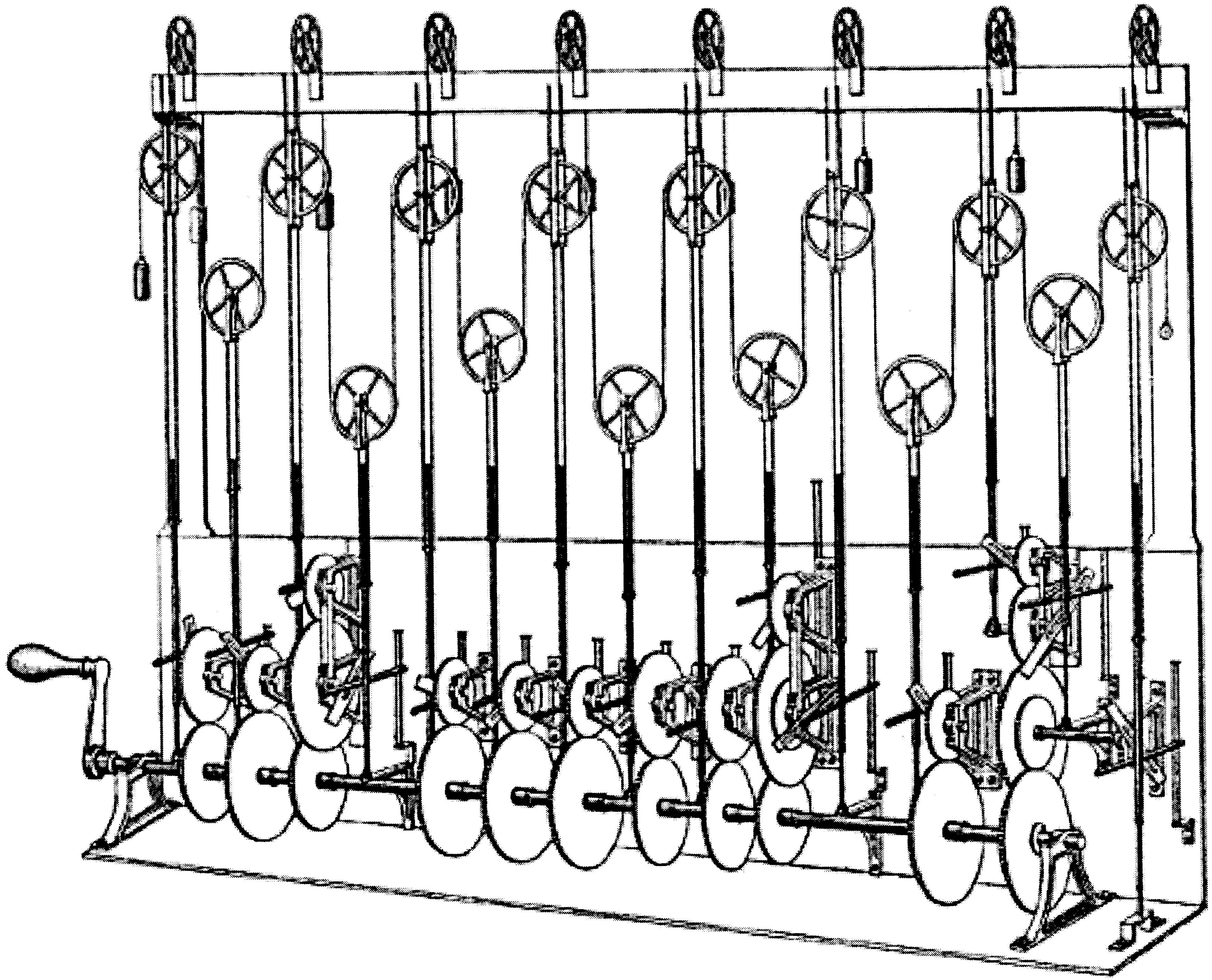
Statistical Rethinking

Week 7: The Generalized Linear Model

Richard McElreath



Tide prediction machine, 1879 design,
by William Thomson (Lord Kelvin, 1824–1907)

Generalized Linear Models

- Goal: Connect linear model to outcome variable
  - Would be better to ditch linear model, too
- Can model multivariate relationships and non-linear responses
- Building blocks of multilevel models
- Strategy:
  1. Pick an outcome distribution
  2. Model its parameters using links to linear models
  3. Compute posterior

Generalized Linear Models

- (1) Pick an outcome distribution
    - Distances and durations: exponential, gamma (*survival* or *event history*)
    - Counts: Poisson, binomial, multinomial, geometric
    - Monsters: Ranks and ordered categories
    - Mixtures: Beta-binomial, gamma-Poisson, zero-inflated processes

Generalized Linear Models

- (2) Model parameters with a *link*

$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma), \\
\mu_i &= \alpha + \beta x_i.
\end{aligned}
$$

Generalized Linear Models

- (2) Model parameters with a *link*

same units

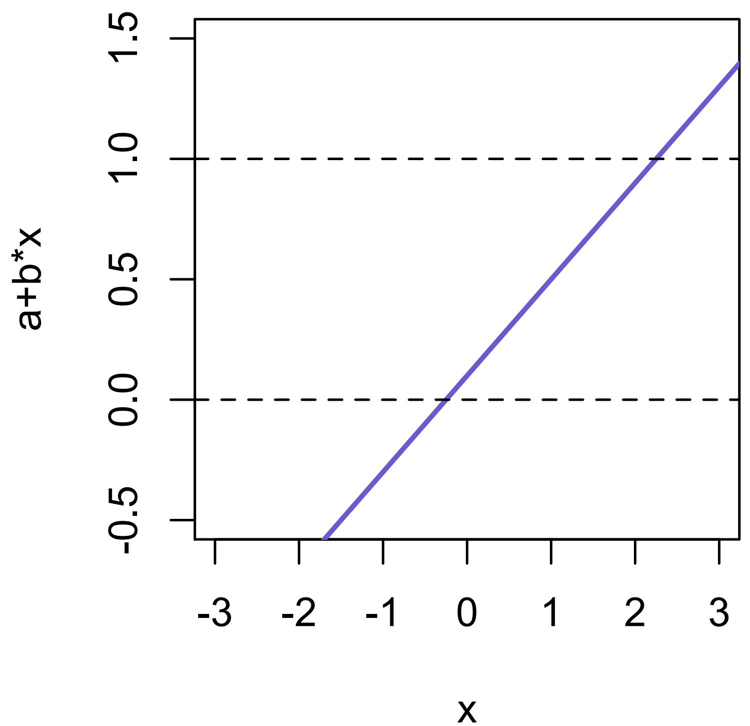
$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma), \\
\mu_i &= \alpha + \beta x_i.
\end{aligned}
$$

Generalized Linear Models

same units

$$y_i \sim \text{Normal}(\mu_i, \sigma),$$
$$\mu_i = \alpha + \beta x_i.$$

$$y_i \sim \text{Binomial}(n_i, p_i)$$
$$p_i \; ? \; \alpha + \beta x_i$$



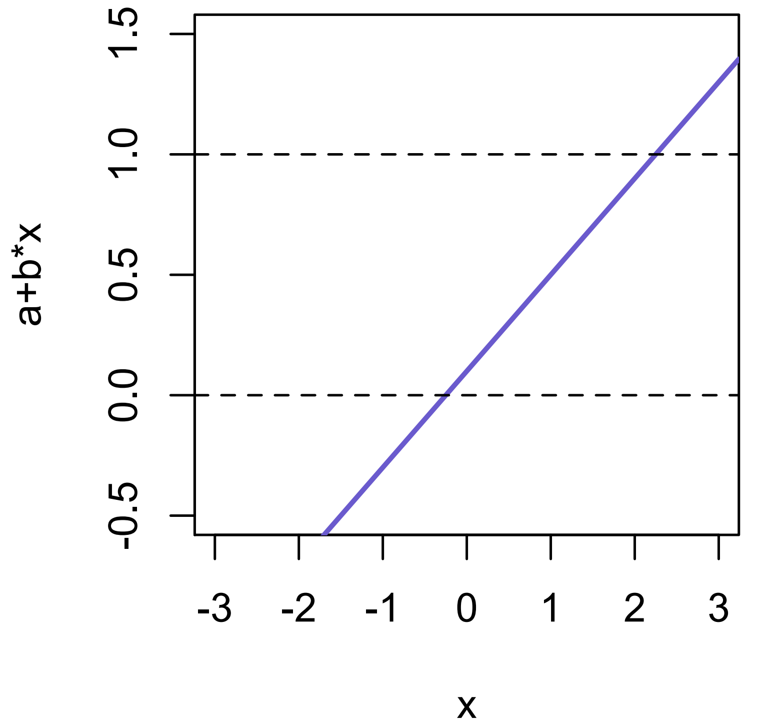
Generalized Linear Models

same units

$$y_i \sim \text{Normal}(\mu_i, \sigma),$$
$$\mu_i = \alpha + \beta x_i.$$

count

$$y_i \sim \text{Binomial}(n_i, p_i)$$
$$p_i \; ? \; \alpha + \beta x_i$$



Generalized Linear Models

same units

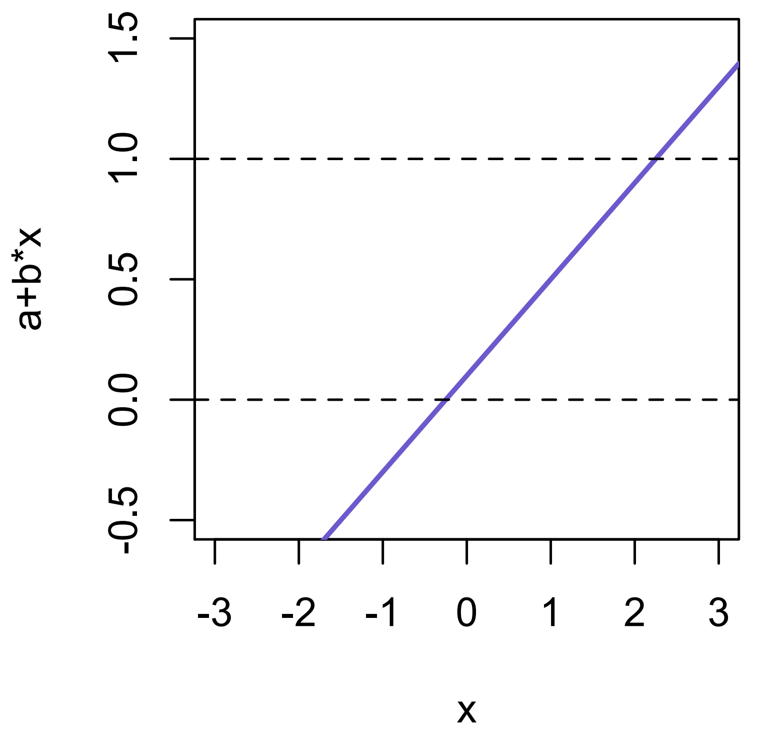
$$y_i \sim \text{Normal}(\mu_i, \sigma),$$
$$\mu_i = \alpha + \beta x_i.$$

count

$$y_i \sim \text{Binomial}(n_i, p_i)$$

probability

$$p_i \,?\, \alpha + \beta x_i$$



Generalized Linear Models

same units

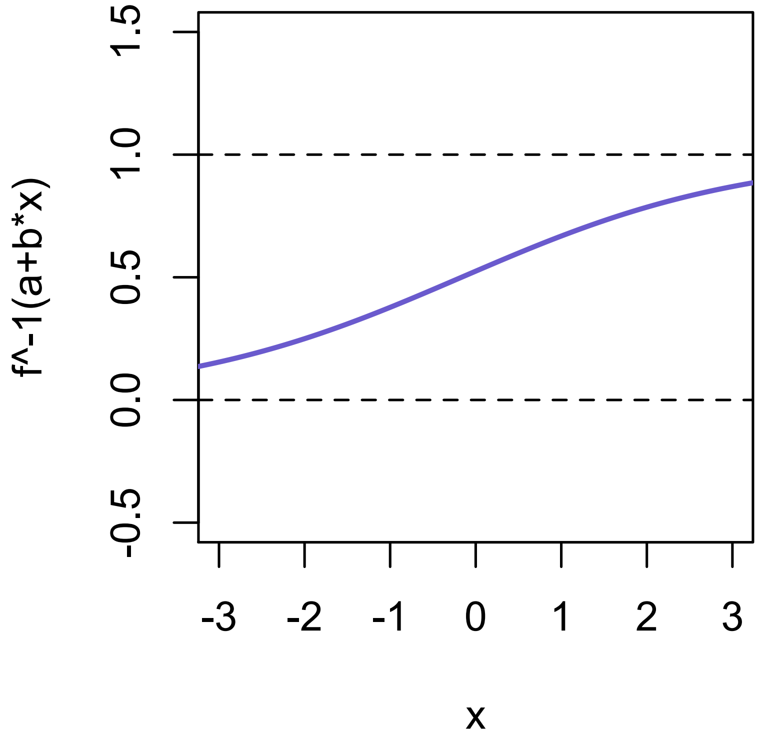
$$y_i \sim \mathrm{Normal}(\mu_i, \sigma),$$

$$\mu_i = \alpha + \beta x_i.$$

count

$$y_i \sim \mathrm{Binomial}(n_i, p_i)$$

$$f(p_i) = \alpha + \beta x_i$$



Generalized Linear Models

same units

$$y_i \sim \text{Normal}(\mu_i, \sigma),$$

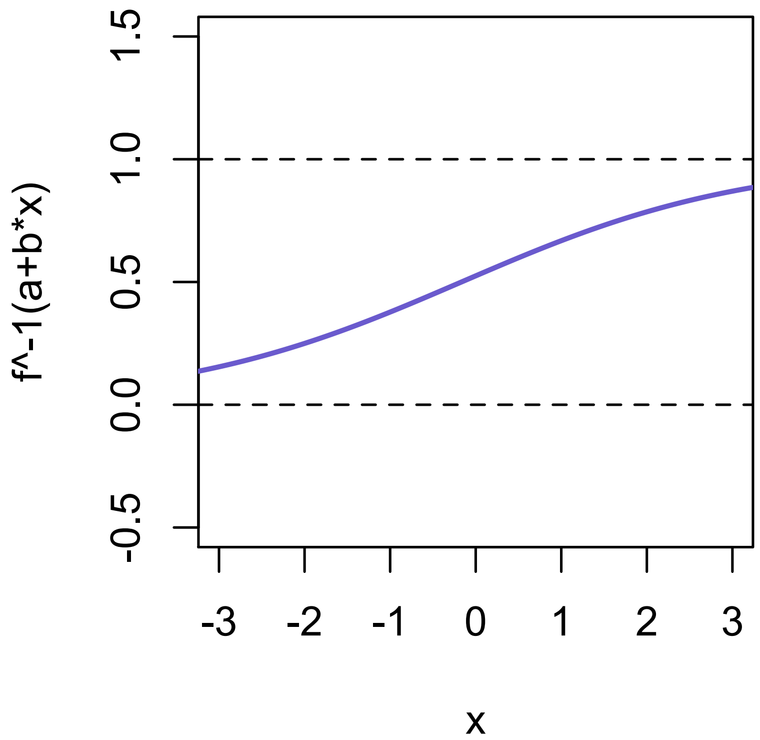
$$\mu_i = \alpha + \beta x_i.$$

count

$$y_i \sim \text{Binomial}(n_i, p_i)$$

$$f(p_i) = \alpha + \beta x_i$$

link function

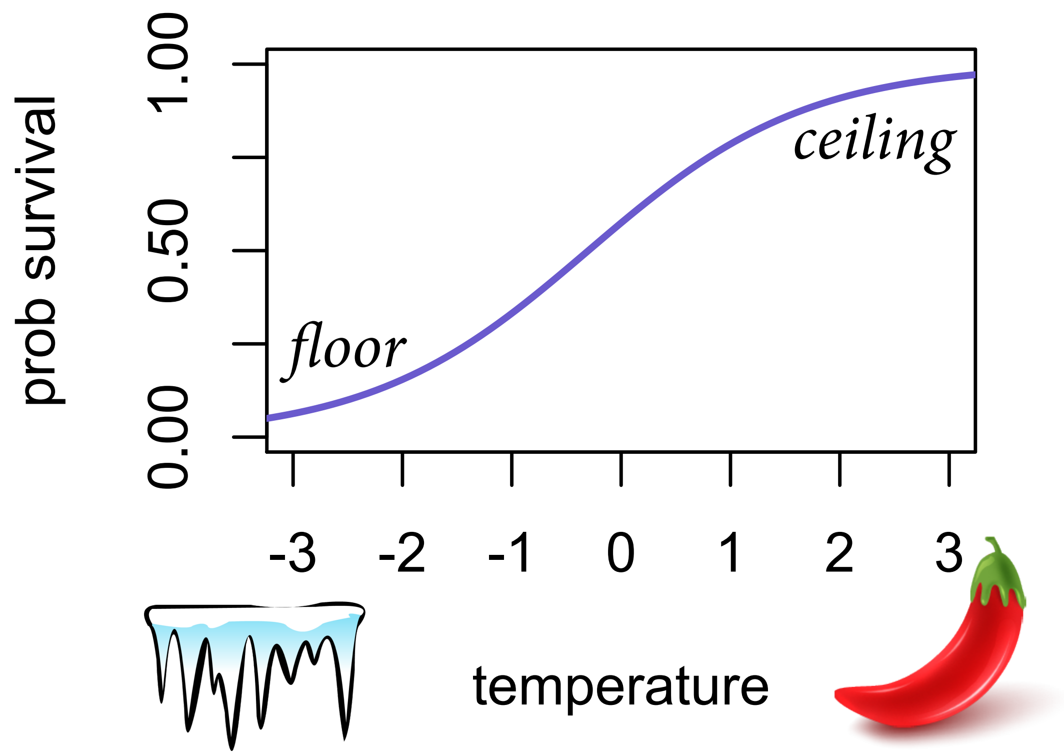


Generalized Linear Models

- (3) Compute posterior
  - Search is harder
  - Interpretation is harder
  - Links matter
  - Quadratic approximation often works, but not always
  - Safest to rely on MCMC

Everything interacts

- There are *floor* and *ceiling* effects



Everything interacts

- Linear regression:

$$\mu = \alpha + \beta x \qquad\qquad \partial\mu/\partial x = \beta.$$

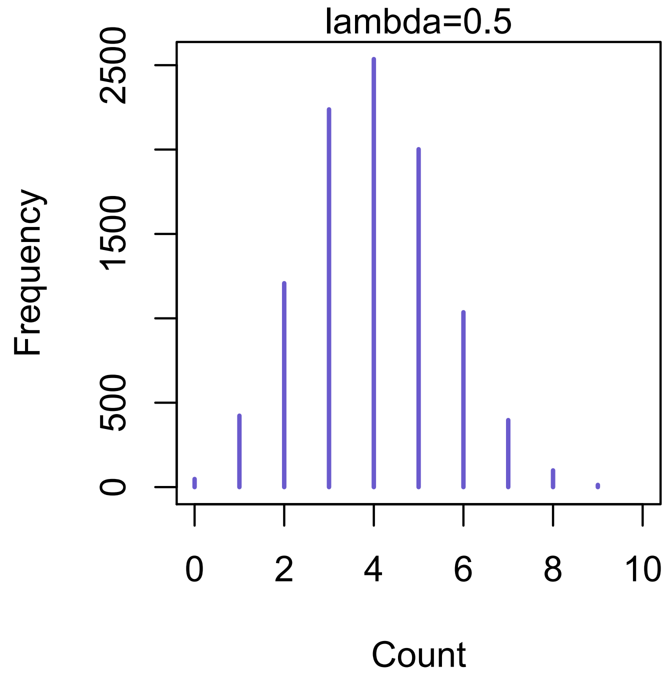
- Logistic regression:

$$p = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \qquad\qquad \frac{\partial p}{\partial x} = \frac{\beta}{2\big(1 + \cosh(\alpha + \beta x)\big)}$$

Binomial distribution

- Counts of a specific event out of n possibilities
- Constant expected value
- Maxent: Binomial

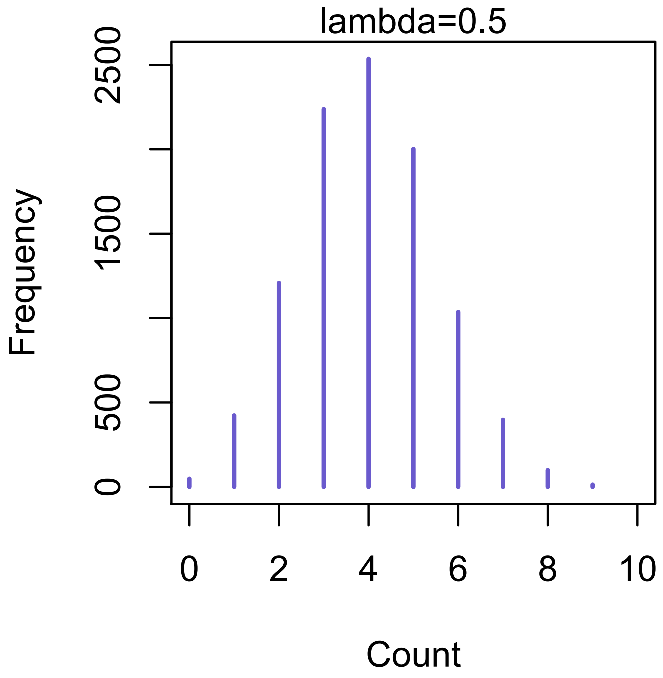
$$y \sim \text{Binomial}(n, p)$$



Binomial distribution

- Counts of a specific event out of n possibilities
- Constant expected value
- Maxent: Binomial

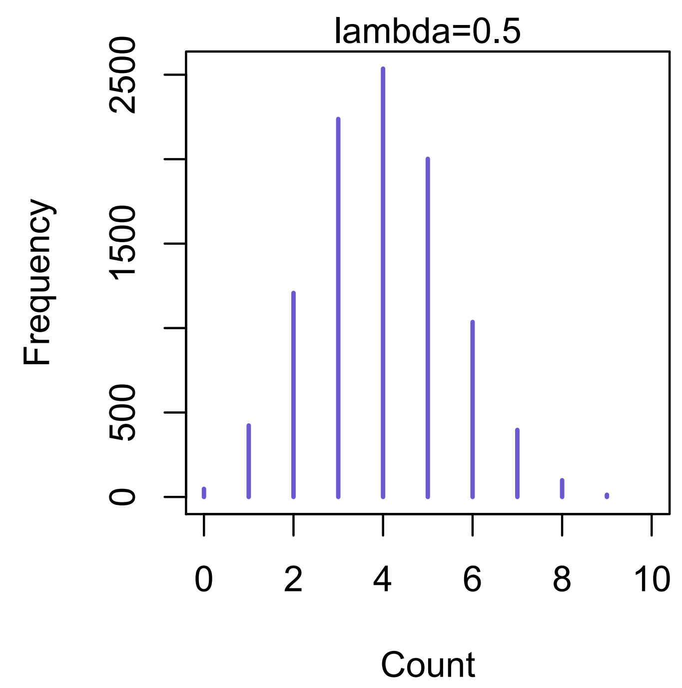
$$\underset{\textit{count "successes"}}{y} \sim \text{Binomial}(n, p)$$



Binomial distribution

- Counts of a specific event out of n possibilities
- Constant expected value
- Maxent: Binomial

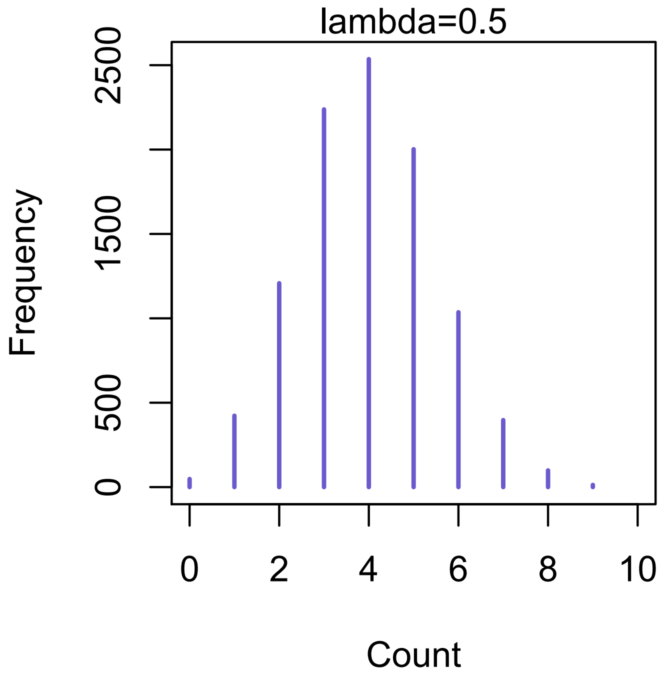
$$\underbrace{y}_{\textit{count “successes”}} \sim \text{Binomial}(\underbrace{n}_{\textit{number of trials}}, p)$$



Binomial distribution

- Counts of a specific event out of n possibilities
- Constant expected value
- Maxent: Binomial

$$\underbrace{y}_{\textit{count “successes”}} \sim \text{Binomial}(\underbrace{n}_{\textit{number of trials}}, \underbrace{p}_{\textit{probability of success}})$$



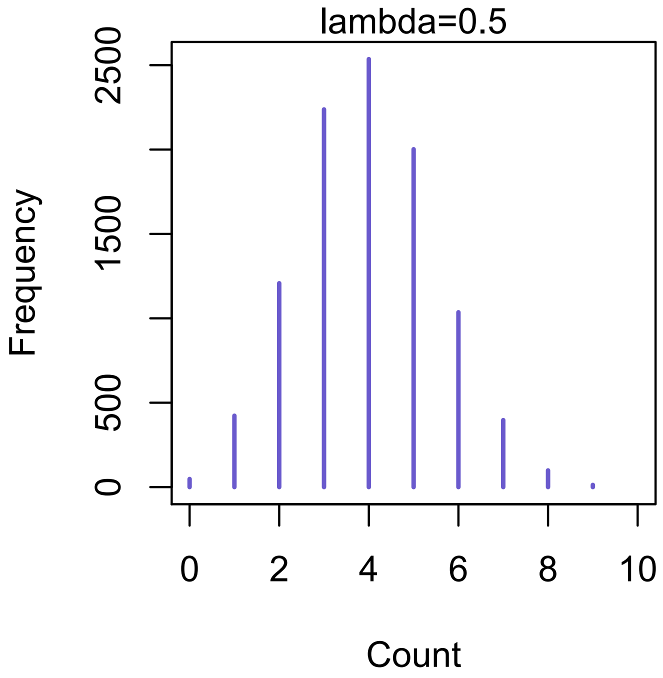
Binomial distribution

- Counts of a specific event out of n possibilities
- Constant expected value
- Maxent: Binomial

$$y \sim \text{Binomial}(n, p)$$

$$\text{E}(y) = np$$
$$\text{var}(y) = np(1 - p)$$

Mean and variance not independent



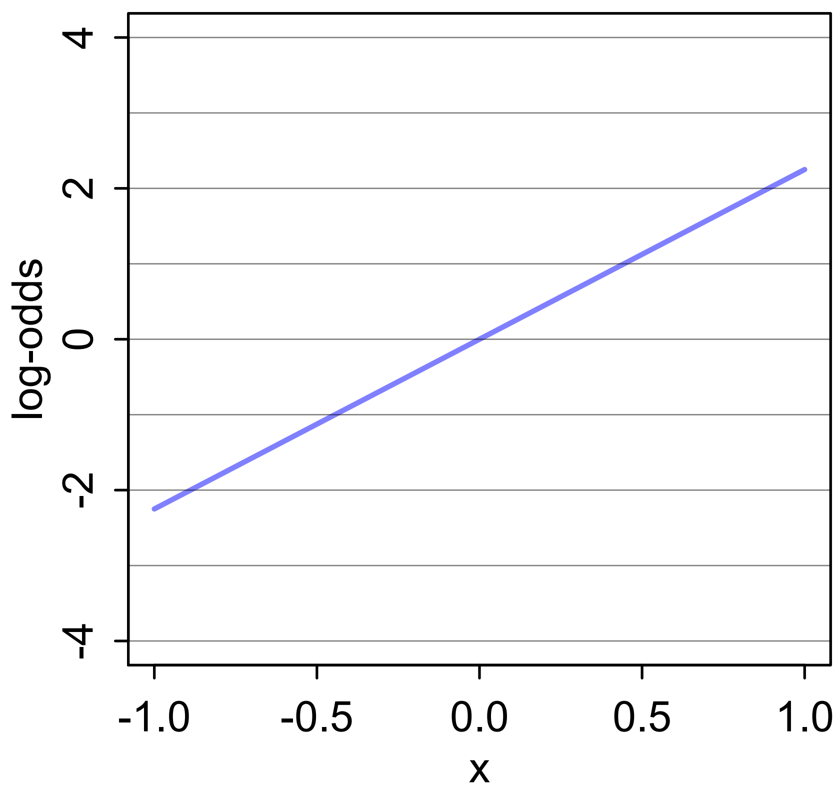
Need a link

$$y \sim \mathrm{Binomial}(n, p)$$

- y and p on different scales
- y: count
- p: probability
- Want to model p as function of predictor variables
- Must bound it to [0,1] interval

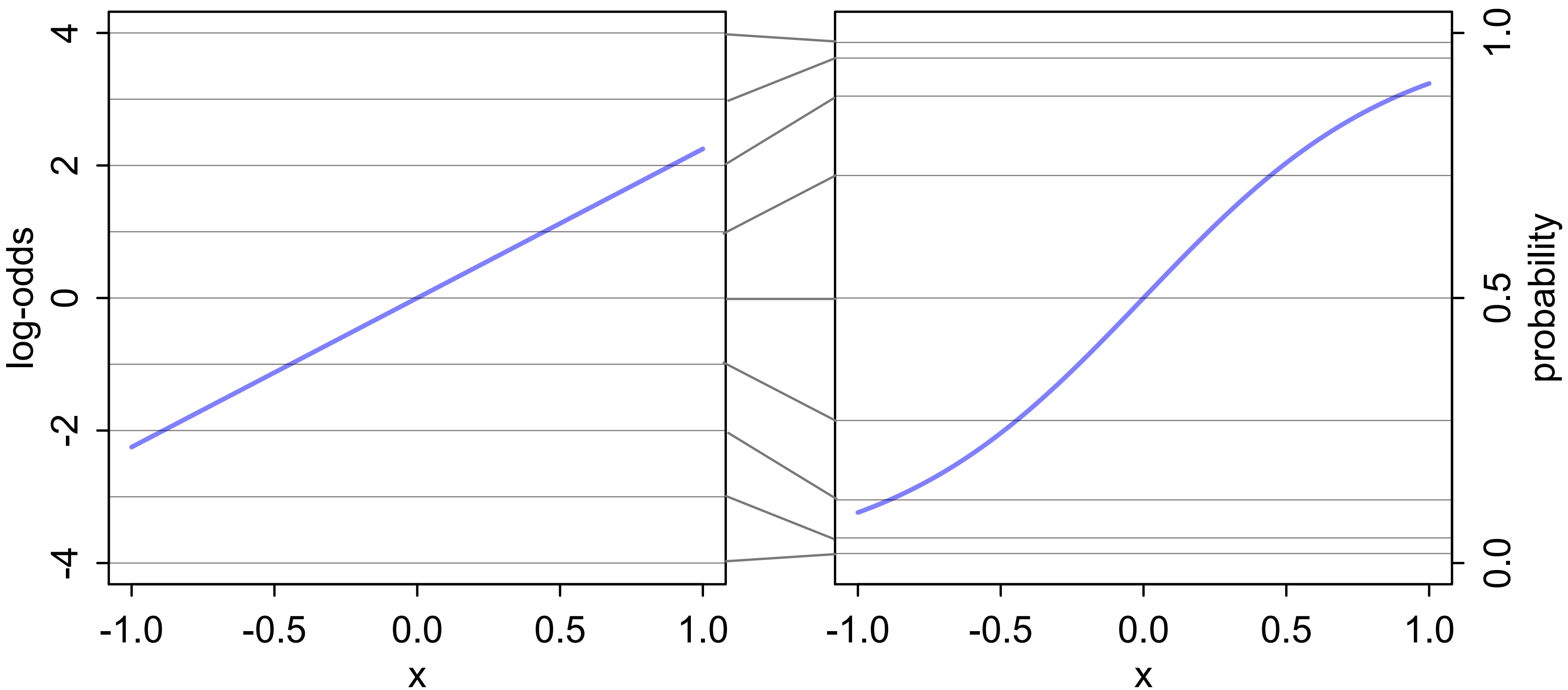
Logit link

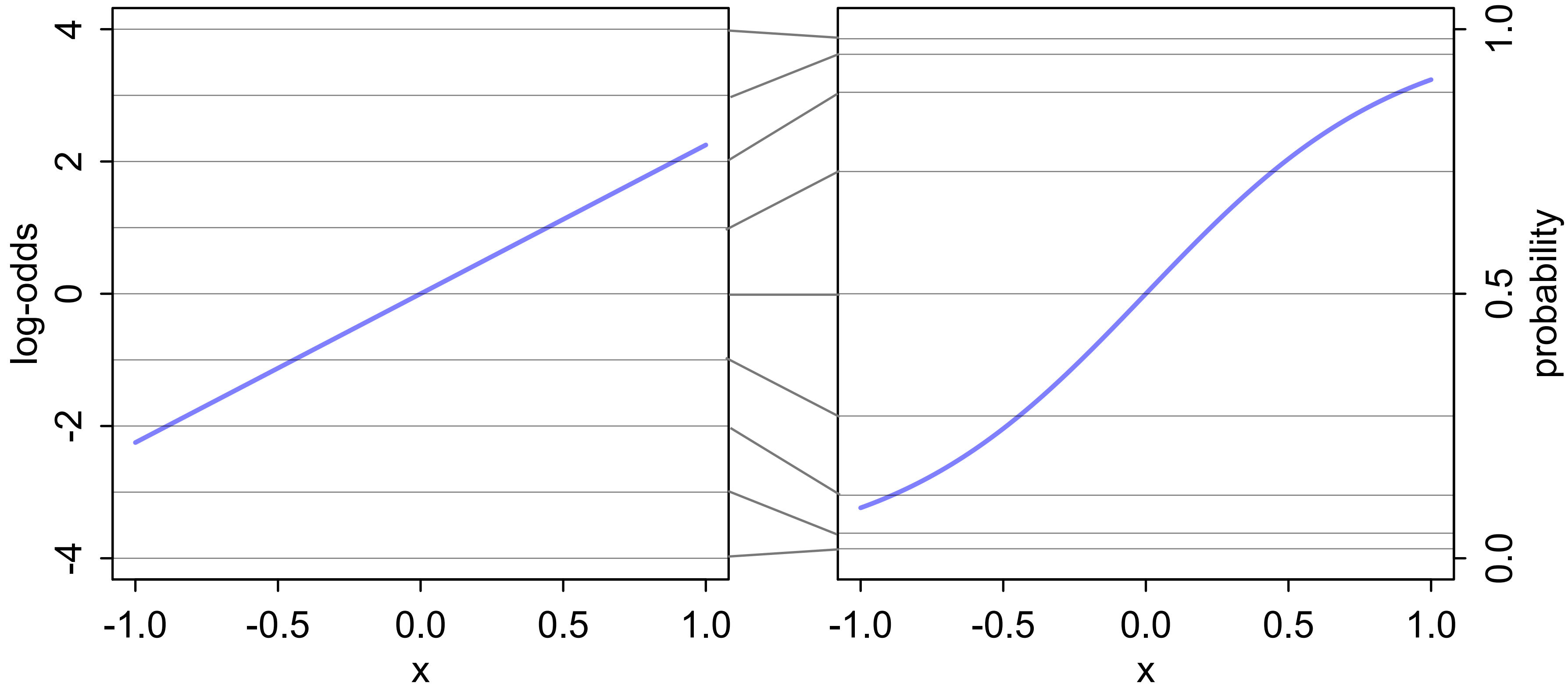
- Goal: map linear model to [0,1]



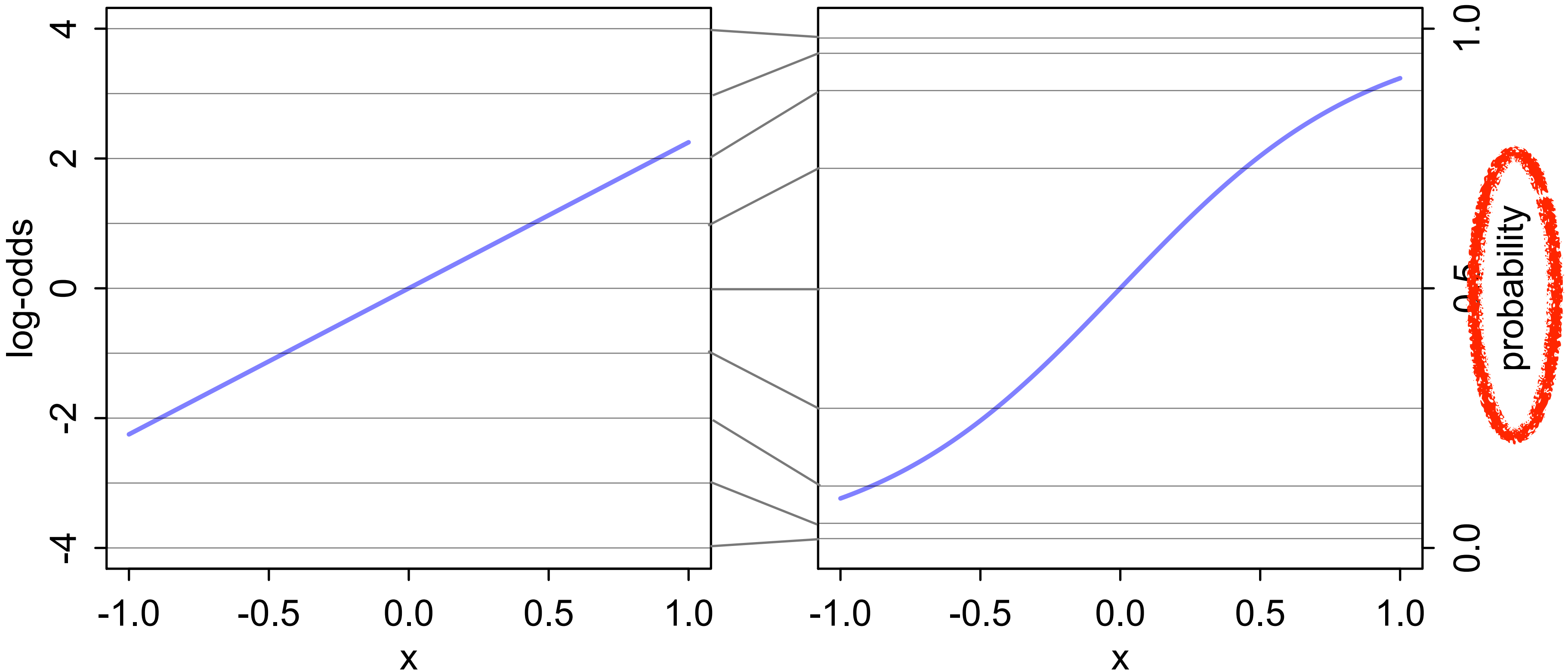
Logit link

- Goal: map linear model to [0,1]

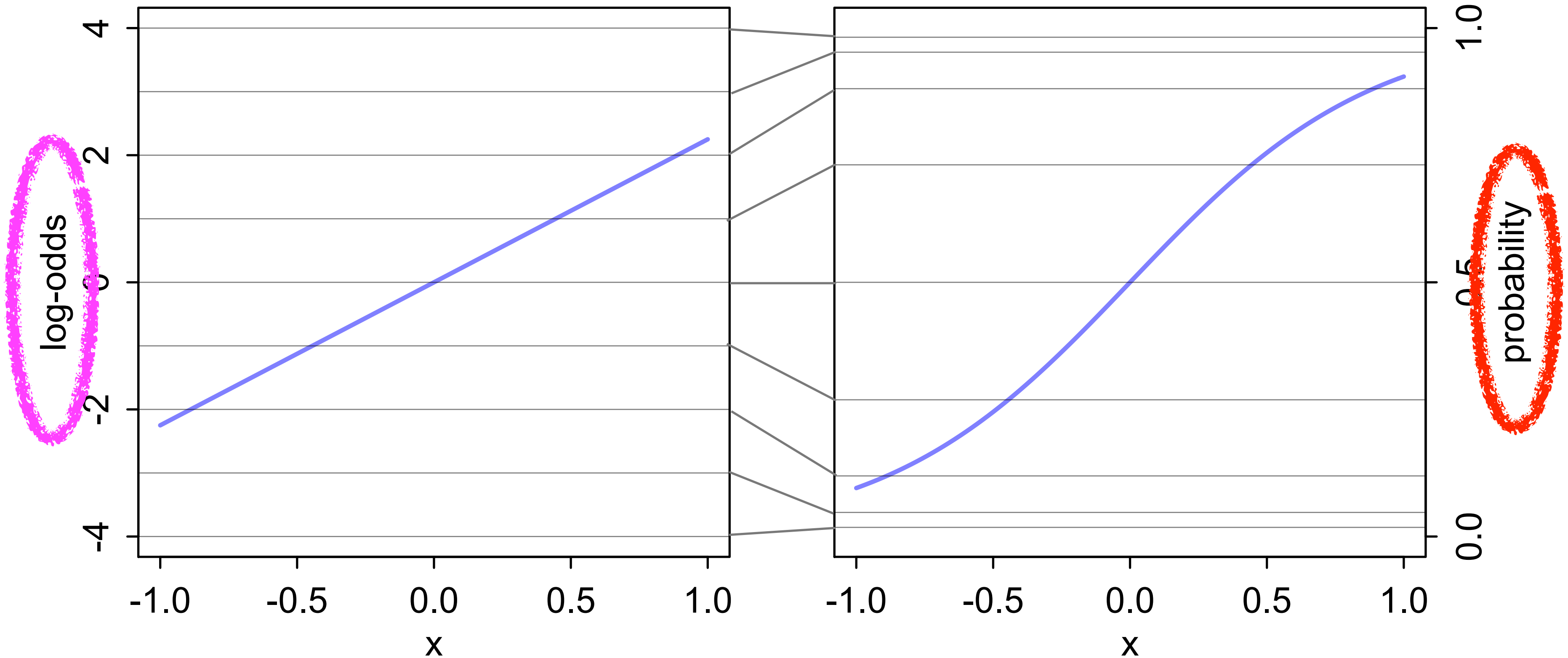




$$y_i \sim \text{Binomial}(n, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta x_i$$

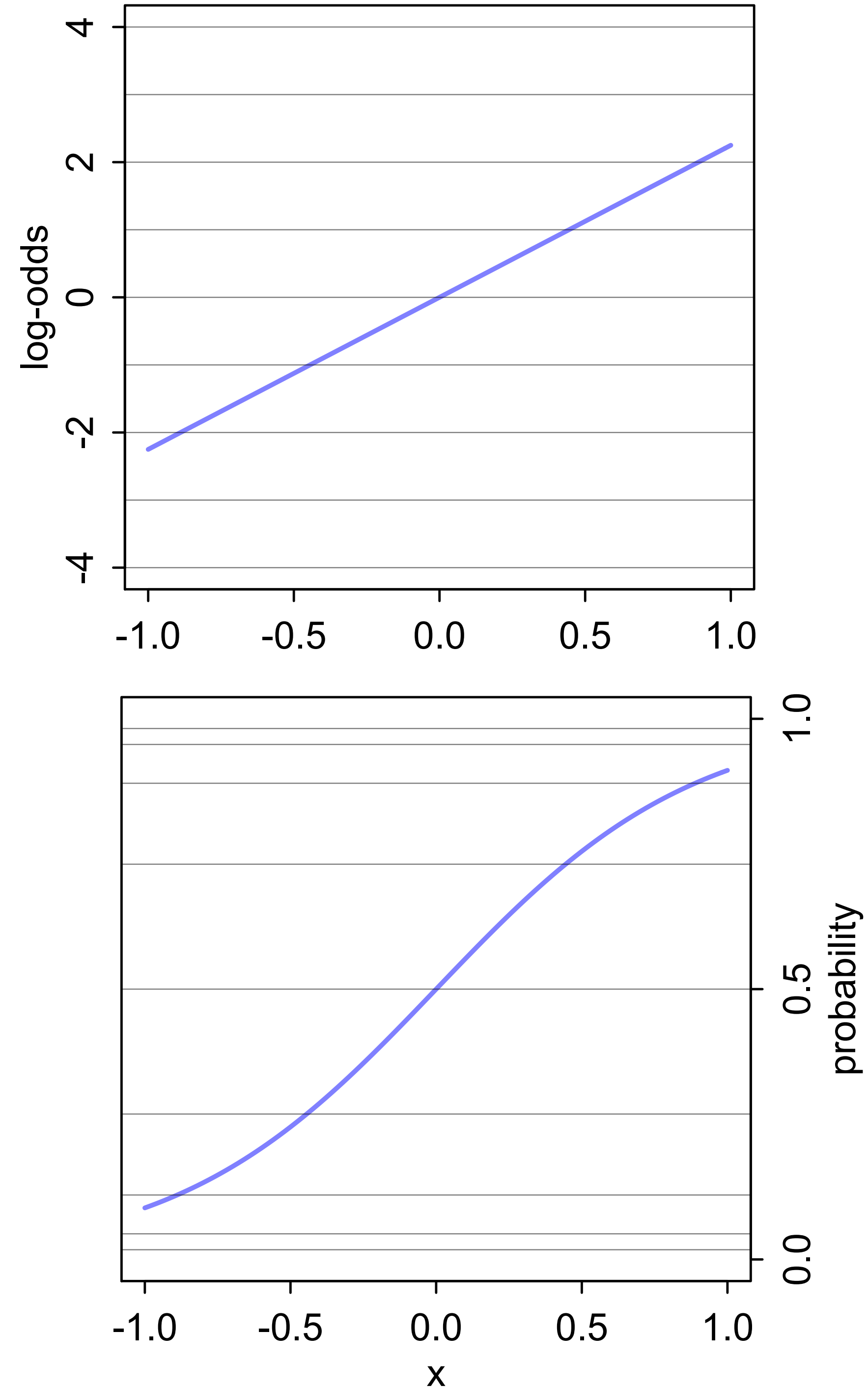


$$y_i \sim \text{Binomial}(n, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta x_i$$



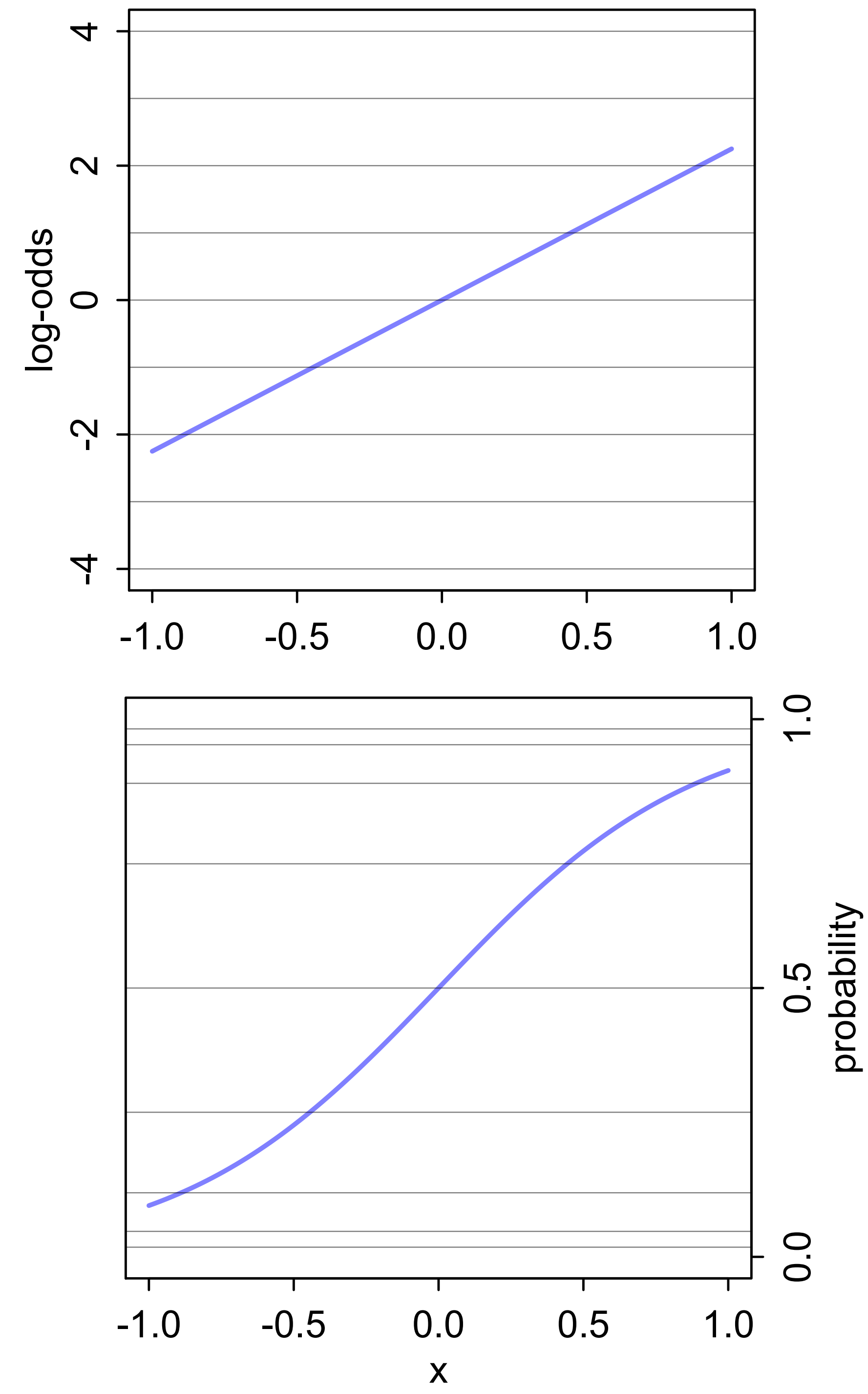
$$y_i \sim \text{Binomial}(n, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta x_i$$

$$y_i \sim \text{Binomial}(n, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta x_i$$



$$y_i \sim \text{Binomial}(n, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta x_i$$

$$\text{logit}(p_i) = \log \frac{p_i}{1 - p_i} = \alpha + \beta x_i$$



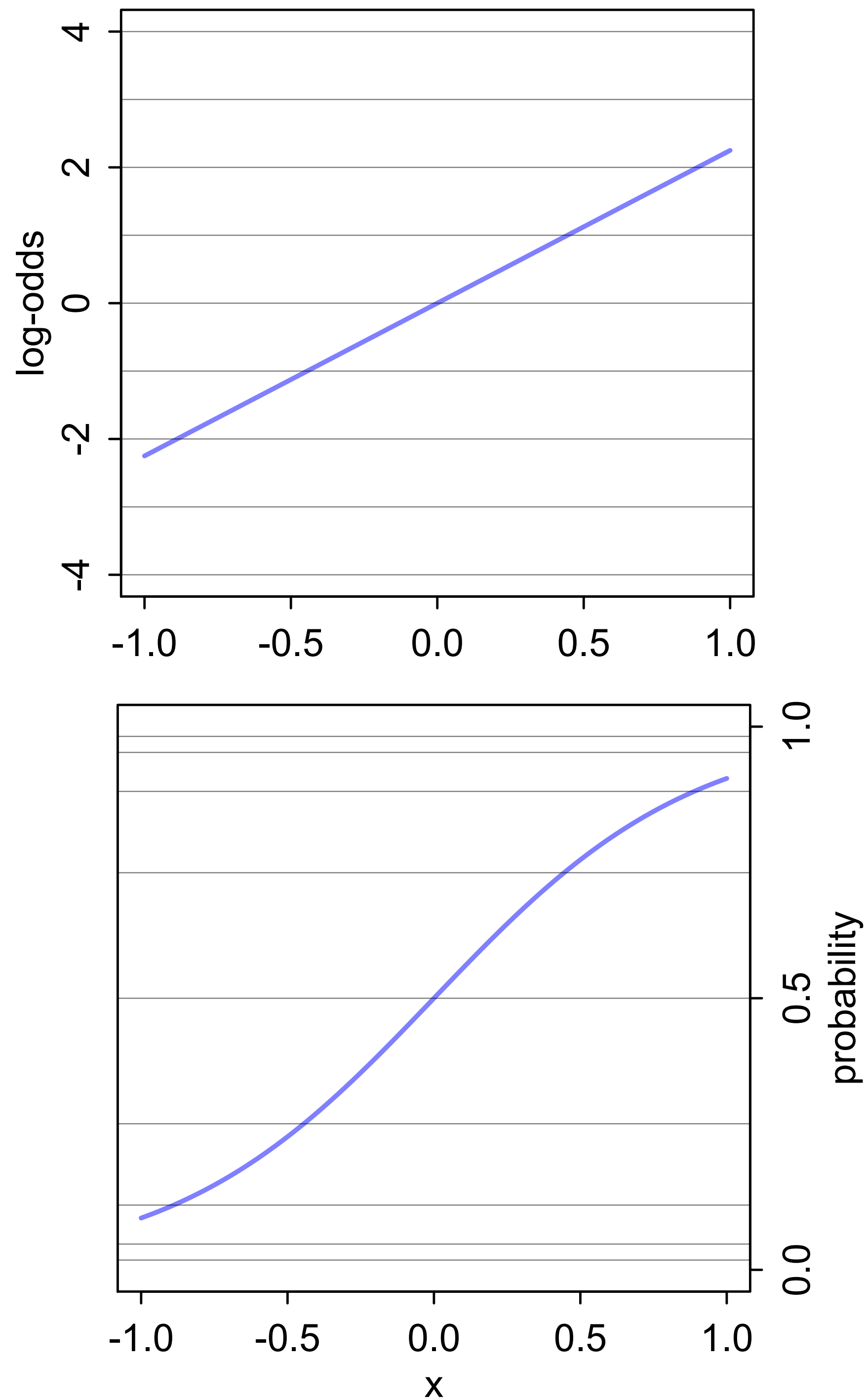
$$y_i \sim \text{Binomial}(n, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta x_i$$

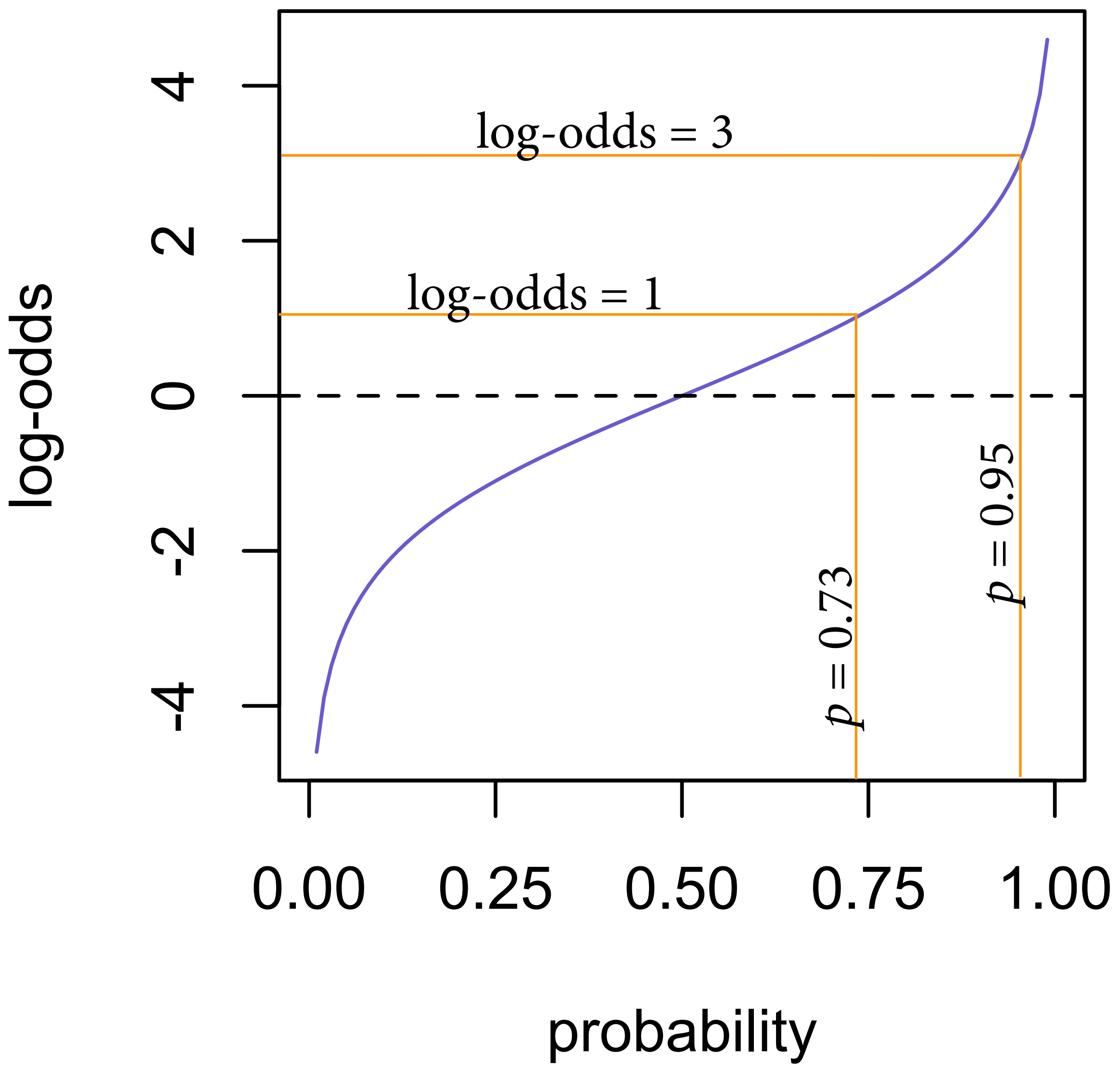
$$\text{logit}(p_i) = \log \frac{p_i}{1 - p_i} = \alpha + \beta x_i$$

Solve for p_i:

$$p_i = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$$

inverse-link is logistic





log-odds = 3
log-odds = 1
p = 0.73
p = 0.95
4
2
0
-2
-4
log-odds
0.00
0.25
0.50
0.75
1.00
probability

Logit link

$$y_i \sim \text{Binomial}(n, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta x_i$$

- Where does this thing come from?
- Several good answers:
  - “Natural” link inside probability formula
  - log-odds is fundamental parameter
  - See Overthinking box, pages 279–280
- Other links sometimes justified
  - Probit (common in economics)
  - Complementary-log-log (cloglog, common where?)

Example: UCB admissions

R code 10.22

```
library(rethinking)
data(UCBadmit)
d <- UCBadmit
```

- Numbers accepted/rejected to 6 PhD programs at UC Berkeley (largest depts in 1973)
- Evidence of gender discrimination? Dean was afraid of lawsuit.
- Call in the statisticians!

Example: UCB admissions

R code 10.22

```
library(rethinking)
data(UCBadmit)
d <- UCBadmit
```

	dept	applicant.gender	admit	reject	applications
1	A	male	512	313	825
2	A	female	89	19	108
3	B	male	353	207	560
4	B	female	17	8	25
5	C	male	120	205	325
6	C	female	202	391	593
7	D	male	138	279	417
8	D	female	131	244	375
9	E	male	53	138	191
10	E	female	94	299	393
11	F	male	22	351	373
12	F	female	24	317	341

Trials vary by row

```
d$male <- ifelse( d$applicant.gender=="male" , 1 , 0 )
m10.6 <- map(
    alist(
        admit ~ dbinom( applications , p ) ,
        logit(p) <- a + bm*male ,
        a ~ dnorm(0,10) ,
        bm ~ dnorm(0,10)
    ) ,
    data=d )
```

R code 10.23

	dept	applicant.gender	admit	reject	applications
1	A	male	512	313	825
2	A	female	89	19	108
3	B	male	353	207	560
4	B	female	17	8	25
5	C	male	120	205	325
6	C	female	202	391	593
7	D	male	138	279	417
8	D	female	131	244	375
9	E	male	53	138	191
10	E	female	94	299	393
11	F	male	22	351	373
12	F	female	24	317	341

$$n_{\text{admit},i} \sim \text{Binomial}(n_i, p_i)$$
$$\text{logit}(p_i) = \alpha + \beta_m m_i$$
$$\alpha \sim \text{Normal}(0, 10)$$
$$\beta_m \sim \text{Normal}(0, 10)$$

With and without males

R code 10.23

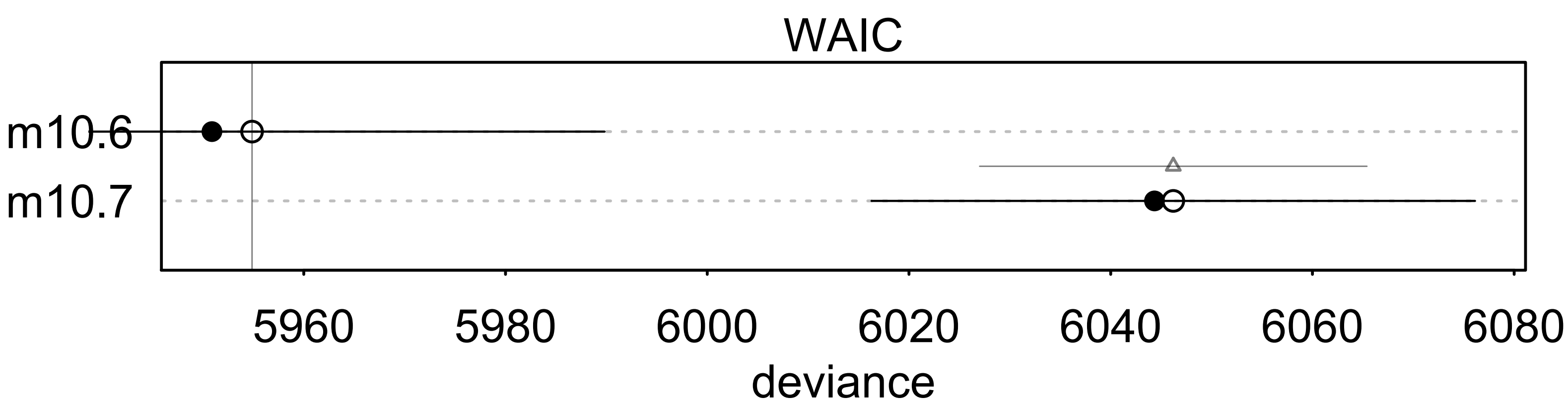
```
d$male <- ifelse( d$applicant.gender=="male" , 1 , 0 )
m10.6 <- map(
    alist(
        admit ~ dbinom( applications , p ) ,
        logit(p) <- a + bm*male ,
        a ~ dnorm(0,10) ,
        bm ~ dnorm(0,10)
    ) ,
    data=d )
m10.7 <- map(
    alist(
        admit ~ dbinom( applications , p ) ,
        logit(p) <- a ,
        a ~ dnorm(0,10)
    ) ,
    data=d )
```

Compare

```
compare( m10.6 , m10.7 )
```

R code 10.24

```
         WAIC pWAIC dWAIC weight    SE   dSE
m10.6 5954.9     2   0.0      1 34.98    NA
m10.7 6046.3     1  91.5      0 29.93 19.13
```



Proportional change in odds

- How to interpret these coefficients?
- exp(estimate) gives proportional change in odds
- Is a *relative* effect size

```r
precis(m10.6)
```

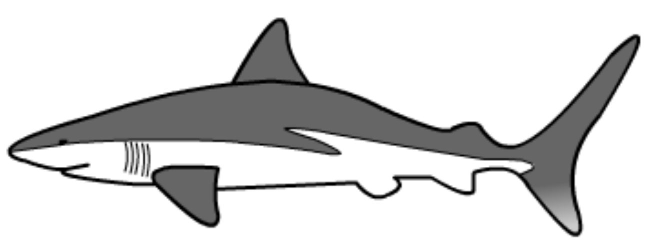
R code 10.25

```
    Mean StdDev  5.5% 94.5%
a  -0.83   0.05 -0.91 -0.75
bm  0.61   0.06  0.51  0.71
```

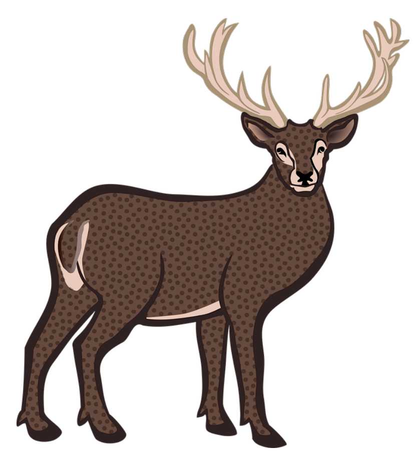
exp(0.61) ≈ 1.84 => male has 184% odds of female

Relative and absolute effects

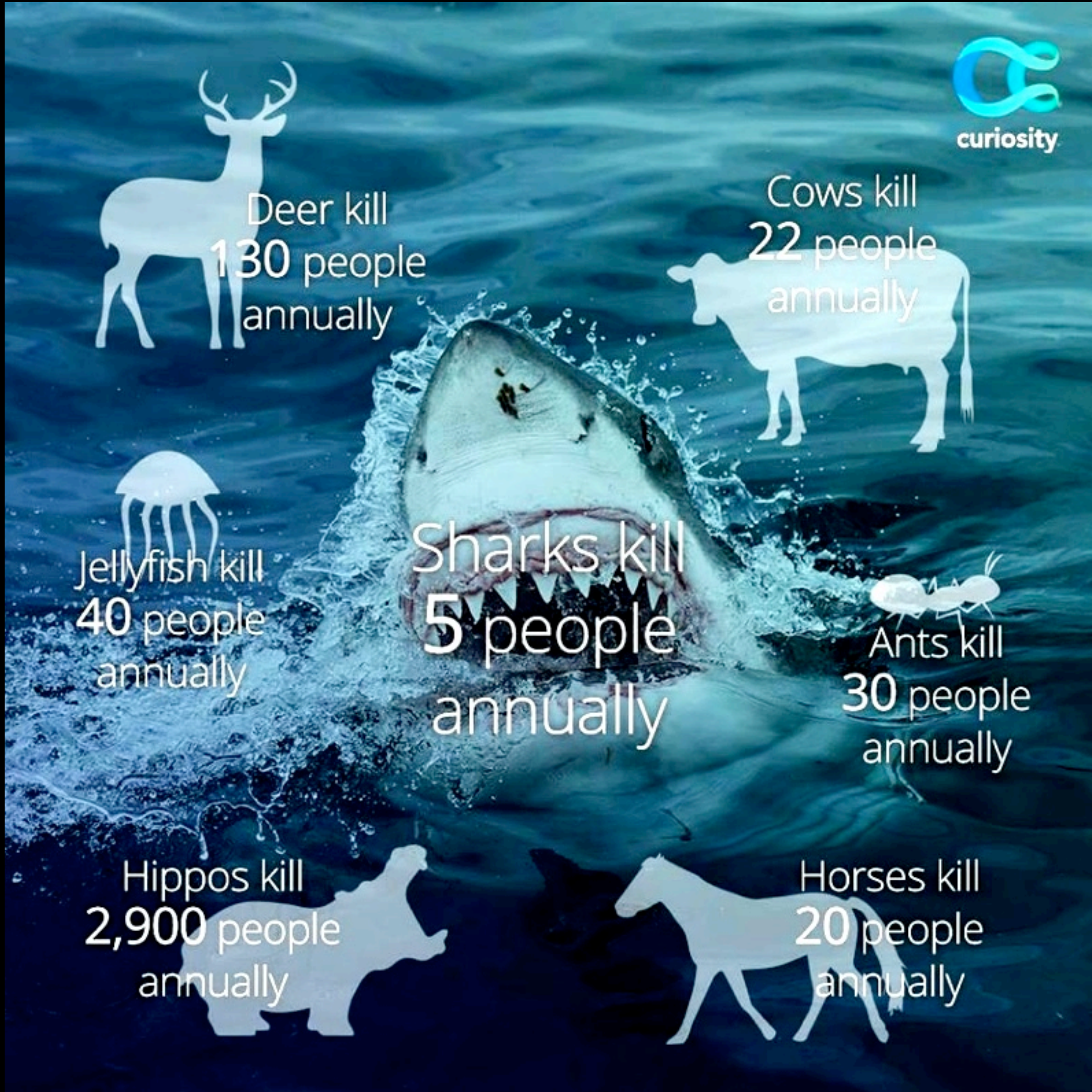
- Parameters on *relative* effect scale
- Predictions on *absolute* effect scale
- Using relative effects may exaggerate importance of predictor
  - Good for scaring people, getting published
  - Not so good for public health, scientific progress
  - But needed for causal inference



relative shark



absolute deer


curiosity
Deer kill 130 people annually
Cows kill 22 people annually
Jellyfish kill 40 people annually
Sharks kill 5 people annually
Ants kill 30 people annually
Hippos kill 2,900 people annually
Horses kill 20 people annually

Risk communication

- Many people mistake relative risk for absolute risk
- Example:
    - 1/1000 women develop blood clots
    - 3/1000 women on birth control develop blood clots
    - => 200% increase in blood clots!
    - Change in probability is only 0.002
    - Pregnancy much more dangerous than blood clots

Daily Mail .com

Home | U.K. | News | Sports | U.S. Showbiz | Australia | Fem

News Home | Arts | Headlines | Pictures | Most read | News Board | Wire

EXCLUSIVE: Grandfather of 'the

ISIS attacks Iraqi base where 320 US

Has Homer Sir actually been i

Deadly risk of pill used by 1m w GP in Britain told to warn abou popular contraceptive

- Bestselling brands of birth control tablets linked to
- They are believed to double the risk compared to ol
- 'Third-generation' contraceptives caused 14 deaths
- UK doctors have been ordered to alert women to the

Compute probabilities

- *Absolute* effect size is on outcome scale:

```
precis(m10.6)
```

R code 10.25

```
    Mean StdDev  5.5% 94.5%
a  -0.83   0.05 -0.91 -0.75
bm  0.61   0.06  0.51  0.71
```

$$
\begin{aligned}
n_{\text{admit},i} &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha + \beta_m m_i \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta_m &\sim \text{Normal}(0, 10)
\end{aligned}
$$

```
logistic( -0.83 )
logistic( -0.83 + 0.61 )
```

```
[1] 0.3036451
[1] 0.4452208
```

Compute probabilities

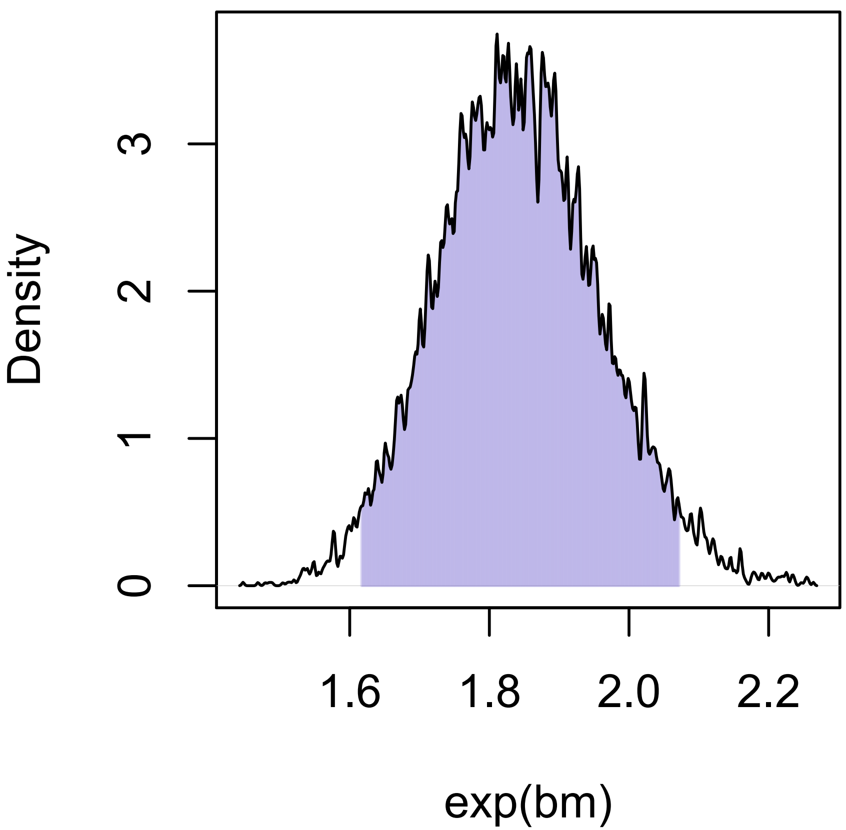
- Compute the contrast (difference in probability of admission):

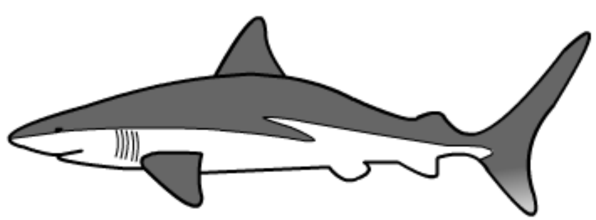
R code 10.26

```
post <- extract.samples( m10.6 )
p.admit.male <- logistic( post$a + post$bm )
p.admit.female <- logistic( post$a )
diff.admit <- p.admit.male - p.admit.female
quantile( diff.admit , c(0.025,0.5,0.975) )
```

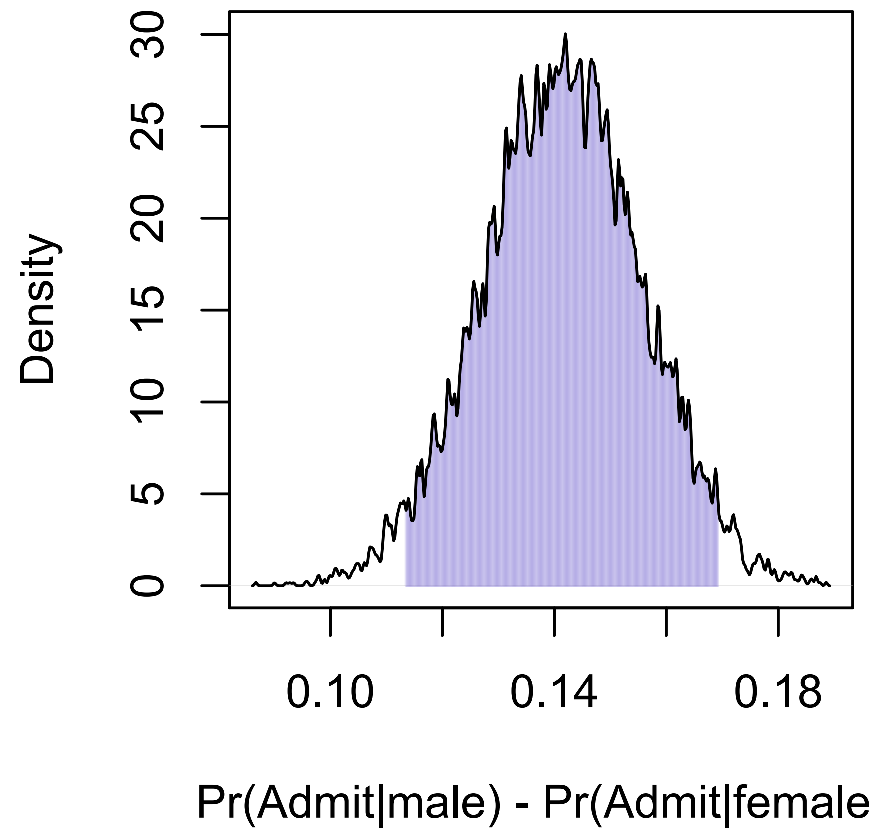
```
     2.5%       50%     97.5%
0.1132778 0.1413527 0.1693274
```

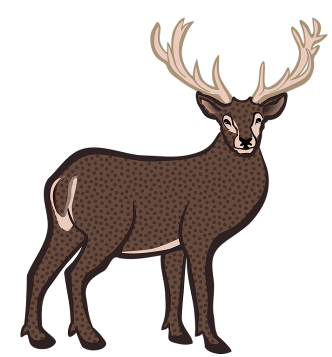
Odds ratios
(relative risk)

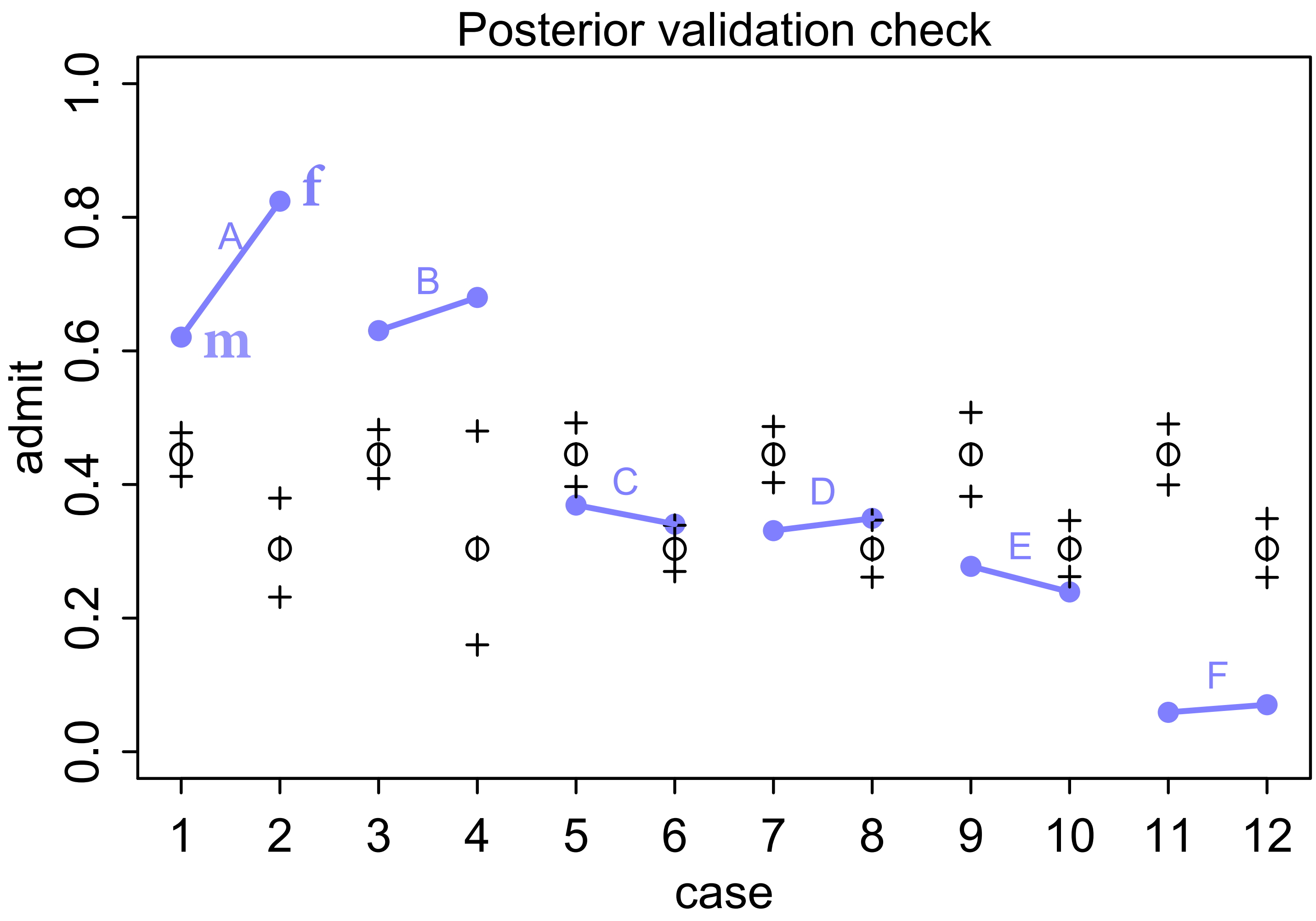




Probability
(difference in absolute risk)







Females admitted more in all but 2 departments!

Figure 10.5

Departments vary

- Overall admission rates vary a lot across departments
- Use unique intercepts to control for that variation

$$
\begin{aligned}
n_{\text{admit},i} &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{DEPT}[i]} + \beta_m m_i \\
\alpha_{\text{DEPT}} &\sim \text{Normal}(0, 10) \\
\beta_m &\sim \text{Normal}(0, 10)
\end{aligned}
$$

Departments vary

$$
\begin{aligned}
n_{\text{admit},i} &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha + \beta_m m_i \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta_m &\sim \text{Normal}(0, 10)
\end{aligned}
$$

Q: What are the average probabilities of admission for females and males across all departments?

$$
\begin{aligned}
n_{\text{admit},i} &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{DEPT}[i]} + \beta_m m_i \\
\alpha_{\text{DEPT}} &\sim \text{Normal}(0, 10) \\
\beta_m &\sim \text{Normal}(0, 10)
\end{aligned}
$$

Q: What is the average difference in probability of admission for females and males within departments?

Departments vary

	dept	dept_id
1	A	1
2	A	1
3	B	2
4	B	2
5	C	3
6	C	3
7	D	4
8	D	4
9	E	5
10	E	5
11	F	6
12	F	6

```
d$dept_id <- coerce_index( d$dept )
m10.8 <- map(
    alist(
        admit ~ dbinom( applications , p ) ,
        logit(p) <- a[dept_id] ,
        a[dept_id] ~ dnorm(0,10)
    ) ,
    data=d )
m10.9 <- map(
    alist(
        admit ~ dbinom( applications , p ) ,
        logit(p) <- a[dept_id] + bm*male ,
        a[dept_id] ~ dnorm(0,10) ,
        bm ~ dnorm(0,10)
    ) ,
    data=d )
```

Departments vary

R code 10.29

```
compare( m10.6 , m10.7 , m10.8 , m10.9 )
```

```
        WAIC pWAIC dWAIC weight    SE   dSE
m10.8 5200.9     6   0.0   0.56 57.02    NA
m10.9 5201.4     7   0.5   0.44 57.06  2.48
m10.6 5954.8     2 753.9   0.00 34.98 48.53
m10.7 6046.3     1 845.4   0.00 29.95 52.37
```

Departments vary

R code 10.29

```
compare( m10.6 , m10.7 , m10.8 , m10.9 )
```

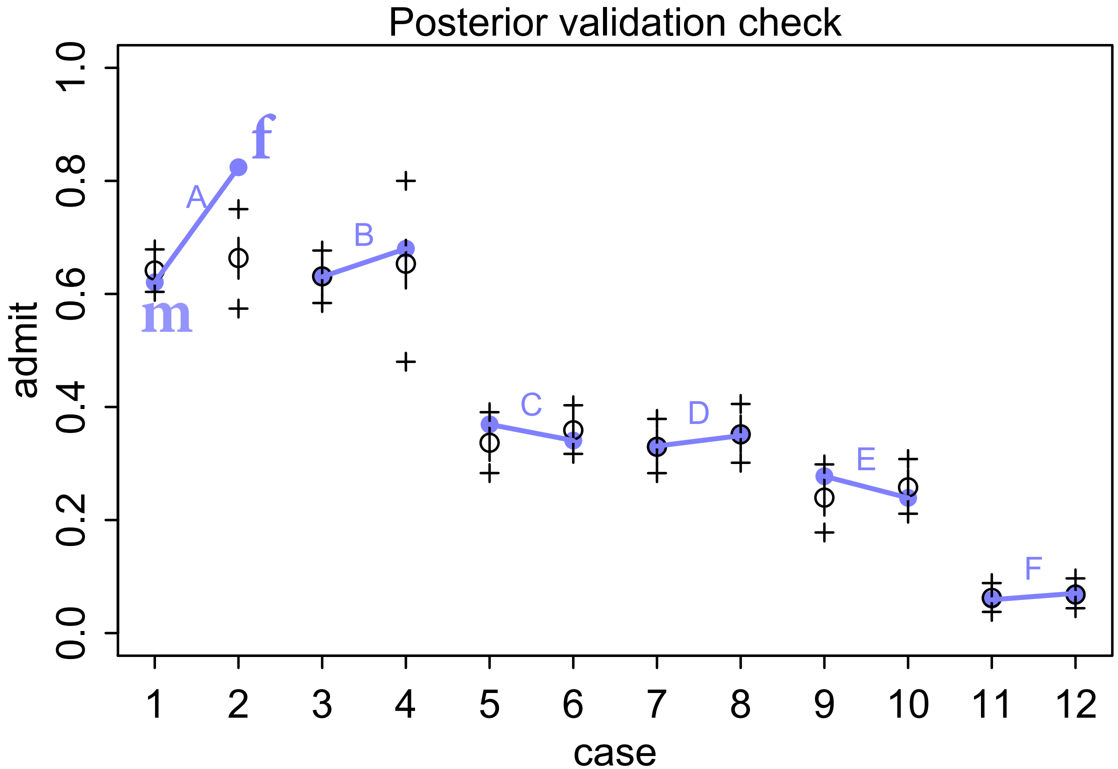
```
        WAIC pWAIC dWAIC weight    SE   dSE
m10.8 5200.9     6   0.0   0.56 57.02    NA
m10.9 5201.4     7   0.5   0.44 57.06  2.48
m10.6 5954.8     2 753.9   0.00 34.98 48.53
m10.7 6046.3     1 845.4   0.00 29.95 52.37
```

R code 10.30

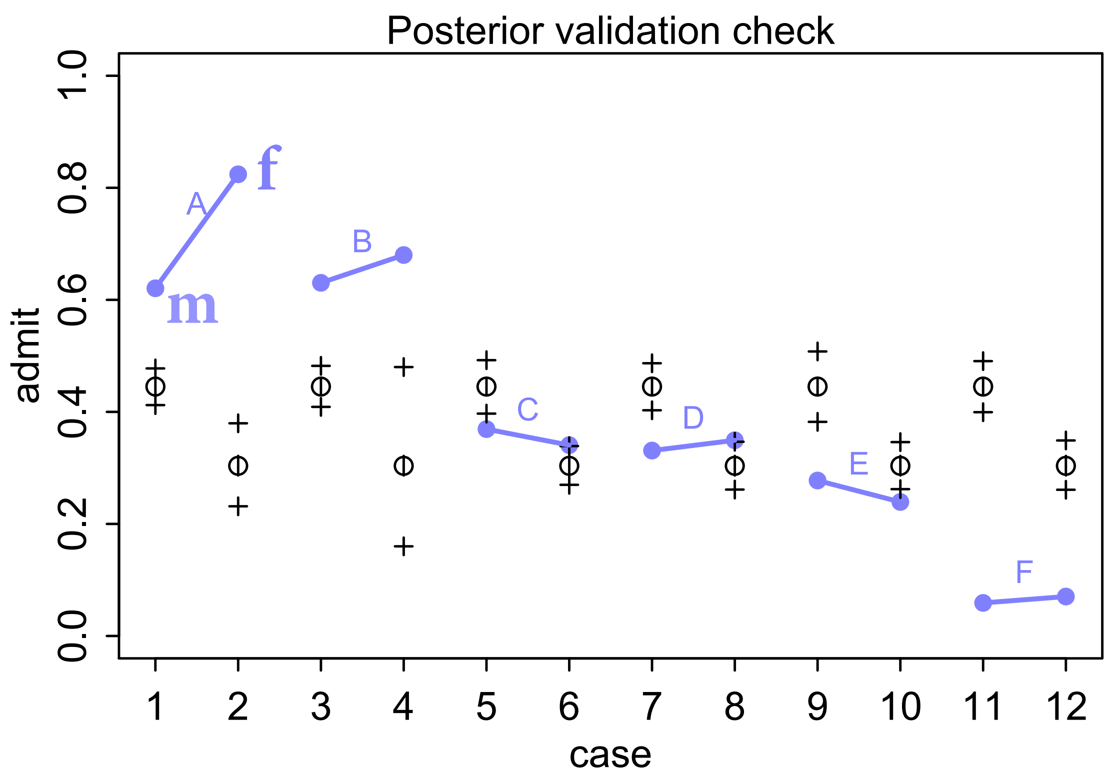
```
precis( m10.9 , depth=2 )
```

```
       Mean StdDev  5.5% 94.5%
a[1]   0.68   0.10  0.52  0.84
a[2]   0.64   0.12  0.45  0.82
a[3]  -0.58   0.07 -0.70 -0.46
a[4]  -0.61   0.09 -0.75 -0.48
a[5]  -1.06   0.10 -1.22 -0.90
a[6]  -2.62   0.16 -2.88 -2.37
bm    -0.10   0.08 -0.23  0.03
```

With dummies

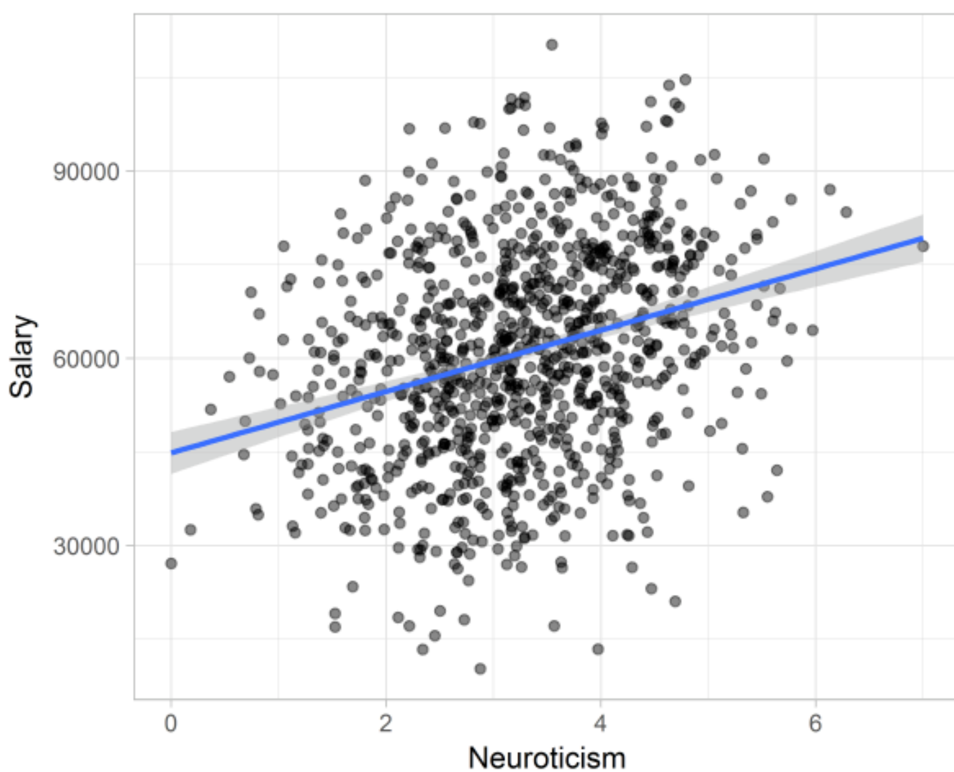


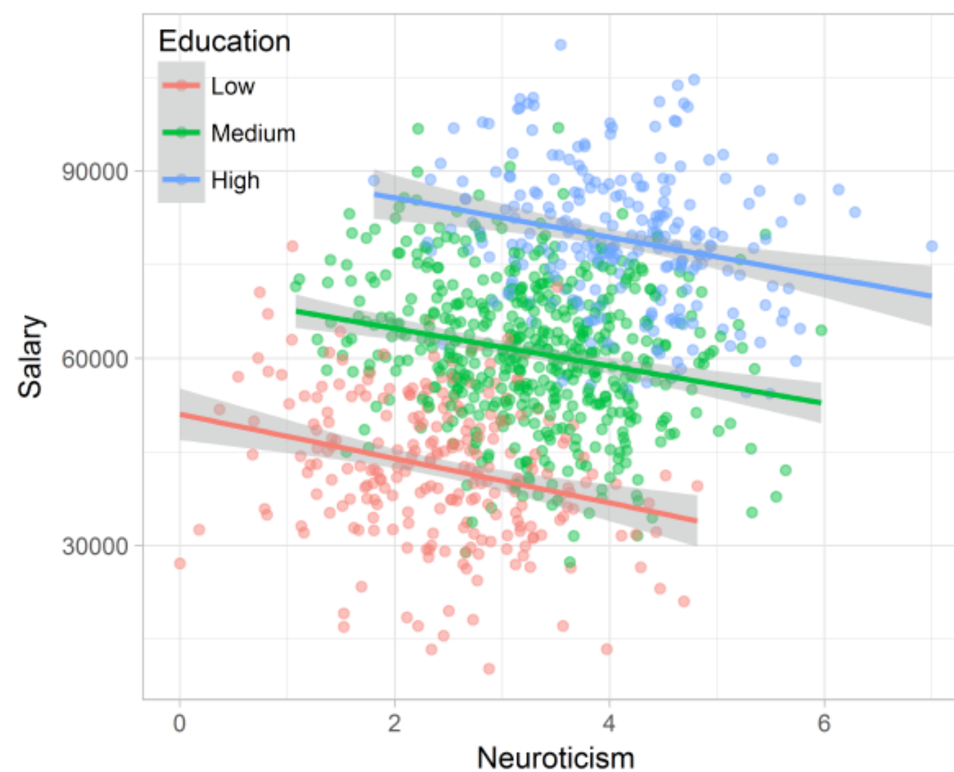
Without



Simpson's Paradox

- Trend reverses when additional predictor added
- Can indicate confound => win!
- Can also indicate collider => lose!





https://paulvanderlaken.com/2017/09/27/simpsons-paradox-two-hr-examples-with-r-code/