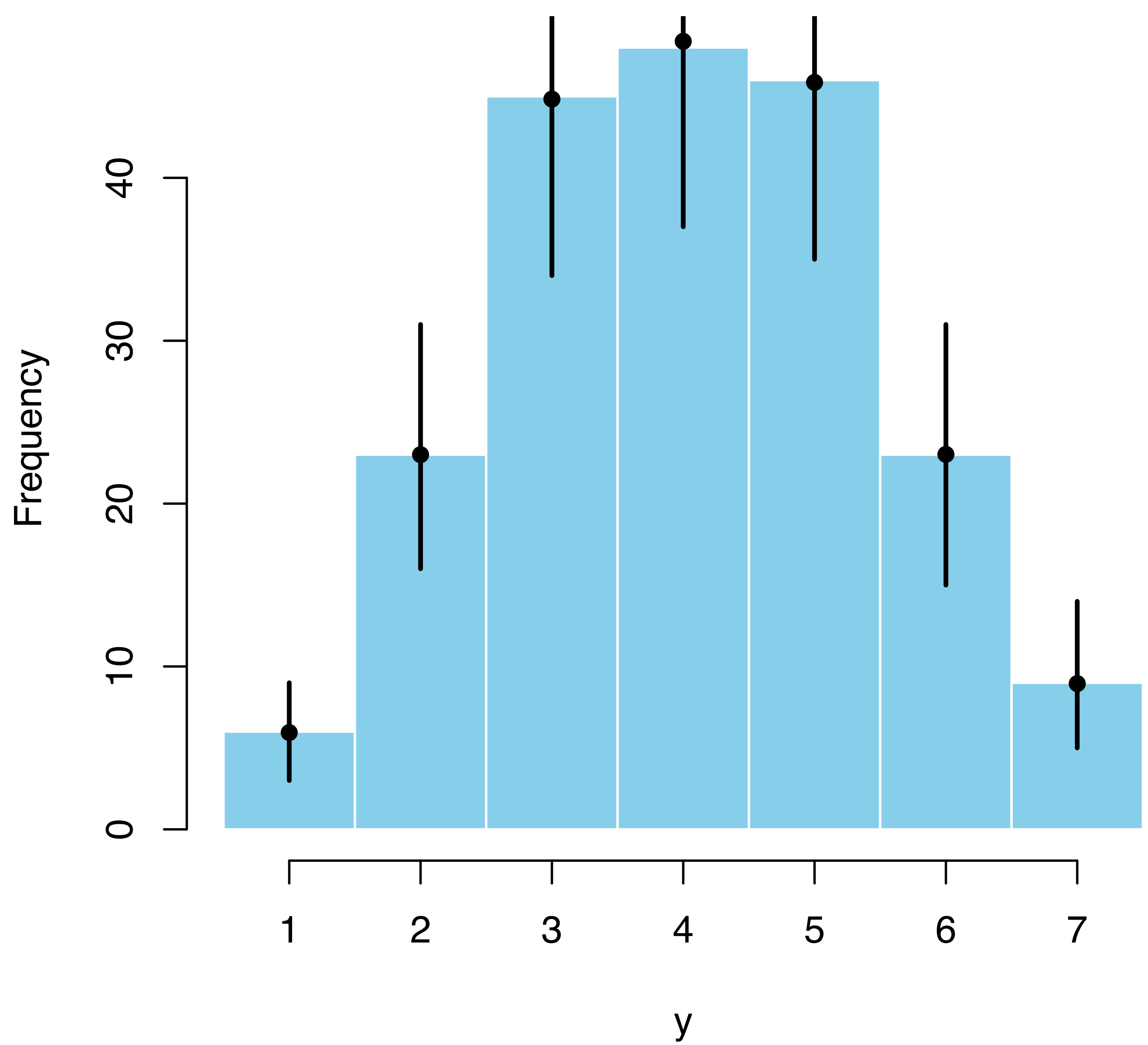
An ordered logistic likelihood with mean; cut points estimates from the data.

```
model {
    // Definitions
    vector[N] mu;

    // Likelihood
    for (i in 1:N) {
      mu[i] = b0 + (b1 * x1[i]) + (b2 * x2[i]);
      y[i] ~ ordered_logistic(mu[i], c);
    }

    // Priors
    b0 ~ normal(0, 1);
    b1 ~ normal(0, 1);
    b2 ~ normal(0, 1);
  }
```

Evaluate Results



Posterior Predictive Check



Count
Predicted Variable:

Poisson regression

Goals and
General Idea

Goals
Contingency tables

- When we have count data distributed across a range of different categories
  - Are counts in one category higher than another?
  - Is the count in one category contingent upon the the level in another category (or plural)?

	Eye Colour			
Hair Colour	**Blue**	**Brown**	**Green**	**Hazel**
Black	20	68	5	15
Blond	94	7	16	10
Brunette	84	119	29	54
Red	17	26	14	14

Data from Snee (1974) *The American Statistician* **28**: 9-12, as presented in Kruschke (2015).

Goals
Contingency tables

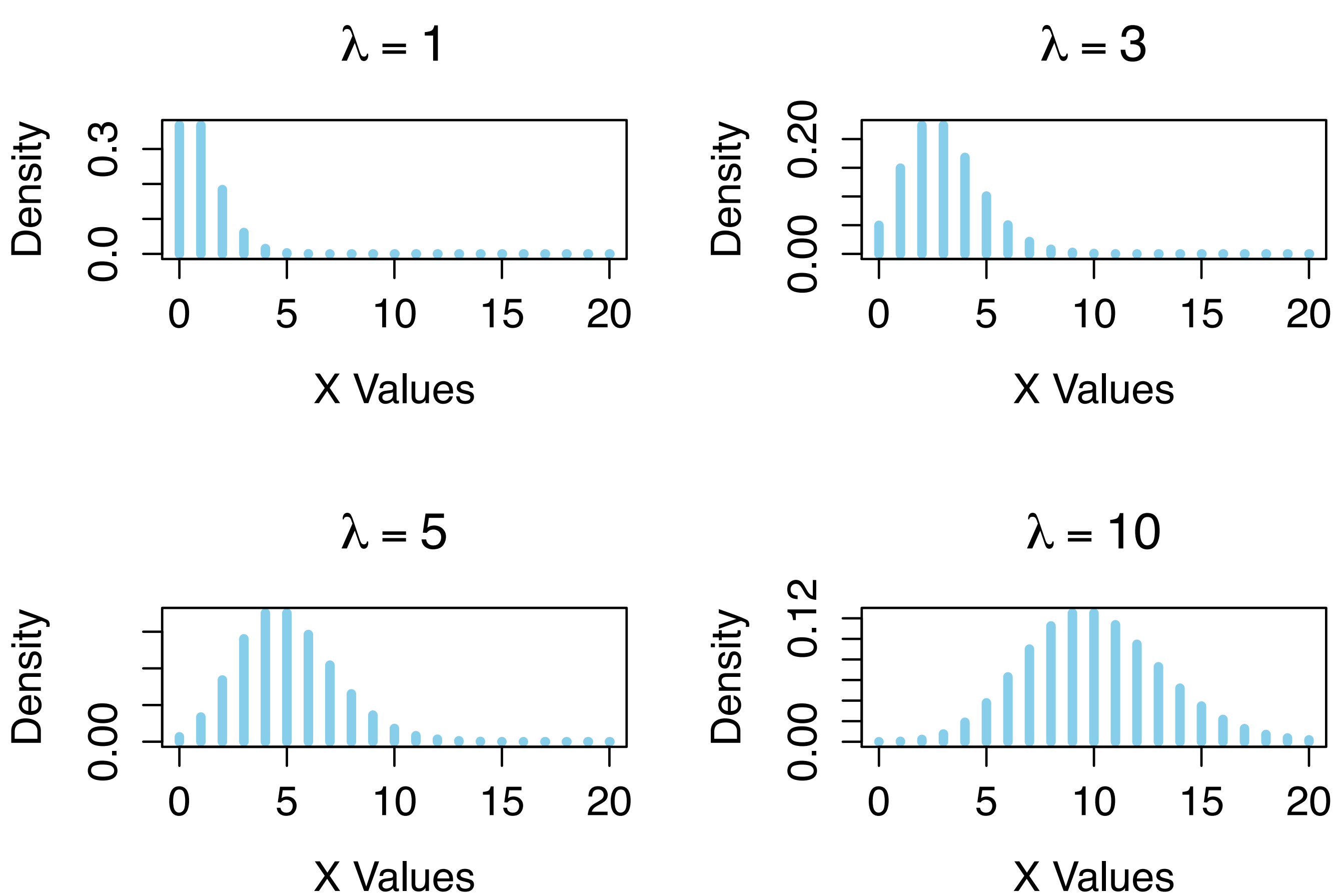
- When we have count data distributed across a range of different categories
  - Are counts in one category higher than another?
  - Is the count in one category contingent upon the the level in another category (or plural)?
- Often addressed with chi-square or Exact test analyses

Goals
Count predicted variable

- Any time our predicted variable is a count
  - Abundance
  - Counts of individuals (or things) with different traits/characteristics
  - etc.

Distribution and Links

- When modelling count data, it is appropriate to use the Poisson distribution
  - Positive integers
  - One parameter - lambda (λ)



Distribution and Links
Contingency table example

	Eye Colour			
Hair Colour	**Blue**	**Brown**	**Green**	**Hazel**
Black	20	68	5	15
Blond	94	7	16	10
Brunette	84	119	29	54
Red	17	26	14	14

- Cell frequencies are representative of underlying cell probabilities
  - Are nominal variables independent of each other?

Distribution and Links
Contingency table example

Hair Colour	Eye Colour Blue	Brown	Green	Hazel
Black	20	68	5	15
Blond	94	7	16	10
Brunette	84	119	29	54
Red	17	26	14	14

- Cell frequencies are representative of underlying cell probabilities
  - Are nominal variables independent of each other?
- If independent

$$68 = Pr(\mathrm{BlackHair}) \ \text{x} \ Pr(\mathrm{BrownEyes})$$

True for all cells

Distribution and Links
Contingency table example

	Eye Colour			
Hair Colour	Blue	Brown	Green	Hazel
Black	20	68	5	15
Blond	94	7	16	10
Brunette	84	119	29	54
Red	17	26	14	14

- Cell frequencies are representative of underlying cell probabilities
  - Are nominal variables independent of each other?
- If independent

$$68 = Pr(\text{BlackHair}) \ \text{x} \ Pr(\text{BrownEyes})$$

True for all cells

- If interaction effects, this will not be the case

Distribution and Links
Contingency table example

	Eye Colour				
Hair Colour	**Blue**	**Brown**	**Green**	**Hazel**	
Black	20	68	5	15	
Blond	94	7	16	10	
Brunette	84	119	29	54	
Red	17	26	14	14	
$f(c)$	**215**	**220**	**64**	**93**	**592** N

marginal frequencies of each eye colour

Distribution and Links
Contingency table example

	Eye Colour				
Hair Colour	**Blue**	**Brown**	**Green**	**Hazel**	
Black	20	68	5	15	
Blond	94	7	16	10	
Brunette	84	119	29	54	
Red	17	26	14	14	
$f(c)$	**215**	**220**	**64**	**93**	**592** N

marginal frequencies of each eye colour

$$Pr(\text{BlueEyes}) = 215 \;/\; 592 = 0.363$$

Distribution and Links
Contingency table example

	Eye Colour				
Hair Colour	**Blue**	**Brown**	**Green**	**Hazel**	
Black	20	68	5	15	**108**
Blond	94	7	16	10	**127**
Brunette	84	119	29	54	**286**
Red	17	26	14	14	**71**
$f(c)$	**215**	**220**	**64**	**93**	**592**
					N

marginal frequencies of each hair colour

marginal frequencies of each eye colour

$Pr(\text{BlueEyes}) = 215 / 592 = 0.363$

Distribution and Links
Contingency table example

Hair Colour	Blue	Brown	Green	Hazel	
	Eye Colour				
Black	20	68	5	15	**108**
Blond	94	7	16	10	**127**
Brunette	84	119	29	54	**286**
Red	17	26	14	14	**71**
$f(c)$	**215**	**220**	**64**	**93**	**592**

marginal frequencies of each hair colour

N

marginal frequencies of each eye colour

$$Pr(\text{BlueEyes}) = 215 / 592 = 0.363$$

$$Pr(\text{BlackHair}) = 108 / 592 = 0.182$$

Distribution and Links
Contingency table example

	Eye Colour				
Hair Colour	Blue	Brown	Green	Hazel	
Black	20	68	5	15	**108**
Blond	94	7	16	10	**127**
Brunette	84	119	29	54	**286**
Red	17	26	14	14	**71**
$f(c)$	**215**	**220**	**64**	**93**	**592**
					N

marginal frequencies of each hair colour

marginal frequencies of each eye colour

Pr(BlueEyes) = 215 / 592 = 0.363

Pr(BlackHair) = 108 / 592 = 0.182

Pr(BlueEyes & BlackHair) = (0.363 x 0.182) x 592 = 39

Hmm...must be an interaction effect

Distribution and Links
Contingency table example

$$Pr(\text{BlueEyes \& BlackHair}) = (0.363 \text{ x } 0.182) \text{ x } 592 = 39$$

- Joint probability is the product of the relevant marginal probabilities
- We're used to dealing with additive combinations
  - Can convert to log scale, then they'll be additive

Segue on Logarithms

- Adding logarithms is the same as multiplying original values

$$10 \times 5 = 50$$

$$log(10) + log(5) = log(50)$$

Segue on Logarithms

- Adding logarithms is the same as multiplying original values

$$10 \times 5 = 50$$

$$log(10) + log(5) = log(50)$$

- Can cancel out logarithms (bring back to original scale), by raising them to the exponent

$$exp(log(10) + log(5)) = 50$$

Distribution and Links

$$y_i \sim Poisson(\lambda)$$

$$\lambda = \exp(\beta_0 + \beta_1[x[1]] + \beta_2[x[2]] + \beta_{1,2}[x[1], x[2]])$$

Distribution and Links

$$y_i \sim Poisson(\lambda)$$

$$\lambda = \exp(\boxed{\beta_0 + \beta_1[x[1]] + \beta_2[x[2]] + \beta_{1,2}[x[1], x[2]]})$$

The “black box” into which we can put any of our previous equations, or others

Distribution and Links

$$y_i \sim Poisson(\lambda)$$

$$\lambda = \exp(\beta_0 + \beta_1[x[1]] + \beta_2[x[2]] + \beta_{1,2}[x[1], x[2]])$$

Depends on rest of model, here the average across all categories of all variables

Distribution and Links

$$y_i \sim Poisson(\lambda)$$

$$\lambda = \exp(\beta_0 + \beta_1[x[1]] + \beta_2[x[2]] + \beta_{1,2}[x[1], x[2]])$$

The deflection away from baseline due to being in each category of our first nominal predictor variable.

Distribution and Links

$$y_i \sim Poisson(\lambda)$$

$$\lambda = \exp(\beta_0 + \beta_1[x[1]] + \boxed{\beta_2}[x[2]] + \beta_{1,2}[x[1], x[2]])$$

The deflection away from baseline due to being in each category of our second nominal predictor variable.

Distribution and Links

$$y_i \sim Poisson(\lambda)$$

$$\lambda = \exp(\beta_0 + \beta_1[x[1]] + \beta_2[x[2]] + \boxed{\beta_{1,2}}[x[1], x[2]])$$

Interaction effects.

Distribution and Links

$$y_i \sim Poisson(\lambda)$$

$$\lambda = \exp(\beta_0 + \beta_1[x[1]] + \beta_2[x[2]] + \beta_{1,2}[x[1], x[2]])$$

Note that coefficient estimates will now be on the log scale
(even though we haven't explicitly specified them as such)

Distribution and Links

- The link between the predictor and predicted variables is based on logarithms
  - Previous models (other than logistic) have been based on identity
  - These types of models are called log linear models

Organize the Data

```
# Y-Data
y = as.integer(haireye$Freq)
N = length(y)
yLogMean = log(mean(y))
yLogSD = log(sd(c(rep(0, N - 1), sum(y))))
```

Organize the Data

```
# Y-Data
y = as.integer(haireye$Freq)
N = length(y)
yLogMean = log(mean(y))
yLogSD = log(sd(c(rep(0, N - 1), sum(y))))
```

Will see why we need these in a minute...

Organize the Data

```
# Y-Data
y = as.integer(haireye$Freq)
N = length(y)
yLogMean = log(mean(y))
yLogSD = log(sd(c(rep(0, N - 1), sum(y))))
```

Largest possible sd would occur if the sum of all values was in one cell, and all others were zero

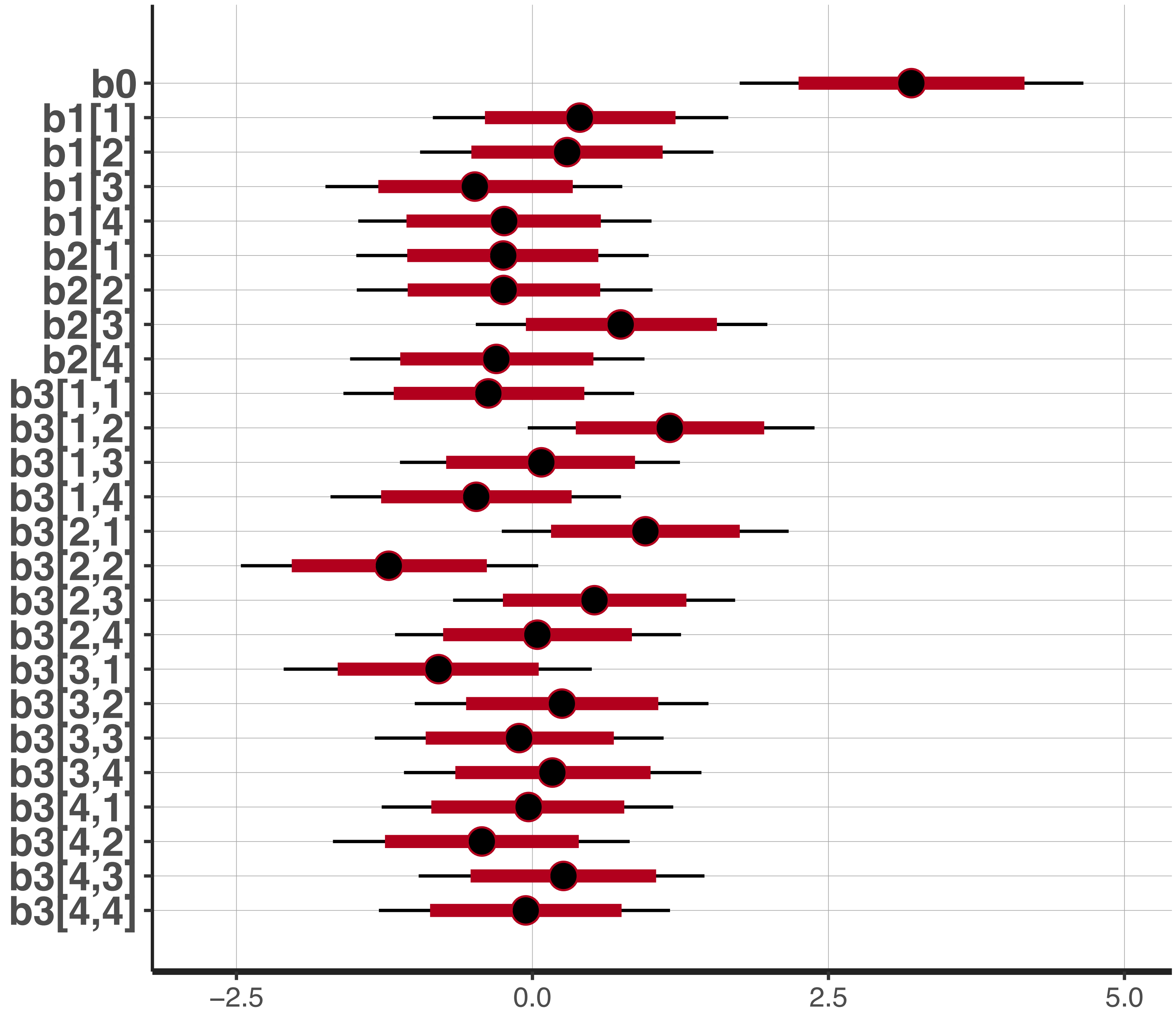
Define the Model

- The **model** block

```
model {
    // Definitions
    vector[N] lambda;

    // Likelihood
    for (i in 1:N) {
      lambda[i] = exp(b0 + b1[eye[i]] + b2[hair[i]] + b3[eye[i], hair[i]]);
      y[i] ~ poisson(lambda[i]);
    }

    // Priors
    b0 ~ normal(yLogMean, yLogSD);

    for (j in 1:nEyeColours) {
      b1[j] ~ normal(0, 1);
    }

    for (j in 1:nHairColours) {
      b2[j] ~ normal(0, 1);
    }

    for (j in 1:nEyeColours) {
      for (k in 1:nHairColours) {
        b3[j, k] ~ normal(0, 1);
      }
    }
  }
```

Plotting Posterior Distributions

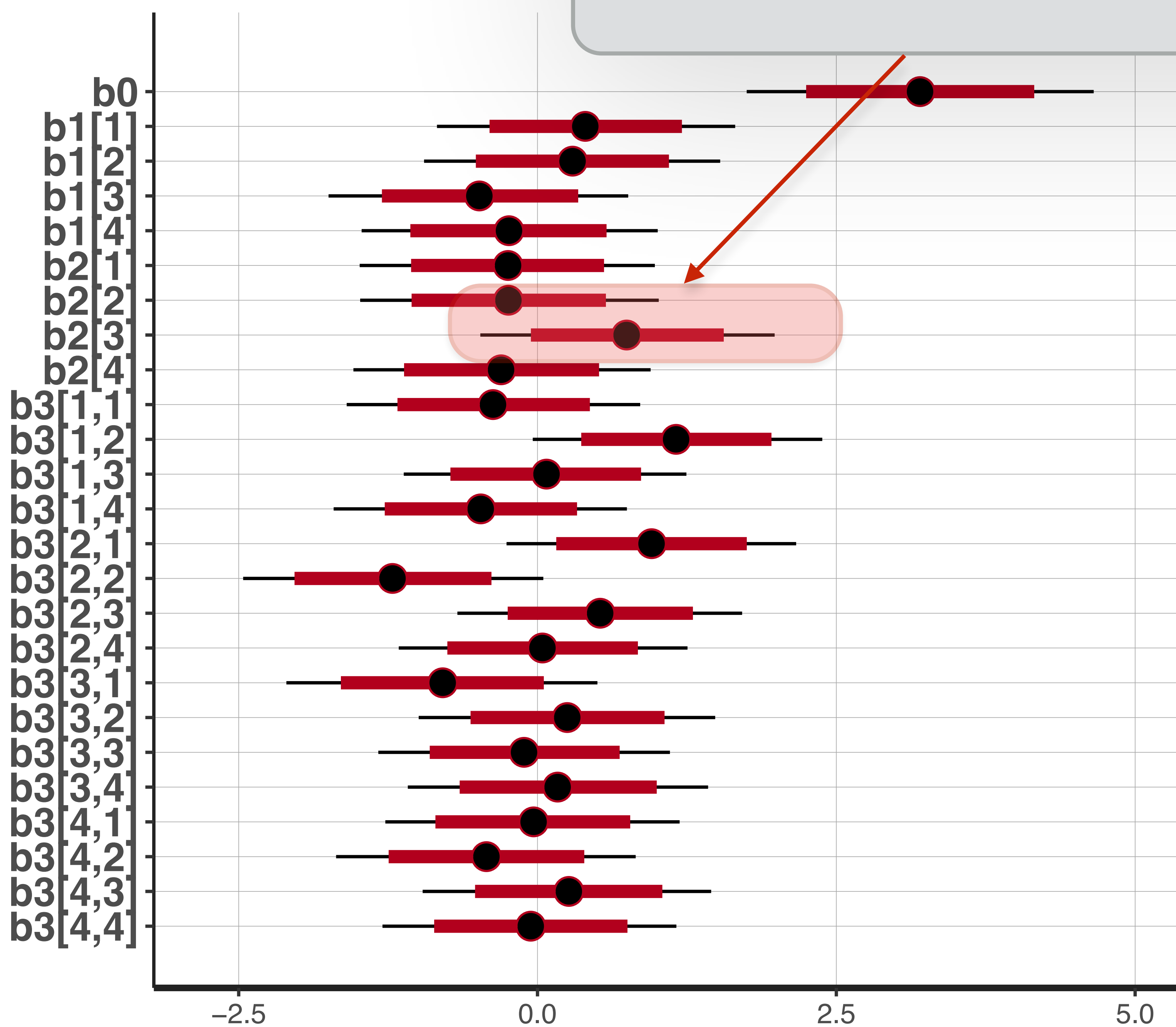
```
stan_plot(stanFit, par = c("b0", "b1", "b2", "b3"))
```



Plotting Posterior Distributions

```
stan_plot(stanFit, par = c("b0", "b1", "b2
```

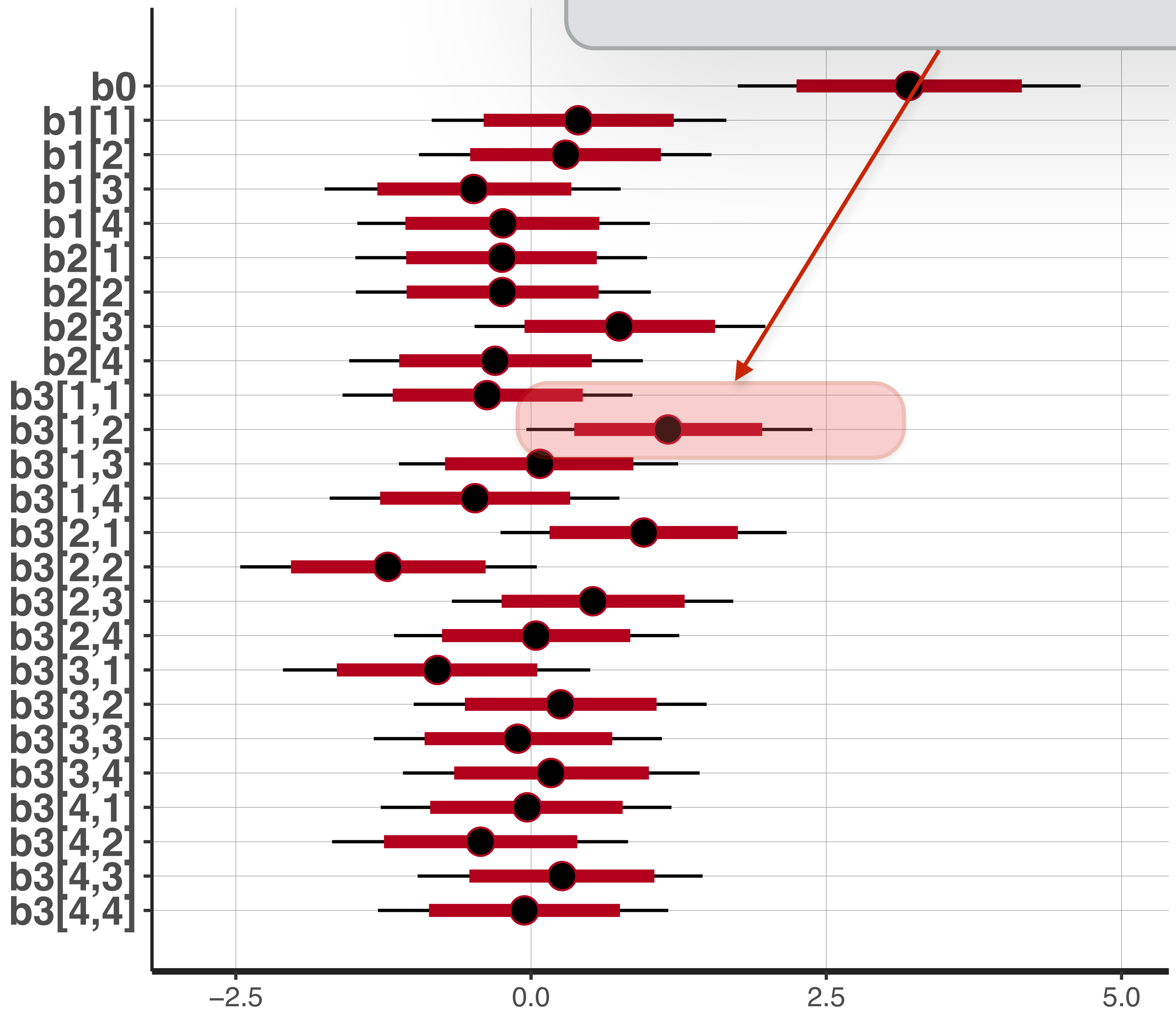
Brunettes



Plotting Posterior Distributions

```
stan_plot(stanFit, par = c("b0", "b1", "b2
```

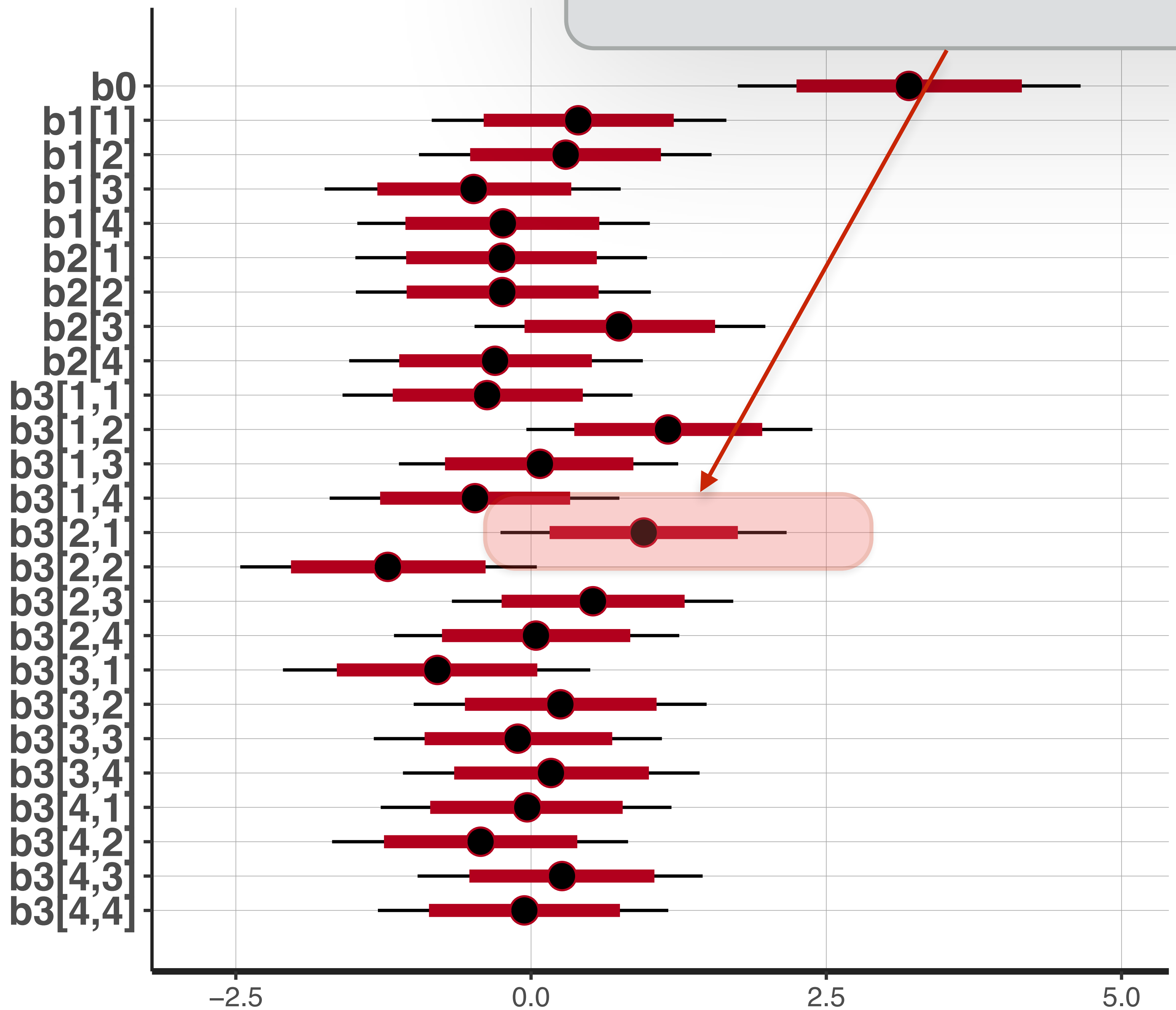
Blue eyes & blonde hair



Plotting Posterior Distributions

```
stan_plot(stanFit, par = c("b0", "b1", "b
```

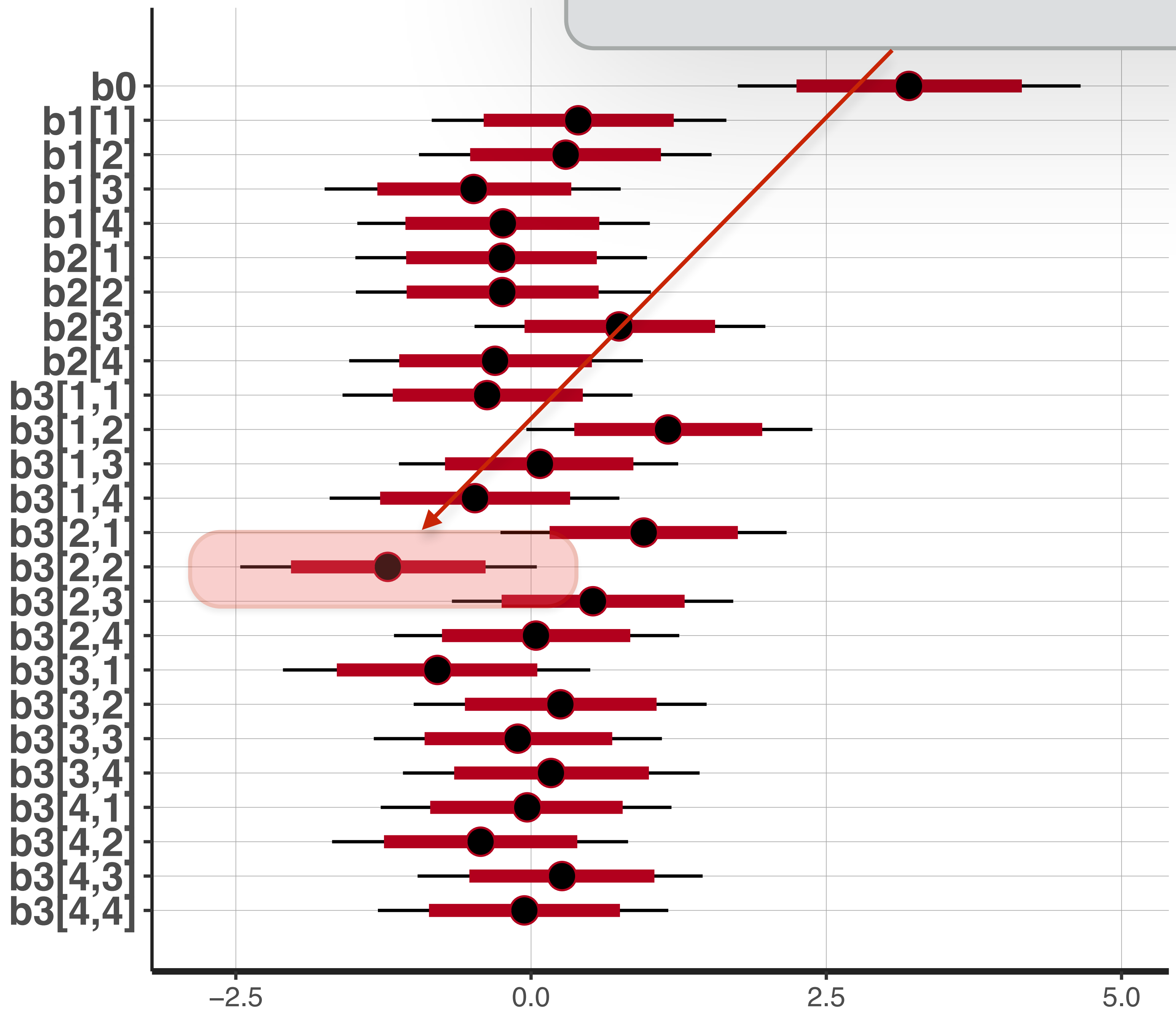
Brown eyes and black hair



Plotting Posterior Distributions

```
stan_plot(stanFit, par = c("b0", "b1", "b
```

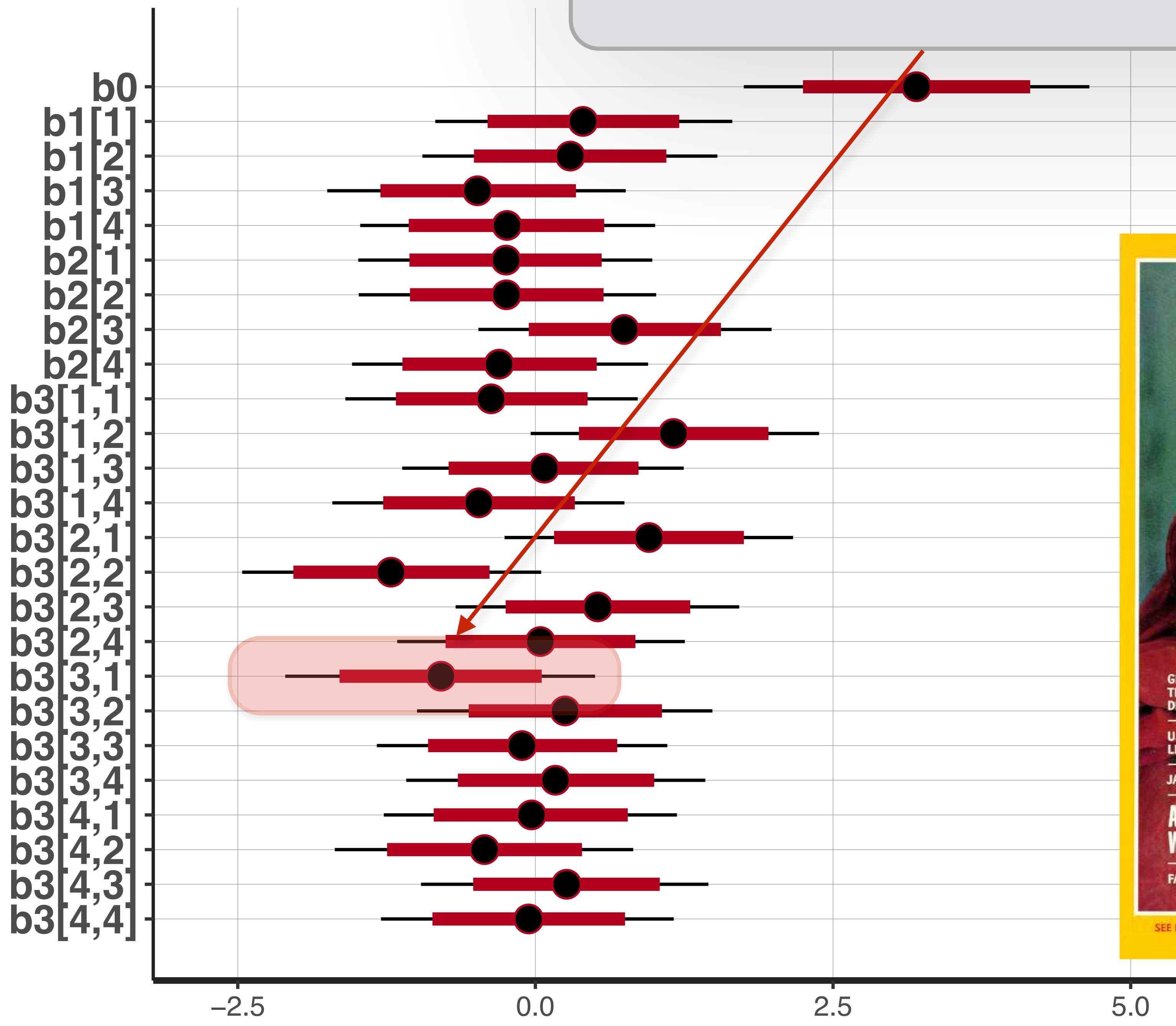
Brown eyes and blonde hair

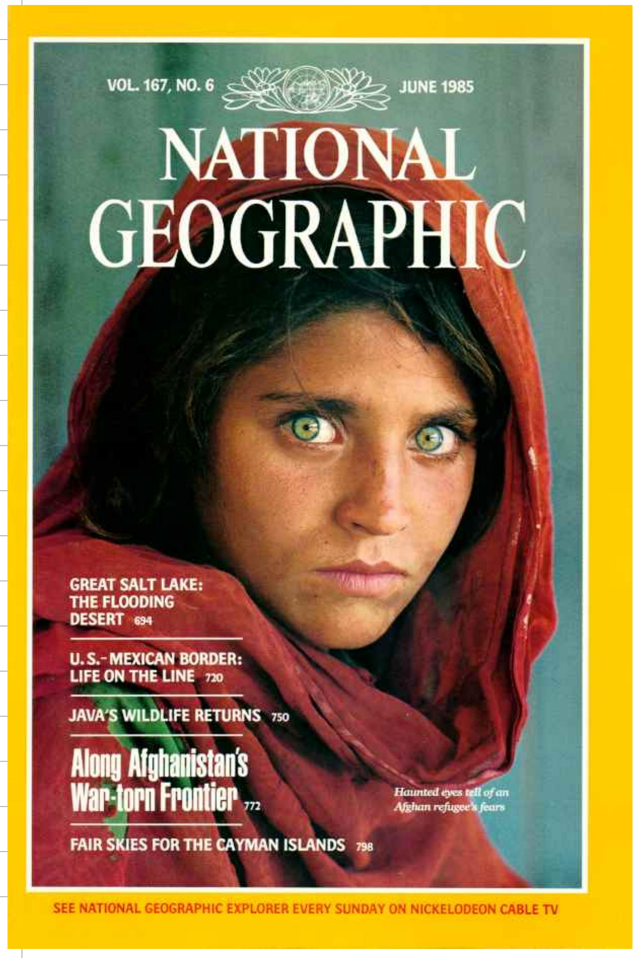


Plotting Posterior Distributions

```
stan_plot(stanFit, par = c("b0", "b1", "b2
```

Green eyes and black hair





Questions?

Thank you!!