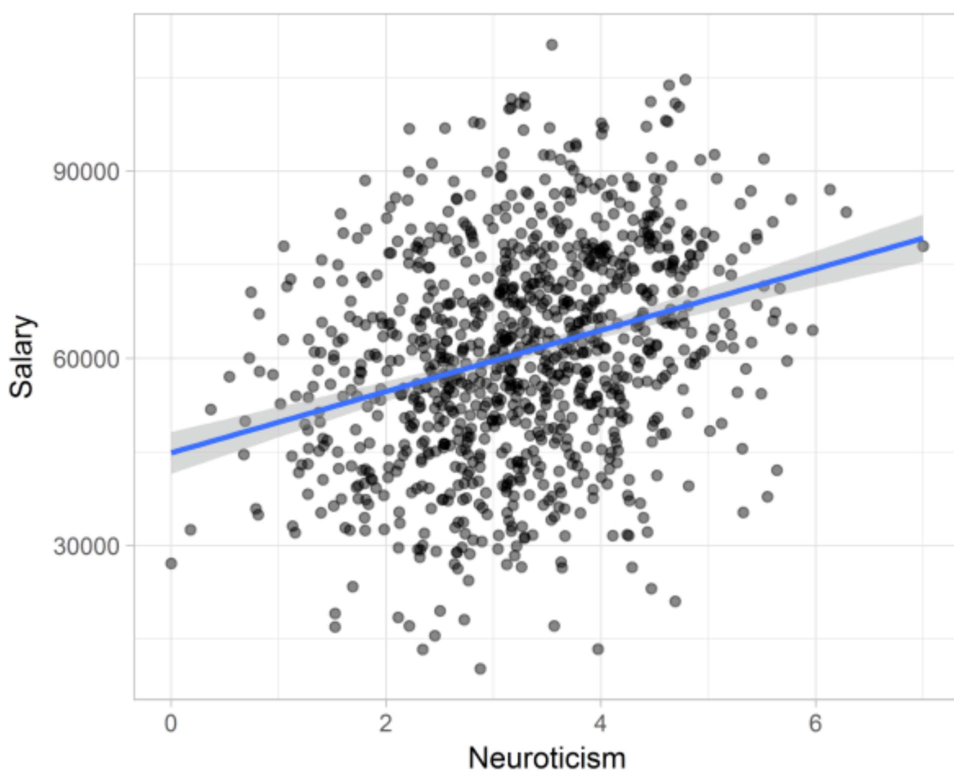
Statistical Rethinking

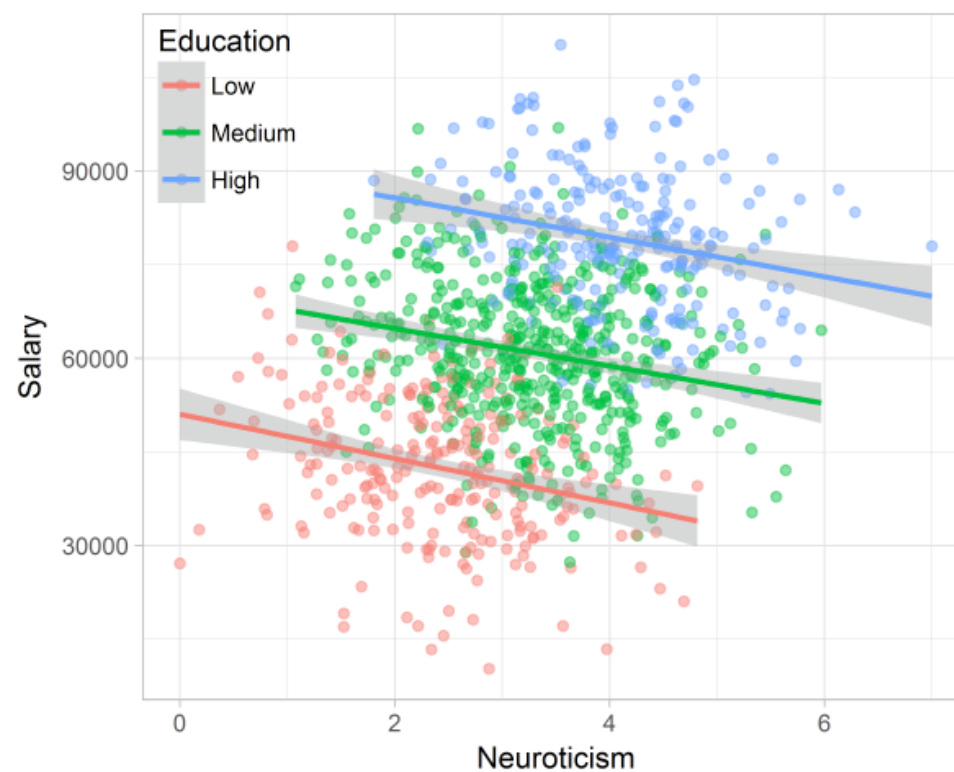
Week 7: GLMs, Monsters and Mixtures

Richard McElreath

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/

Gender contributes to personal research funding success in The Netherlands

Romy van der Lee[1] and Naomi Ellemers[2]

Department of Social and Organizational Psychology, Institute of Psychology, Leiden University, 2300 RB Leiden, The Netherlands

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved August 19, 2015 (received for review May 26, 2015)

We examined the application and review materials of three calls (n = 2,823) of a prestigious grant for personal research funding in a national full population of early career scientists awarded by the Netherlands Organization for Scientific Research (NWO). Results showed evidence of gender bias in application evaluations and success rates, as well as in language use in instructions and evaluation sheets. Male applicants received significantly more competitive "quality of researcher" evaluations (but not "quality of proposal" evaluations) and had significantly higher application success rates than female applicants. Gender disparities were most prevalent in scientific disciplines with the highest number of applications and with equal gender distribution among the applicants (i.e., life sciences and social sciences). Moreover, content analyses of the instructional and evaluation materials revealed the use of gendered language favoring male applicants. Overall, our data reveal a 4% "loss" of women during the grant review procedure, and illustrate the perpetuation of the funding gap, which contributes to the underrepresentation of women in academia.

gender bias | research funding | success rates | academia | STEM

Women are still underrepresented in academia today. Despite various attempts to promote gender equality (e.g., affirmative action initiatives, quotas), female scientists are less likely to get offered tenure, are judged to be less competent, receive less payment and research facilities, and are less likely to be awarded research grants compared with male scientists (1–3). Over time, this type of bias accumulates and contributes to the attrition of women from academia (4); the academic pipeline leaks. Here we report evidence of gender bias in personal research funding for early career scientists.

The importance of equal gender representation is widely ac-

associated with male traits and considered necessary for academic career success (20). Moreover, women still earn on average 18% less than their male colleagues for the same work with similar responsibilities (3). Although the salary gap seems to narrow for early career researchers, women in top academic positions are still substantially underpaid compared with men. Finally, across different career phases, success rates for female scientists applying for research funding tend to be lower than for male scientists (3, 21, 22). Even when overall success rates for men and women are equal, women receive less research funding than men, and are less often listed as principal investigators (23–25). Closing the funding gap is of particular importance, because this may help retain women in academia and foster the closing of other gaps by facilitating negotiations about salaries, research facilities, and promotion opportunities.

Current Study

To investigate the possibility of a funding gap, we examined a national full population of early career researchers who applied for a prestigious personal grant between 2010 and 2012 (Innovational Research Incentives Scheme Veni; n = 2,823, with 42.1% female applicants) awarded by the Netherlands Organization for Scientific Research (NWO). The NWO made available anonymized data from their archives for the purpose of this study, and approved publication of this research.

Our focus in this study was twofold. First, we tested for applicant gender differences in success rates and application evaluations. In doing so, we analyzed applicant gender as a statistical predictor of final success rates and also the success rates at each step in the review procedure (application, preselection, external reviewing,* interviews,

	discipline	gender	applications	awards	rejects
1	Chemical sciences	m	83	22	61
2	Chemical sciences	f	39	10	29
3	Physical sciences	m	135	26	109
4	Physical sciences	f	39	9	30
5	Physics	m	67	18	49
6	Physics	f	9	2	7
7	Humanities	m	230	33	197
8	Humanities	f	166	32	134
9	Technical sciences	m	189	30	159
10	Technical sciences	f	62	13	49
11	Interdisciplinary	m	105	12	93
12	Interdisciplinary	f	78	17	61
13	Earth/life sciences	m	156	38	118
14	Earth/life sciences	f	126	18	108
15	Social sciences	m	425	65	360
16	Social sciences	f	409	47	362
17	Medical sciences	m	245	46	199
18	Medical sciences	f	260	29	231

PNAS

No evidence that gender contributes to personal research funding success in The Netherlands: A reaction to van der Lee and Ellemers

A recent PNAS article (1) argues that success rates for attaining research grants are gender-biased. However, the overall gender effect borders on statistical significance, despite the large sample. Moreover, their conclusion could be a prime example of Simpson's paradox (2, 3); if a higher percentage of women apply for grants in more competitive scientific disciplines (i.e., with low application success rates for both men and women), then an analysis across all disciplines could incorrectly show "evidence" of gender inequality. Indeed, the social sciences and medical sciences are the two fields with a high proportion of female applicants as well as a low application success rate (table S1 in ref. 1). Moreover, multiple

tion of Scientific Research (NWO) consisting of 8,687 individual applications to all grants announced in the period between 2006 and

ni grant). Taking nesting
nd years into account
coefficient = 14.5% in
ivariate analyses of the
show no or just border-
0.062), whereas bivariate
tions show a highly sig-
eems to support the con-
Lee and Ellemers (1).
f grant and social scien-
—separately or together—
vidence to reject the null
equality. Also, no interac-
een gender and these

no convincing evidence
y. However, based on
may not conclude that
equality in NWO grant
Rather, it is too soon to
on changing the eval-
and gender balancing
Science Foundation in
The Netherlands. More in-depth analyses with statistical techniques that overcome the above-mentioned issues are needed before jumping to conclusions about gender inequality in grant awards.

Our analyses are summarized in Table 1 and more detailed analyses are available on request.

Beate Volker[a,1] and Wouter Steenbeek[b]

[a]University of Amsterdam, 1018 WV Amsterdam, The Netherlands; and [b]Netherlands Institute for the Study of Crime and Law Enforcement, 1081 HV Amsterdam, The Netherlands

1 van der Lee R, Ellemers N (2015) Gender contributes to personal research funding success in The Netherlands. *Proc Natl Acad Sci USA* 112(40):12349–12353.

2 Albers C (2015) NWO, gender bias and Simpson's paradox. *Casper Albers' Blog*. Available at blog.casperalbers.nl/science/nwo-gender-bias-and-simpsons-paradox/. Accessed November 5, 2015.

3 Simpson EH (1951) The interpretation of interaction in contingency tables. *J R Stat Soc, B* 13(2):238–241.

Author contributions: B.V. designed research; B.V. performed research; W.S. contributed new reagents/analytic tools; B.V. analyzed data; and B.V. and W.S. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: b.volker@uva.nl.

GLMs need taming

R code 10.34

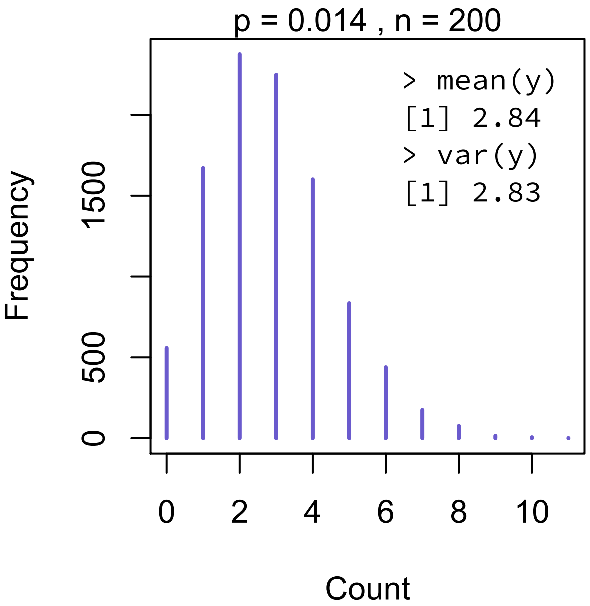
```
# outcome and predictor almost perfectly associated
y <- c( rep(0,10) , rep(1,10) )
x <- c( rep(-1,9) , rep(1,11) )
# fit binomial GLM
m.bad <- glm( y ~ x , data=list(y=y,x=x) , family=binomial )
precis(m.bad)
```

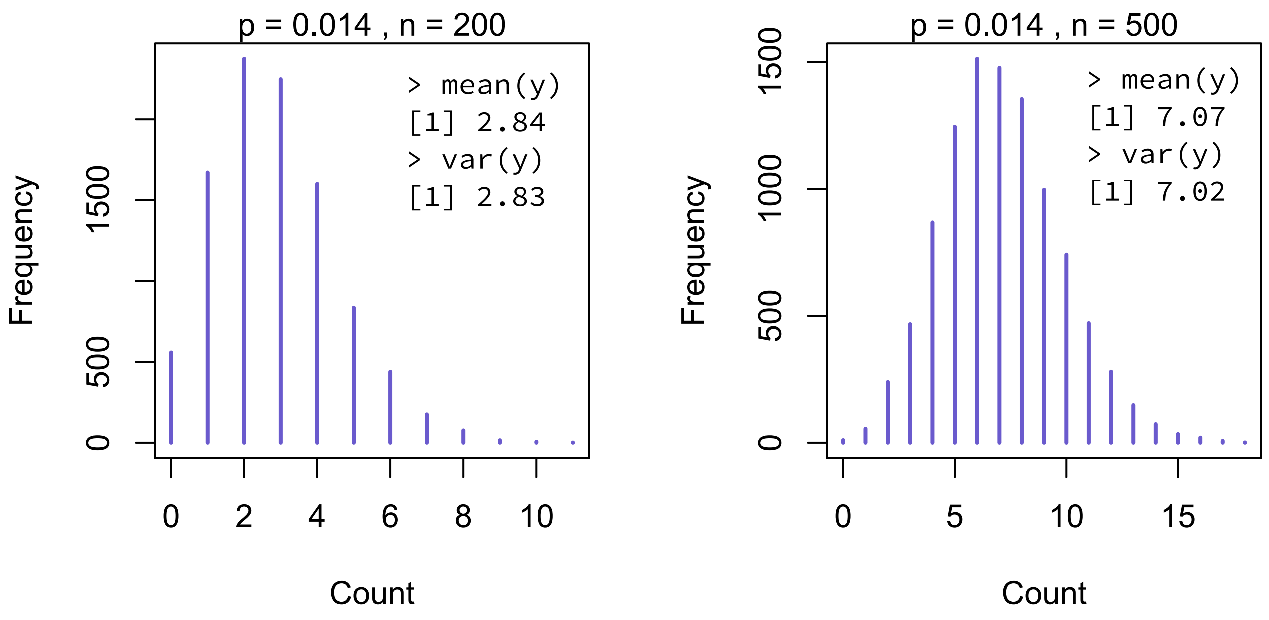
```
             Mean  StdDev     2.5%   97.5%
(Intercept) -9.13 2955.06 -5800.95 5782.68
x           11.43 2955.06 -5780.38 5803.25
```

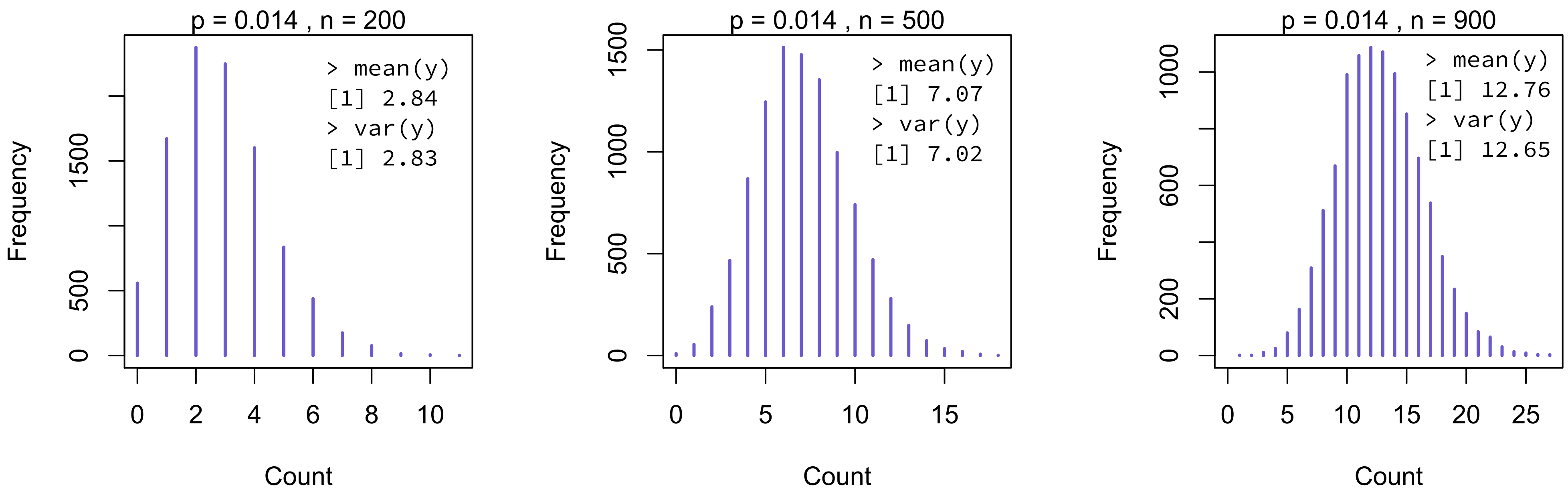
```
   y  x
1  0 -1
2  0 -1
3  0 -1
4  0 -1
5  0 -1
6  0 -1
7  0 -1
8  0 -1
9  0 -1
10 0  1
11 1  1
12 1  1
13 1  1
14 1  1
15 1  1
16 1  1
17 1  1
18 1  1
19 1  1
20 1  1
```

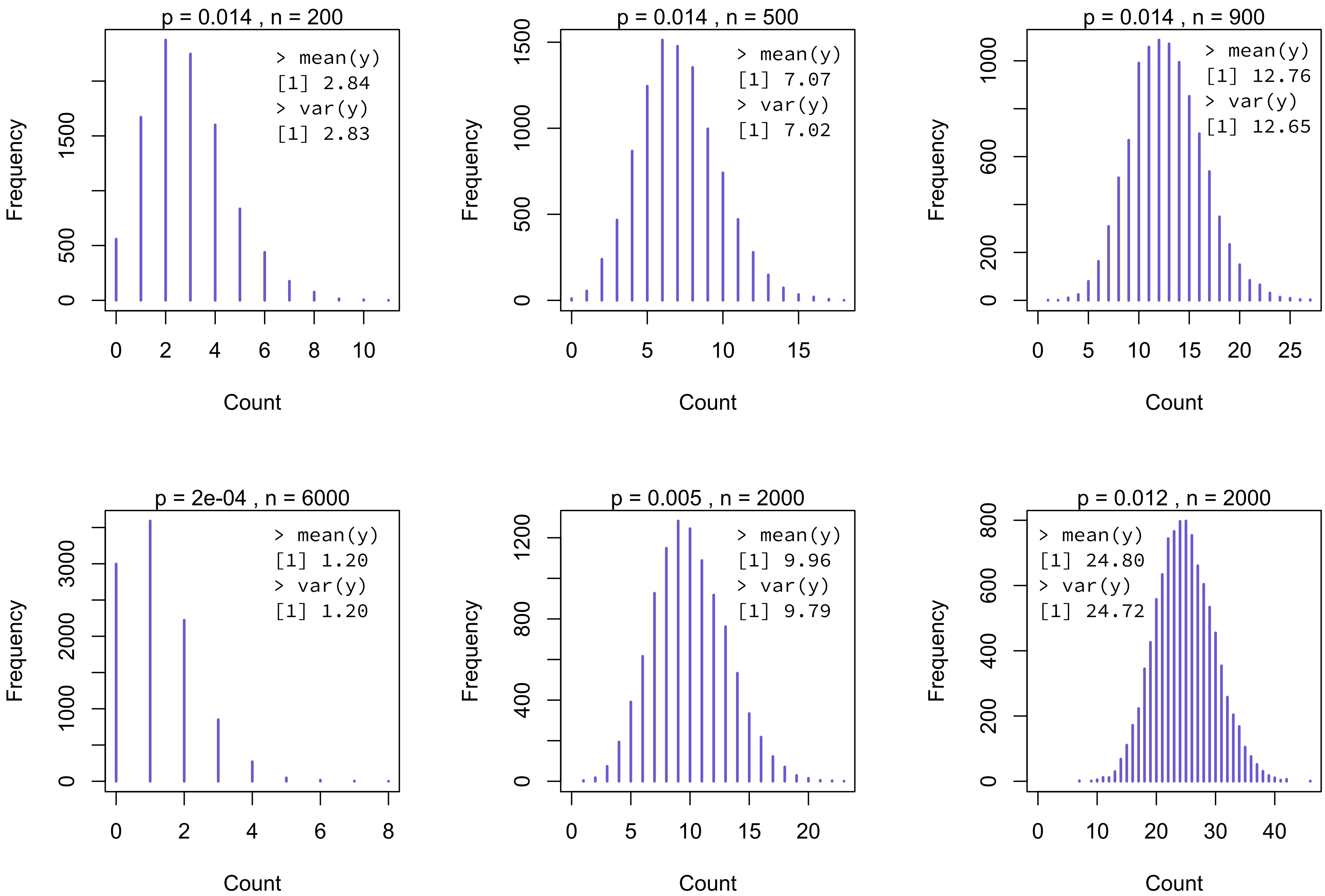
Binomial GLMs

- Predict counts with a fixed maximum
- Use logit link
- Distrust MAP estimation & QA
    - may work, but routinely does not
    - regularization even more important now
- Convert back to probability/count scale to plot predictions
- Focus on *predictions*, not *parameters*


p = 0.014 , n = 200
> mean(y)
[1] 2.84
> var(y)
[1] 2.83
Frequency
0
500
1500
0
2
4
6
8
10
Count


p = 0.014 , n = 200
> mean(y)
[1] 2.84
> var(y)
[1] 2.83
Frequency
0
500
1500
0
2
4
6
8
10
Count
p = 0.014 , n = 500
> mean(y)
[1] 7.07
> var(y)
[1] 7.02
Frequency
0
500
1000
1500
0
5
10
15
Count


p = 0.014 , n = 200
> mean(y)
[1] 2.84
> var(y)
[1] 2.83
Frequency
Count
p = 0.014 , n = 500
> mean(y)
[1] 7.07
> var(y)
[1] 7.02
Frequency
Count
p = 0.014 , n = 900
> mean(y)
[1] 12.76
> var(y)
[1] 12.65
Frequency
Count

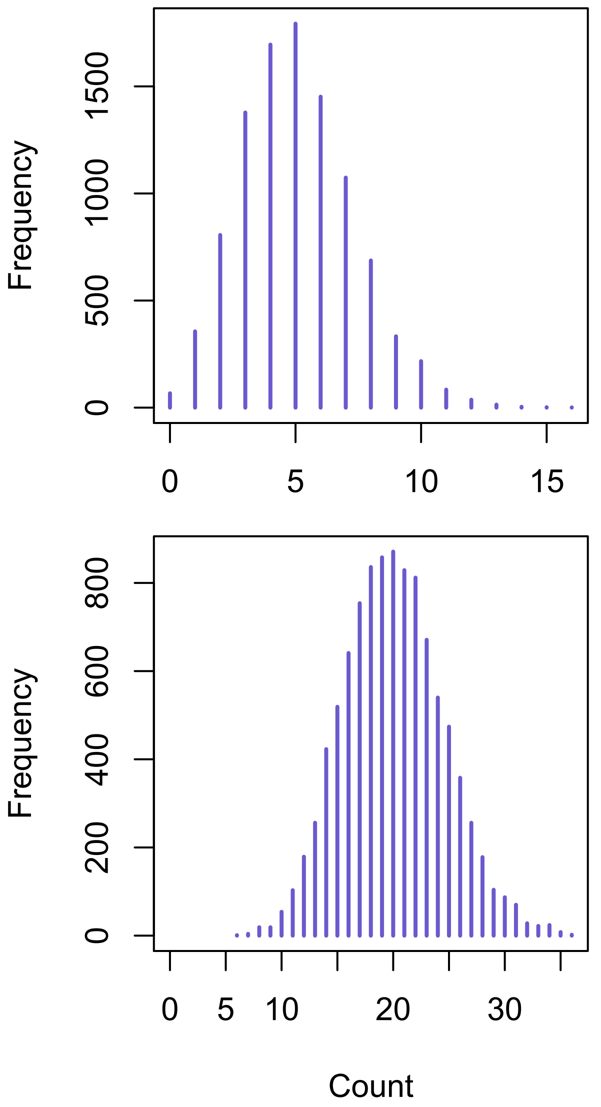

p = 0.014 , n = 200
> mean(y)
[1] 2.84
> var(y)
[1] 2.83
p = 0.014 , n = 500
> mean(y)
[1] 7.07
> var(y)
[1] 7.02
p = 0.014 , n = 900
> mean(y)
[1] 12.76
> var(y)
[1] 12.65
p = 2e-04 , n = 6000
> mean(y)
[1] 1.20
> var(y)
[1] 1.20
p = 0.005 , n = 2000
> mean(y)
[1] 9.96
> var(y)
[1] 9.79
p = 0.012 , n = 2000
> mean(y)
[1] 24.80
> var(y)
[1] 24.72
Frequency
Count

Poisson GLMs

$$y \sim \text{Poisson}(\lambda)$$

$$E(y) = \lambda$$
$$\text{var}(y) = \lambda$$

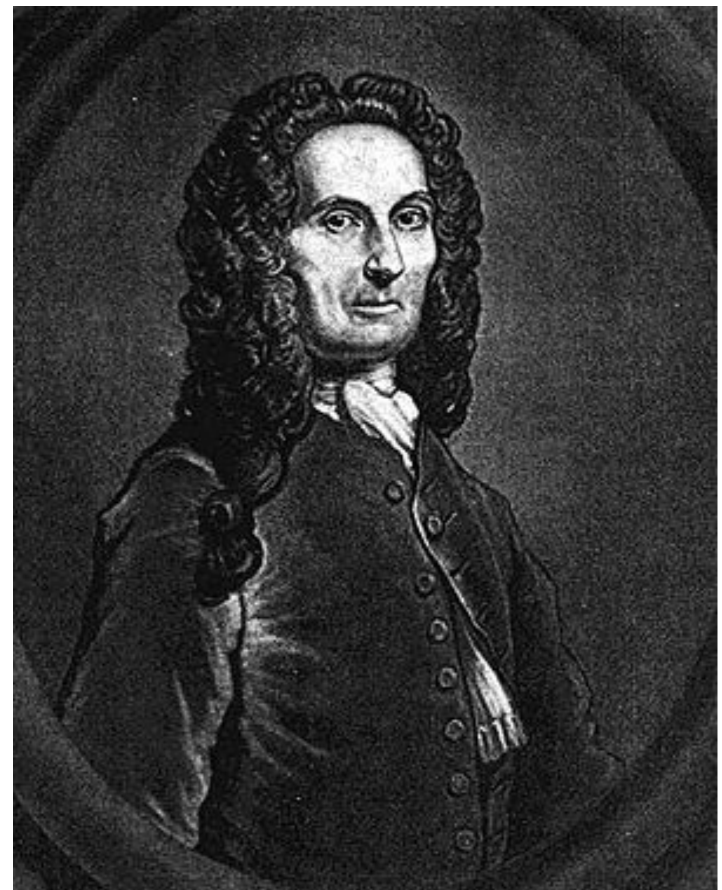
- Counts without upper limit, constant expected value
- Single parameter: events per unit time/distance
- Variance equal to mean



Poisson GLMs

- Examples: Soccer goals, fission events, photons striking a detector, DNA mutations, soldiers killed by horses

Siméon Denis Poisson (1781–1840)



Abraham de Moivre (1667–1754)

entire city, α_i the effect of ground displacement on terrace i, β_j the effect of ground displacement on ridge j, and ε_{ij} the random variation of house height around the acre mean in acre (i,j).

The point is that this model incorporates three sources of variation in a structured way that permits them to be separately dealt with, even if no two plots have the same combination of treatments. By incorporating all this in one model, there is a huge bonus: If one approaches the data ignoring one factor (for example, fertilizer or ridges), the variation due to the omitted factor could dwarf the variation due to the other factor and uncontrolled factors, thus making detection or estimation of the other factor (for example, variety or terraces) impossible. But if both were included (in some applications, Fisher would call this blocking), the effect of both would jump out through the row or column means and their variation and be clearly identifiable. In even a basic additive effects example, the result could be striking; in more complex situations, it could be heroic.

To give one example that shows clearly what could be missed, consider the famous set of data compiled in the 1890s with great effort from massive volumes of Prussian state statistics by Ladislaus von Bortkiewicz and included in his short tract, *Das Gesetz der kleinen Zahlen* (The law of small numbers) in 1898.[7] The data give the numbers of Prussian

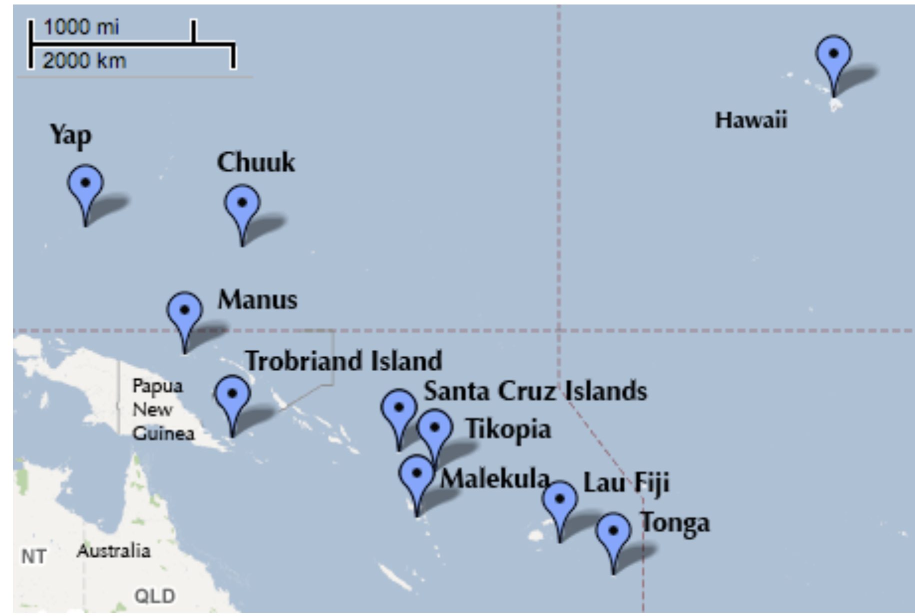
	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
G	—	2	2	1	—	—	1	1	—	3	—	2	1	—	—	1	—	1	—	1
I	—	—	—	2	—	3	—	2	—	—	—	1	1	1	—	2	—	3	1	—
II	—	—	—	2	—	2	—	—	1	1	—	—	2	1	1	—	—	2	—	—
III	—	—	—	1	1	1	2	—	2	—	—	—	1	—	1	2	1	—	—	—
IV	—	1	—	1	1	1	1	—	—	—	—	1	—	—	—	—	1	1	—	—
V	—	—	—	—	2	1	—	—	1	—	—	1	—	1	1	1	1	1	1	—
VI	—	—	1	—	2	—	—	1	2	—	1	1	3	1	1	1	—	3	—	—
VII	1	—	1	—	—	—	1	—	1	1	—	—	2	—	—	2	1	—	2	—
VIII	1	—	—	—	1	—	—	1	—	—	—	—	1	—	—	—	1	1	—	1
IX	—	—	—	—	—	2	1	1	1	—	2	1	1	—	1	2	—	1	—	—
X	—	—	1	1	—	1	—	2	—	2	—	—	—	—	2	1	3	—	1	1
XI	—	—	—	—	2	4	—	1	3	—	1	1	1	1	2	1	3	1	3	1
XIV	1	1	2	1	1	3	—	4	—	1	—	3	2	1	—	2	1	1	—	—
XV	—	1	—	—	—	—	—	1	—	1	1	—	—	—	2	2	—	—	—	—

6.1 Bortkiewicz's data were gathered from the large published Prussian state statistics (three huge volumes each year for this period). He included 14 corps (*G* being the Guard Corps) over 20 years. *(Bortkiewicz 1898)*

cavalry killed by horse kicks in 14 cavalry corps over a 20-year period (see Figure 6.1). Bortkiewicz wanted to demonstrate that the great variability in such small and unpredictable numbers could mask real effects, and he showed that the 280 numbers viewed together were well fit as a set of identically distributed Poisson variables. And indeed they are. But Bortkiewicz lacked the technology of additive models that, if applied here (using a generalized linear model with Poisson variation), clearly shows not only corps-to-corps variation, but also year-to-year variability. The corps and year variations were not large, but the additive model allowed them to be captured by 14 plus 20 separate effects. With 240

Oceanic tool complexity

culture	population	contact	total_tools	mean_TU
Malekula	1100	low	13	3.2
Tikopia	1500	low	22	4.7
Santa Cruz	3600	low	24	4.0
Yap	4791	high	43	5.0
Lau Fiji	7400	high	33	5.0
Trobriand	8000	high	19	4.0
Chuuk	9200	high	40	3.8
Manus	13000	low	28	6.6
Tonga	17500	high	55	5.4
Hawaii	275000	low	71	6.6



(1) Complexity of toolkit proportional to magnitude of population?

(2) Contact with other islands moderates impact?

Dr. Michelle Kline (Simon Fraser U)

Anatomy of Poisson GLM

$$
\begin{aligned}
T_i &\sim \mathrm{Poisson}(\lambda_i) \\
\log \lambda_i &= \alpha + \beta_P \log P_i + \beta_C C_i + \beta_{PC} C_i \log P_i \\
\alpha &\sim \mathrm{Normal}(0, 100) \\
\beta_P &\sim \mathrm{Normal}(0, 1) \\
\beta_C &\sim \mathrm{Normal}(0, 1) \\
\beta_{PC} &\sim \mathrm{Normal}(0, 1)
\end{aligned}
$$

Anatomy of Poisson GLM

total_tools
(*outcome*)

$$T_i \sim \text{Poisson}(\lambda_i)$$
$$\log \lambda_i = \alpha + \beta_P \log P_i + \beta_C C_i + \beta_{PC} C_i \log P_i$$

Anatomy of Poisson GLM

total_tools
(*outcome*)

expected tools
for case i

$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\log \lambda_i = \alpha + \beta_P \log P_i + \beta_C C_i + \beta_{PC} C_i \log P_i$$

Anatomy of Poisson GLM

total_tools
(*outcome*)

expected tools
for case *i*

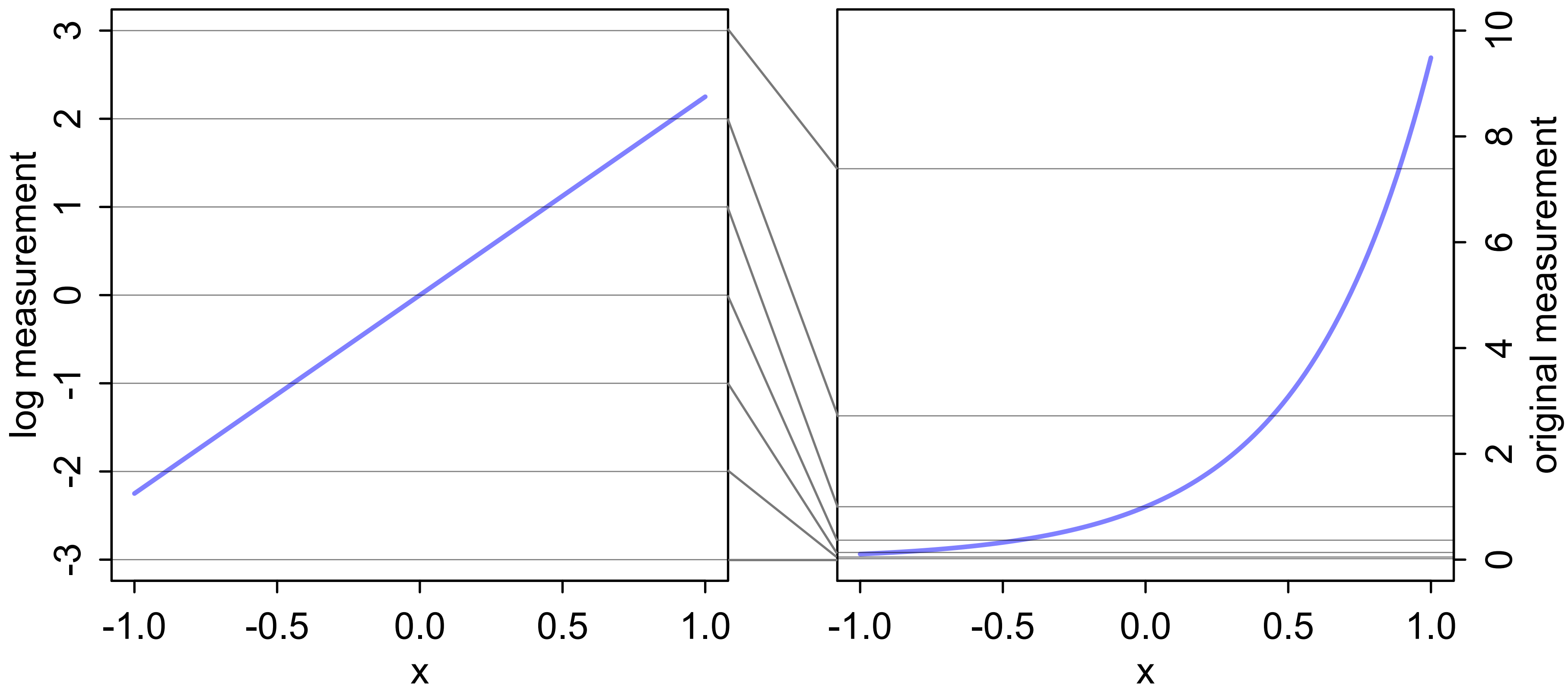
$$T_i \sim \text{Poisson}(\lambda_i)$$

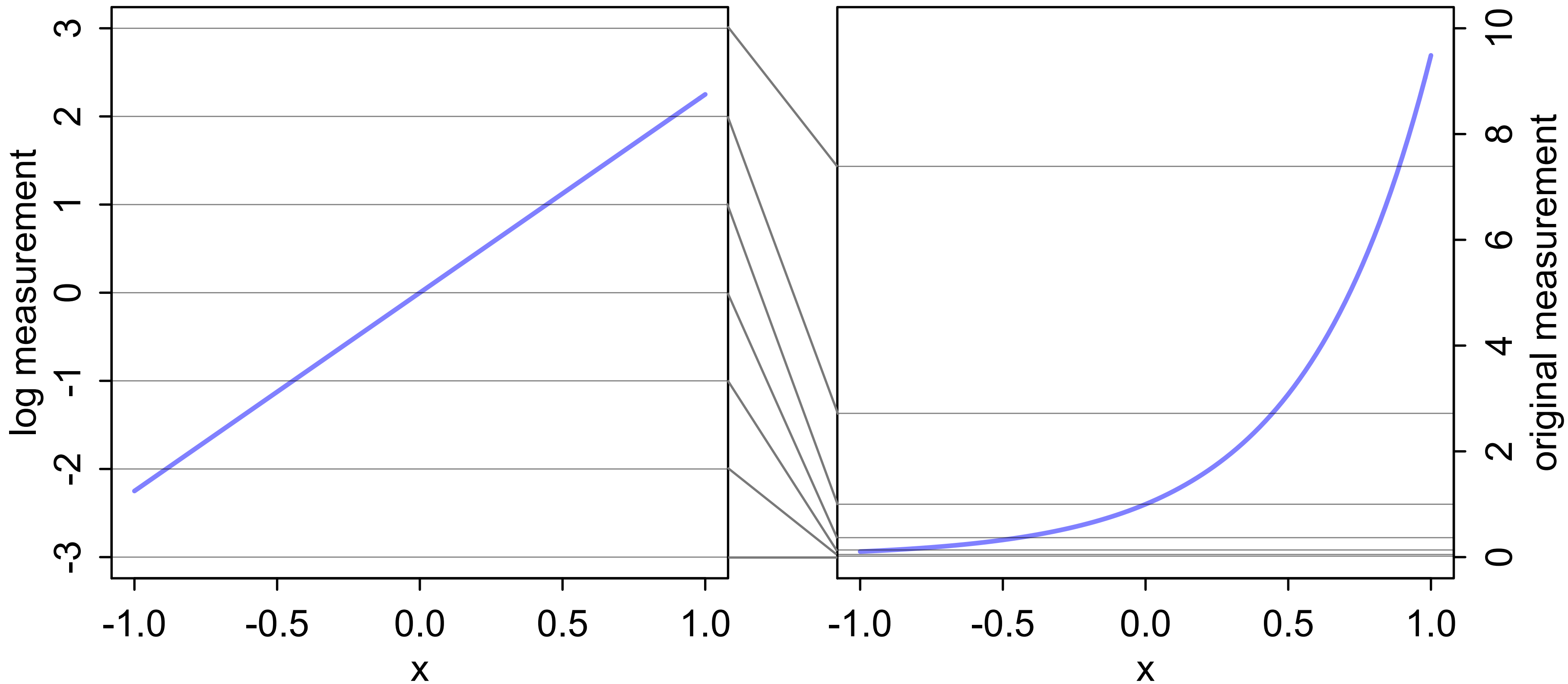
log link

$$\log \lambda_i = \alpha + \beta_P \log P_i + \beta_C C_i + \beta_{PC} C_i \log P_i$$

Log link

- Goal: Map linear model to positive reals





$$y_i \sim \text{Normal}(\mu, \sigma_i)$$
$$\log(\sigma_i) = \alpha + \beta x_i$$

Solve for *sigma*:

$$\sigma_i = \exp(\alpha + \beta x_i)$$

Anatomy of Poisson GLM

total_tools
(*outcome*)

expected tools
for case i

$$T_i \sim \text{Poisson}(\lambda_i)$$

log link

$$\log \lambda_i = \alpha + \beta_P \log P_i + \beta_C C_i + \beta_{PC} C_i \log P_i$$

log population

contact
(0/1)

interaction

Anatomy of Poisson GLM

total_tools
(*outcome*)

expected tools
for case *i*

log link

$$
\begin{aligned}
T_i &\sim \text{Poisson}(\lambda_i) \\
\log \lambda_i &= \alpha + \beta_P \log P_i + \beta_C C_i + \beta_{PC} C_i \log P_i \\
\alpha &\sim \text{Normal}(0, 100) \\
\beta_P &\sim \text{Normal}(0, 1) \\
\beta_C &\sim \text{Normal}(0, 1) \\
\beta_{PC} &\sim \text{Normal}(0, 1)
\end{aligned}
$$

Fitting

R code
10.39

```
d$log_pop <- log(d$population)
d$contact_high <- ifelse( d$contact=="high" , 1 , 0 )
```

R code
10.40

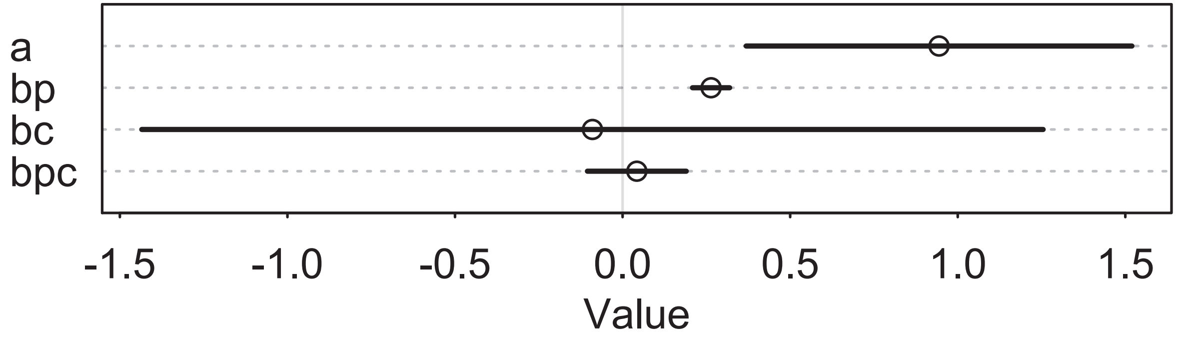
```
m10.10 <- map(
    alist(
        total_tools ~ dpois( lambda ),
        log(lambda) <- a + bp*log_pop +
            bc*contact_high + bpc*contact_high*log_pop,
        a ~ dnorm(0,100),
        c(bp,bc,bpc) ~ dnorm(0,1)
    ),
    data=d )
```

Beware marginal estimates

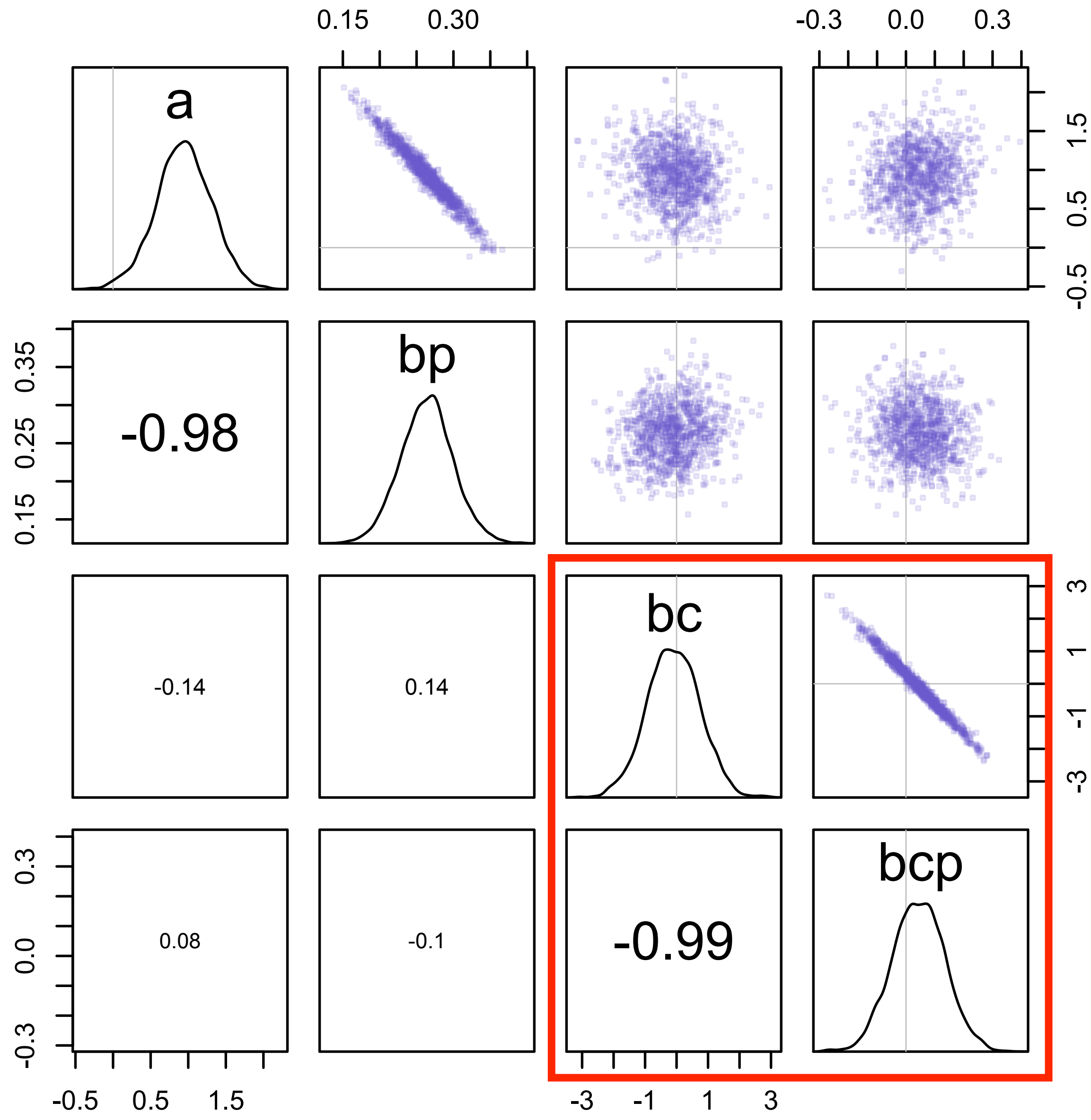
```
precis(m10.10,corr=TRUE)
plot(precis(m10.10))
```

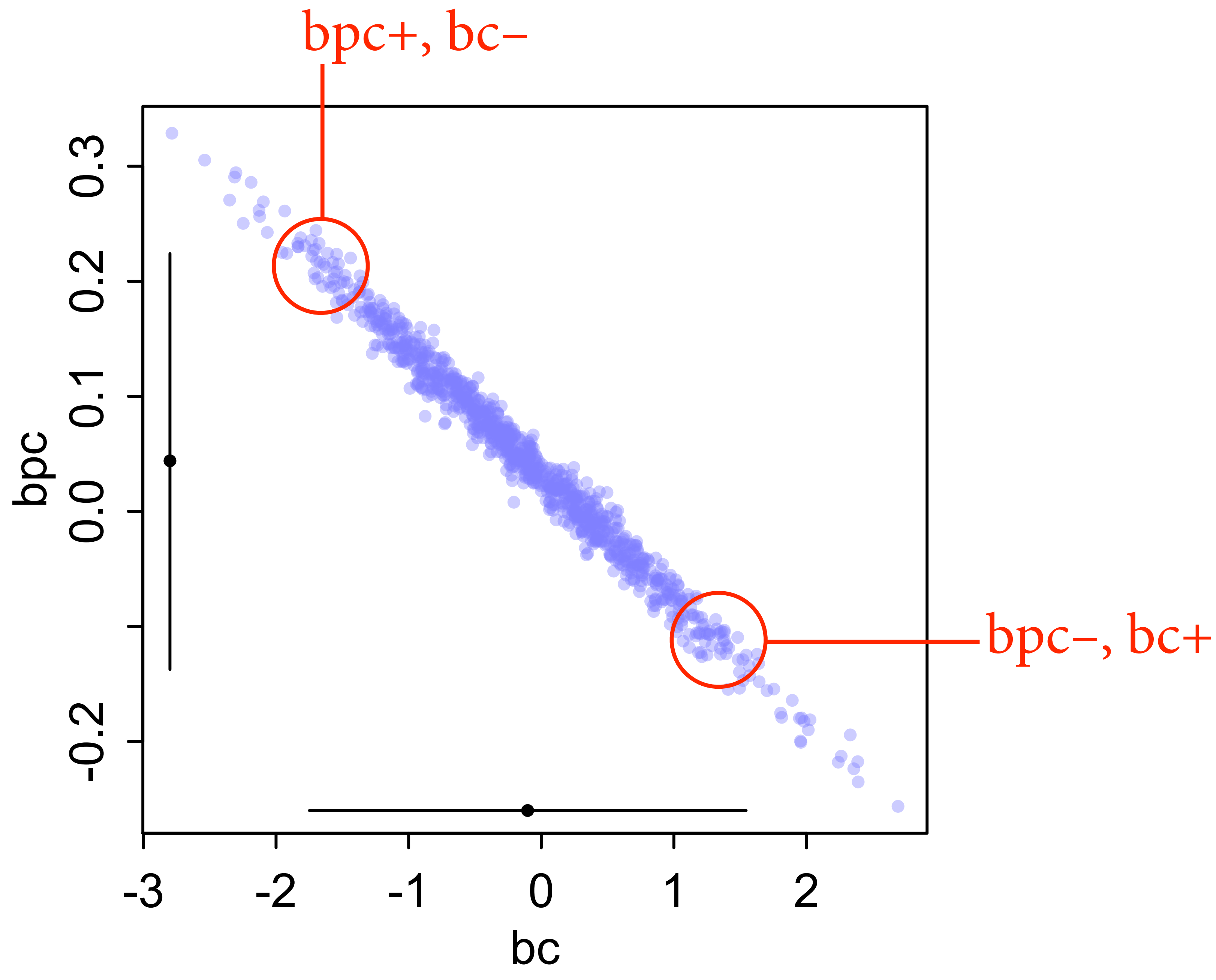
R code 10.42

```
     Mean  StdDev  5.5%  94.5%      a     bp     bc    bpc
a     0.94   0.36   0.37   1.52   1.00 -0.98 -0.13   0.07
bp    0.26   0.03   0.21   0.32  -0.98  1.00  0.12  -0.08
bc   -0.09   0.84  -1.43   1.25  -0.13  0.12  1.00  -0.99
bpc   0.04   0.09  -0.10   0.19   0.07 -0.08 -0.99   1.00
```



Pairs plot (Stan samples)



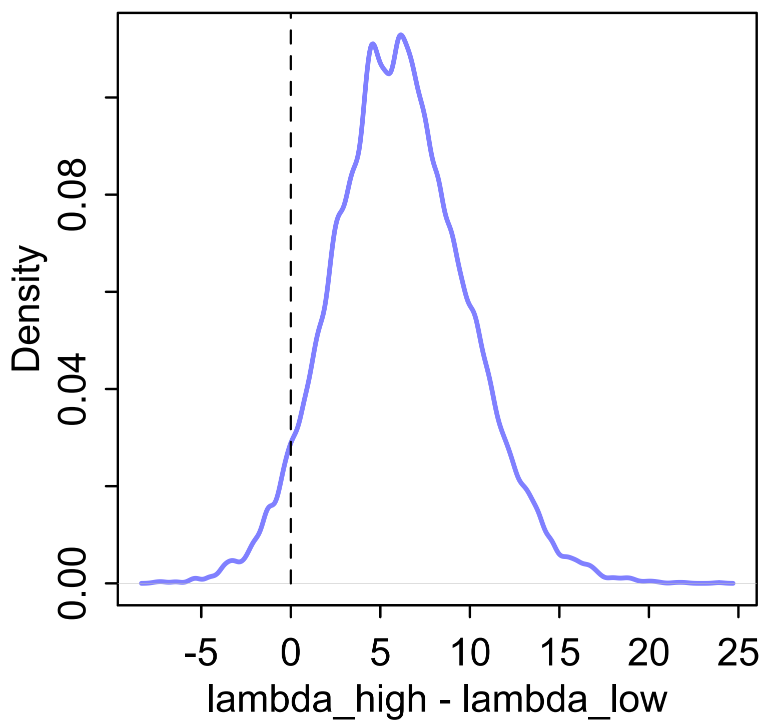

bpc+, bc−
bpc−, bc+
bpc
0.3
0.2
0.1
0.0
-0.2
bc
-3
-2
-1
0
1
2

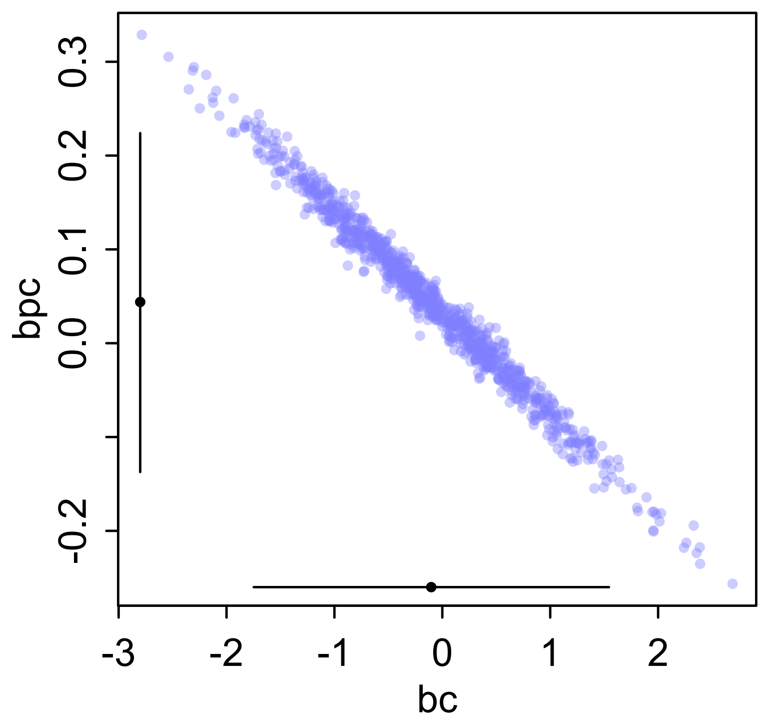
Focus on predictions

```r
diff <- lambda_high - lambda_low
sum(diff > 0)/length(diff)
```

R code 10.43

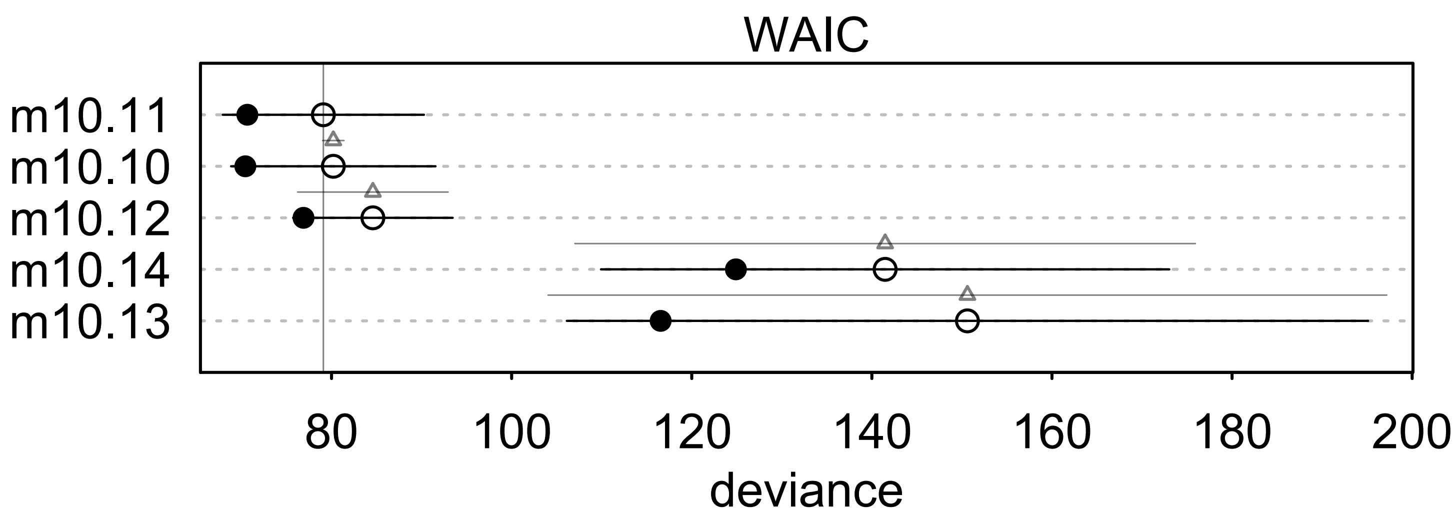
```
[1] 0.9527
```





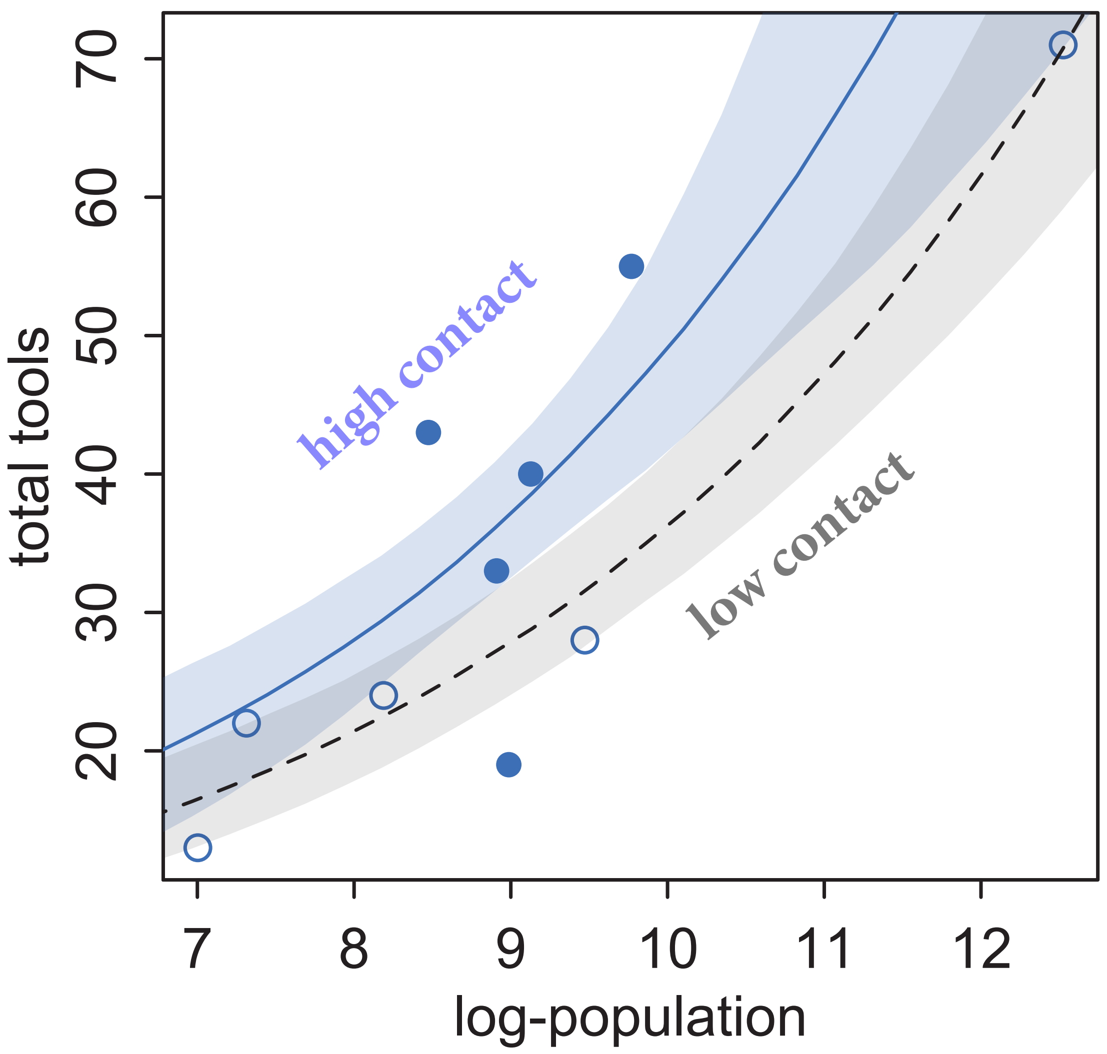
Model comparison

		WAIC	pWAIC	dWAIC	weight	SE	dSE
log pop, contact	m10.11	79.0	4.2	0.0	0.62	11.19	NA
interaction	m10.10	80.1	4.9	1.2	0.35	11.42	1.28
log pop only	m10.12	84.6	3.8	5.6	0.04	8.91	8.47
null (intercept only)	m10.14	141.5	8.2	62.5	0.00	31.53	34.42
contact only	m10.13	149.8	16.7	70.8	0.00	43.96	46.01



Since WAIC constructed over predictions, automatically accounts for posterior correlations

Prediction ensemble



Poisson GLMs

- For counts without obvious upper bound
- log link is customary; linear model of *magnitude*
  - Beware exploding exponential predictions
- Focus on *predictions*, not *parameters*
  - Convert back to count scale to interpret/plot
- Use offset to adjust *exposure* duration/distance
- Predictions tend to be *under-dispersed* relative to data
  - Common problem for both binomial and Poisson GLMs => un-modeled heterogeneity

Additional count distributions

- Multinomial: generalized binomial, more than 2 un-ordered outcomes
    - Tricky to use and understand
- Geometric: number of trials until specific event
    - Common event-history (survival) distribution
- Mixtures, coping with heterogeneity:
    - Beta-binomial: varying probabilities
    - gamma-Poisson: varying rates
    - many others (e.g. Dirichlet-multinomial)

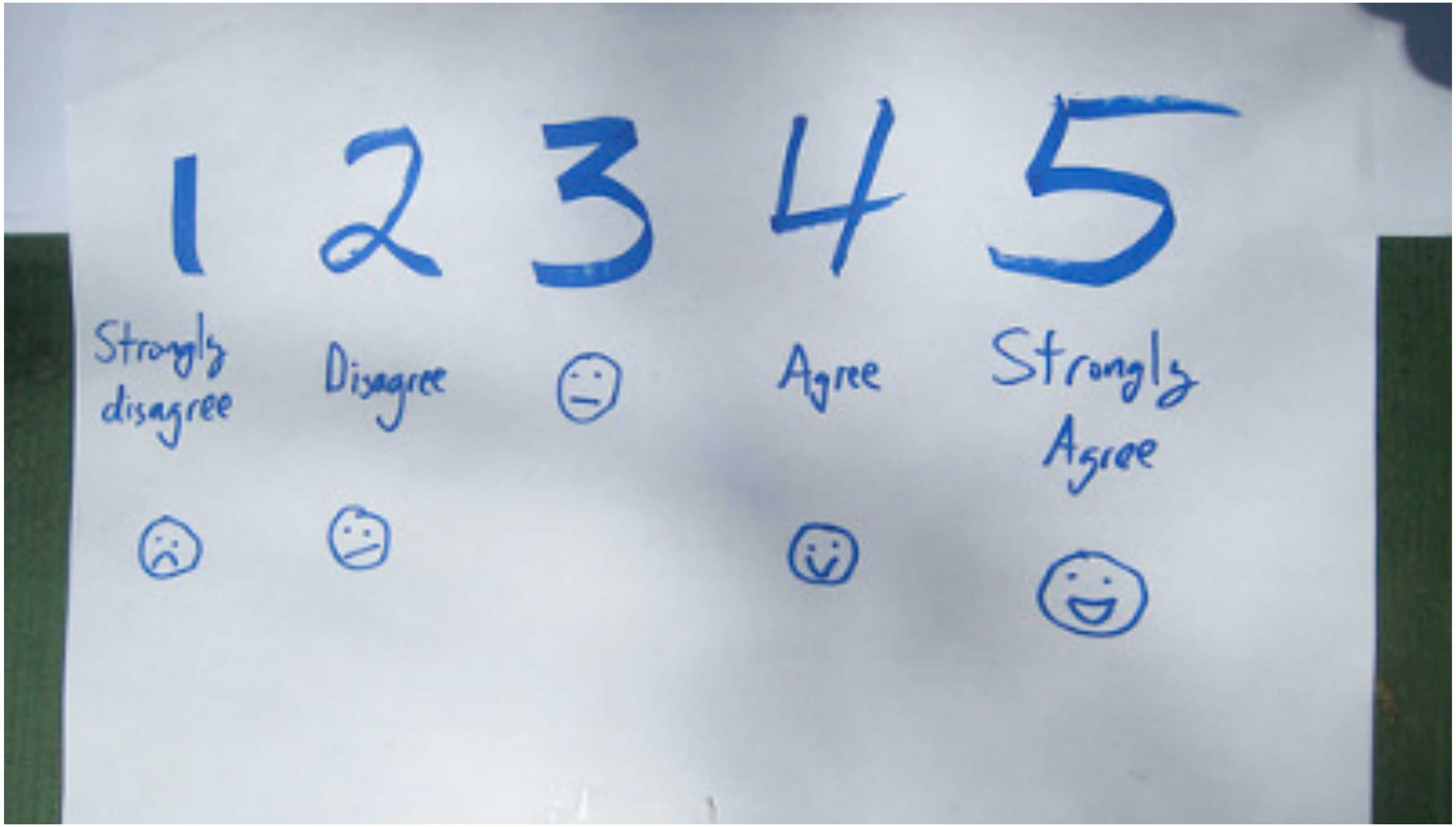

But the night before the child was born...
I'm leaving...
Where are you going?
Our son won't be safe in this world. I'm going where I can protect him.
Here. He'll need this.
Your cape?
WAIT!
His skin feels rough!
GILLS?
...FINS?!

Monsters & mixtures

- More complicated GLMs:
  - *Monsters*: Specialized, complex distributions
    - ordered categories, ranks
  - *Mixtures*: Blends of stochastic processes
    - Varying means, probabilities, rates
    - Varying process: zero-inflation, hurdles

Ordered categories

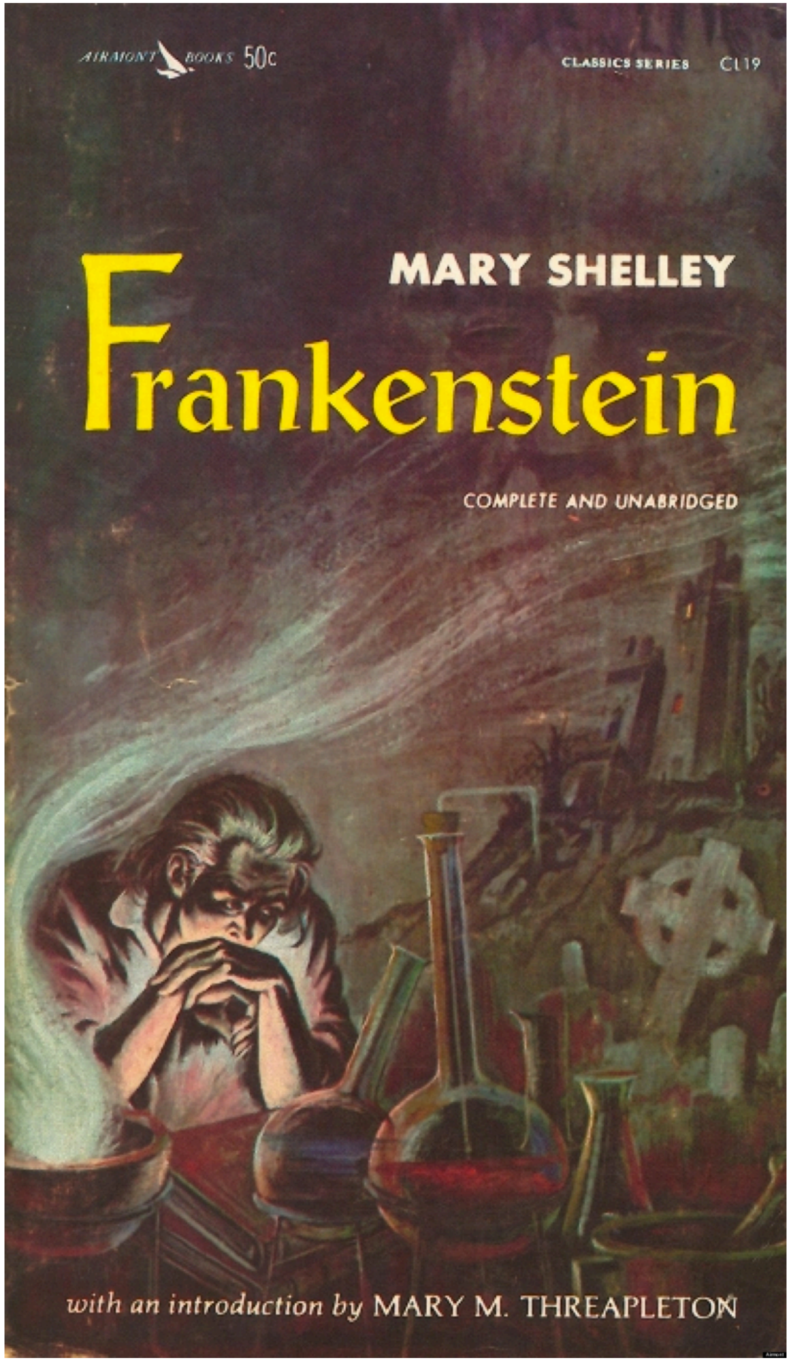
- How much do you like this class? (1–7)
- How important is income of a potential spouse? (1–7)
- How often do you see bats around Leipzig? (never, sometimes, frequently)



- Discrete outcomes
- Defined minimum and maximum
- Defined order
- “Distances” between categories unknown

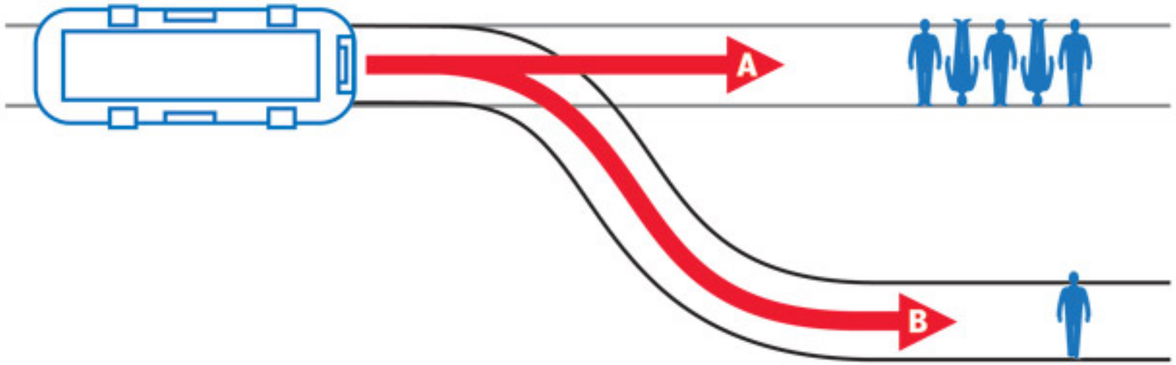
Ordered categories

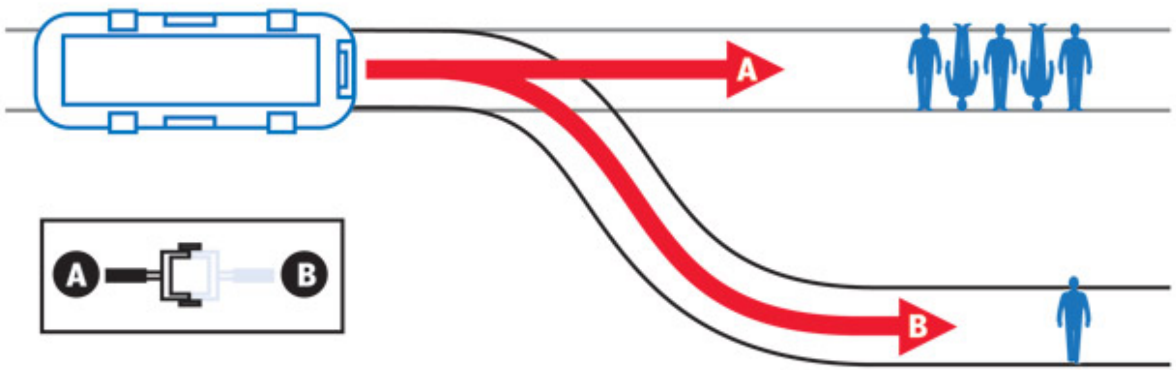
- Hard to model
  - Not continuous
  - Not counts
- Solution: ordered logistic regression
  - categorical model with a fancy link function
- Good example of making a monster

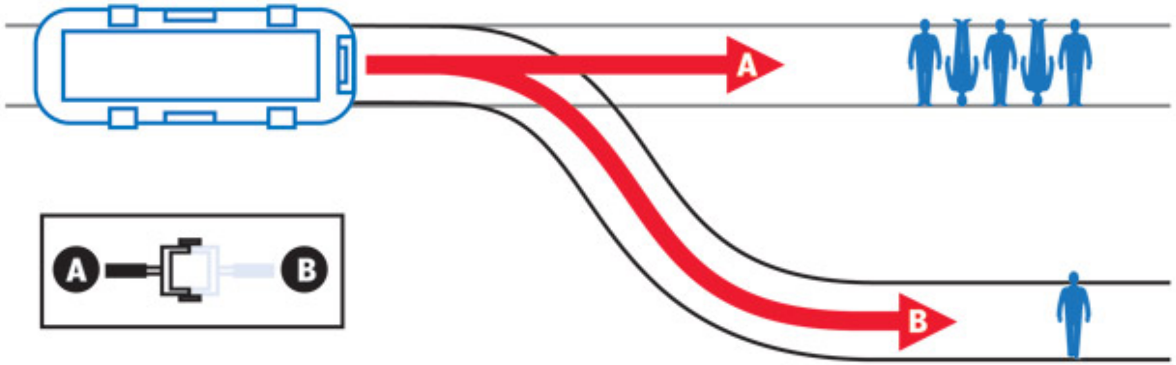



A

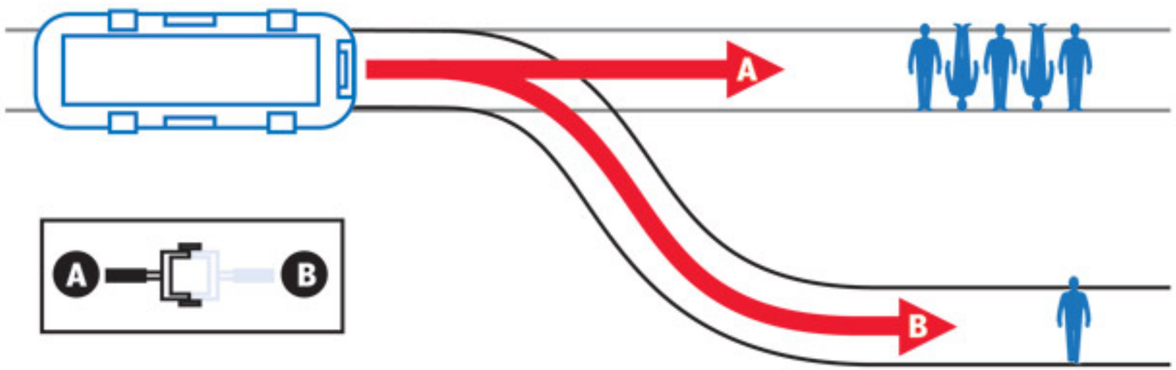

A


A
B


A
B
A
B

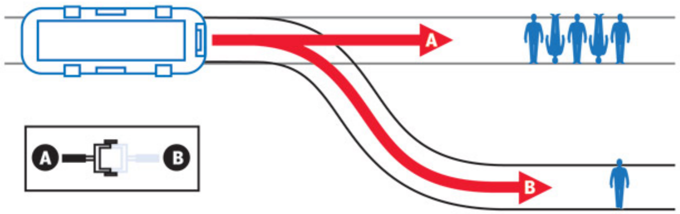


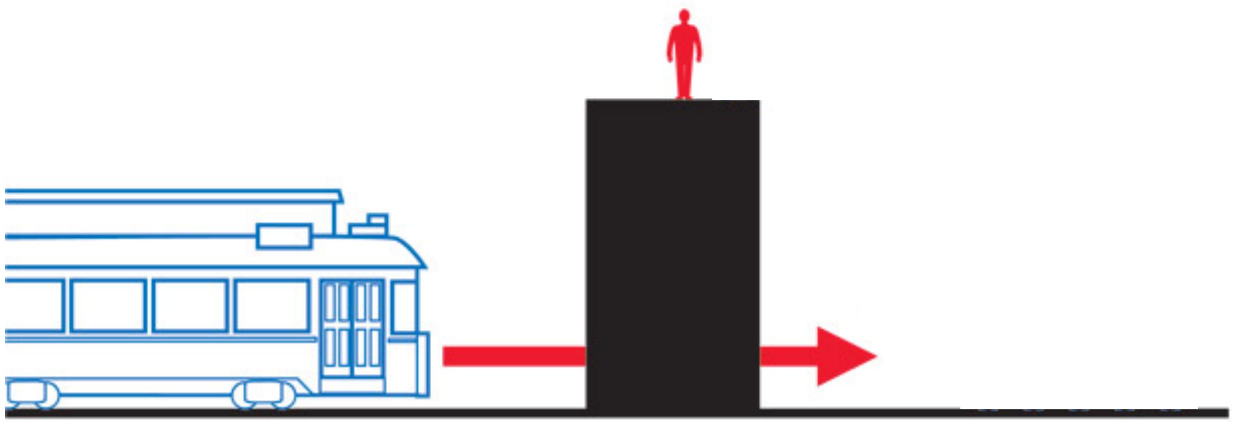
How morally permissible is it
to pull the lever?

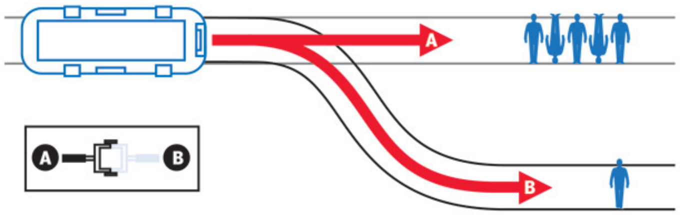


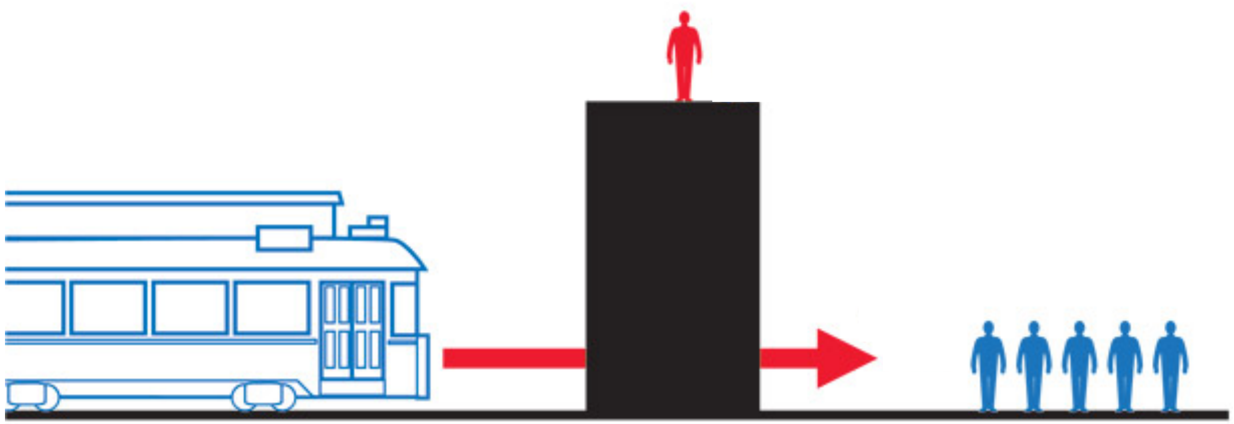
How morally permissible is it
to pull the lever?

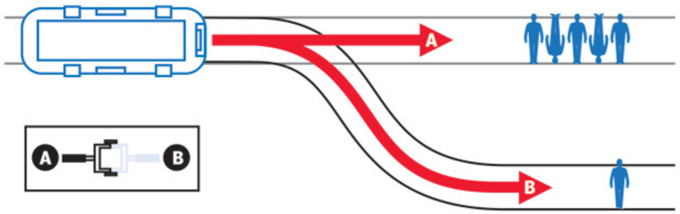
never 1 2 3 4 5 6 7 always

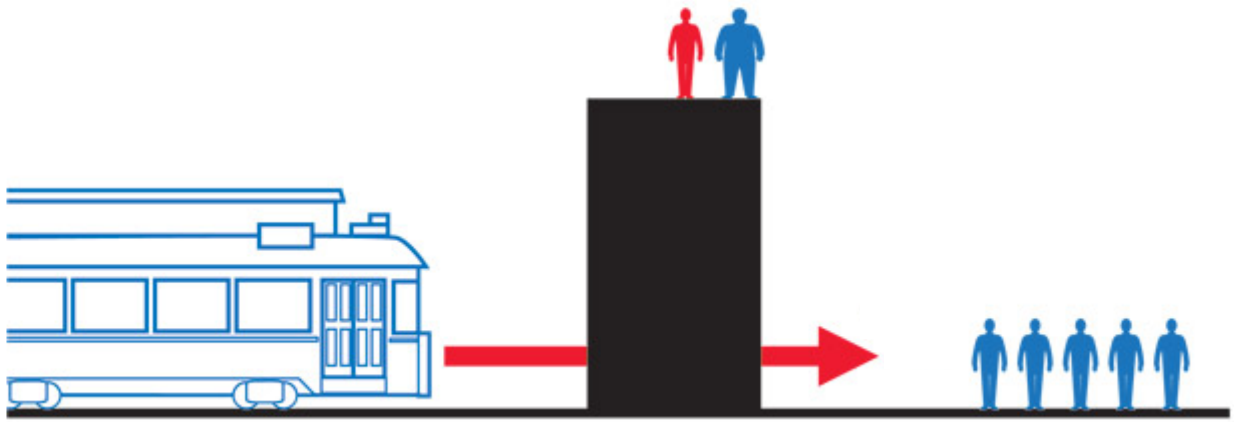

A
B
A
B

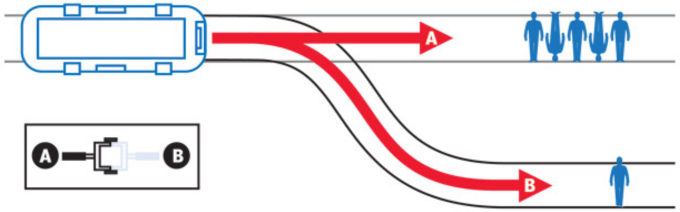



A
B
A
B

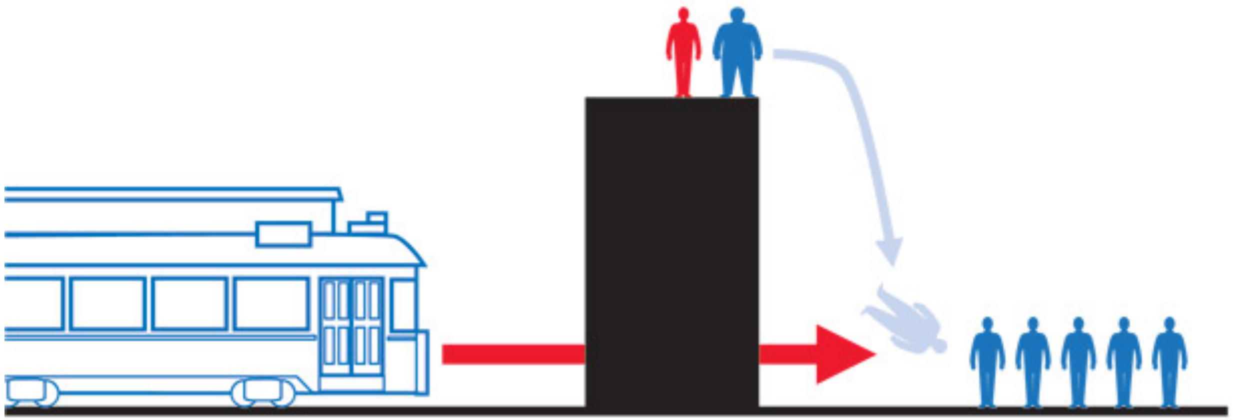


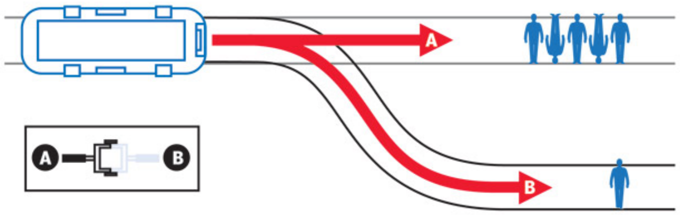

A
B

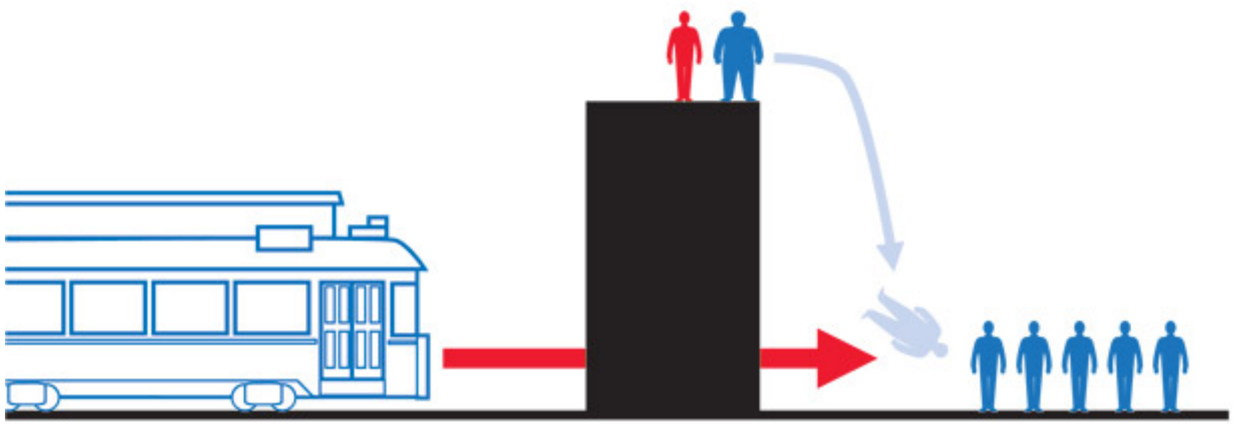




A
B
A
B

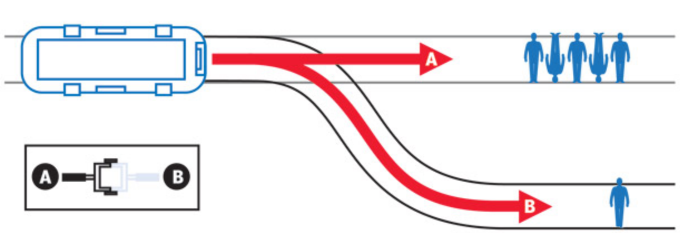


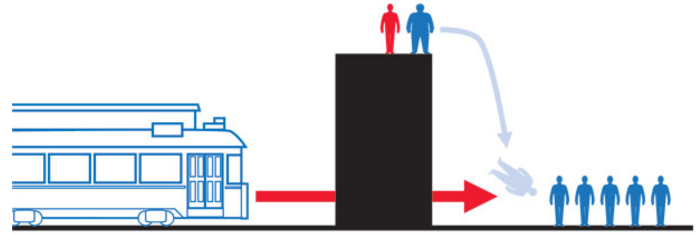


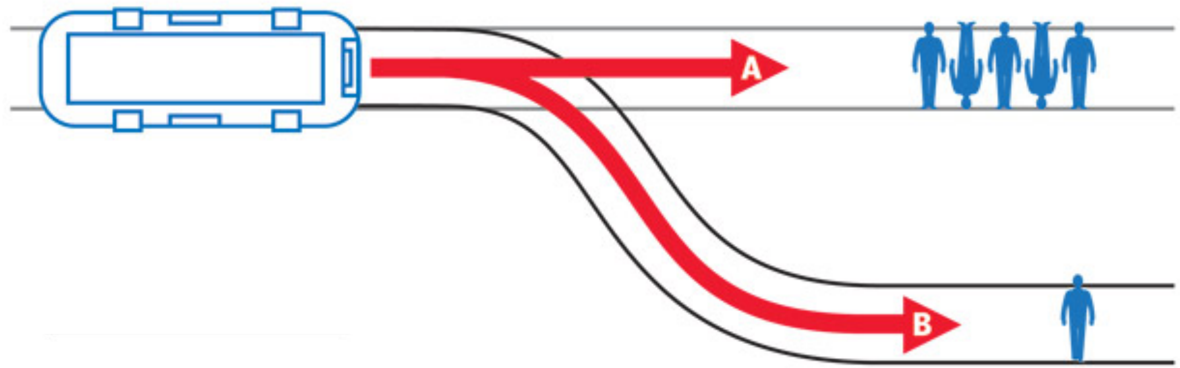


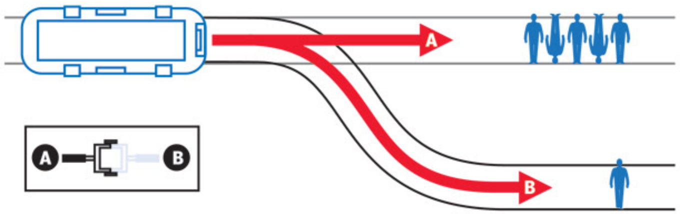
How morally permissible is it
to push the man?

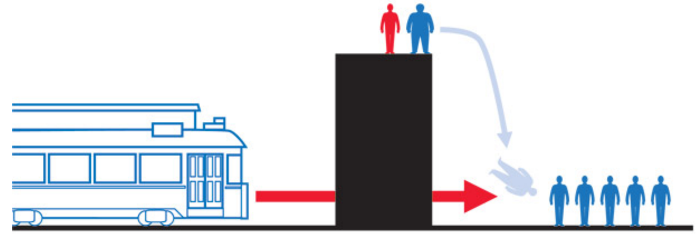
never 1 2 3 4 5 6 7 always

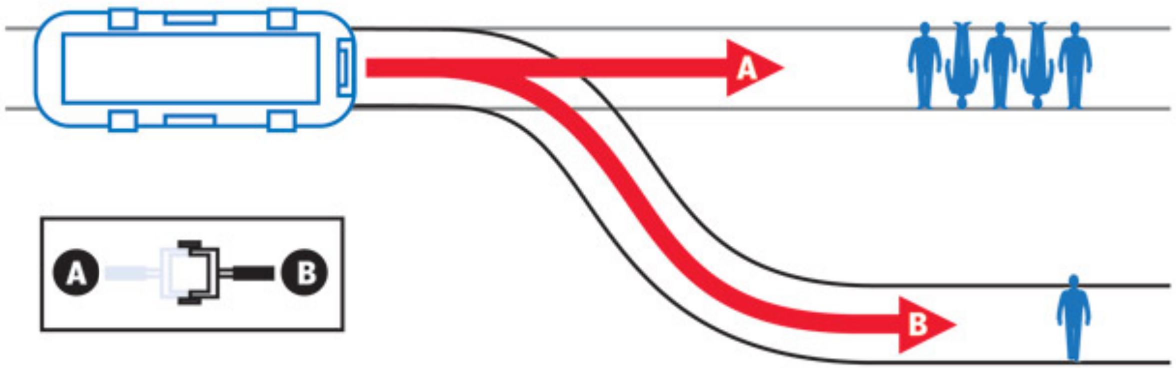

A
B
A
B

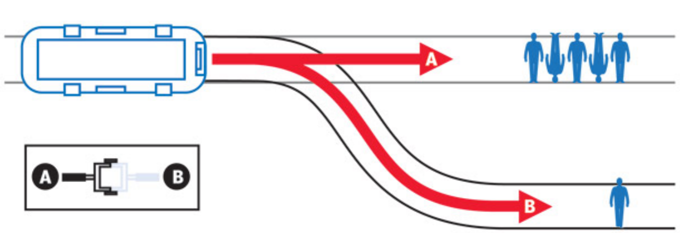


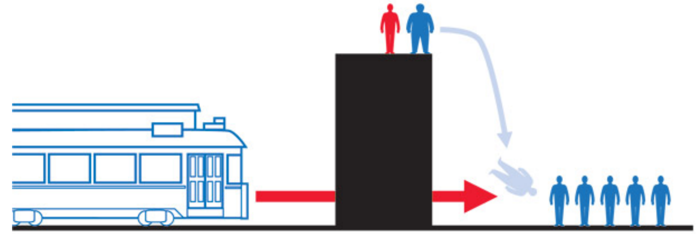

A
B

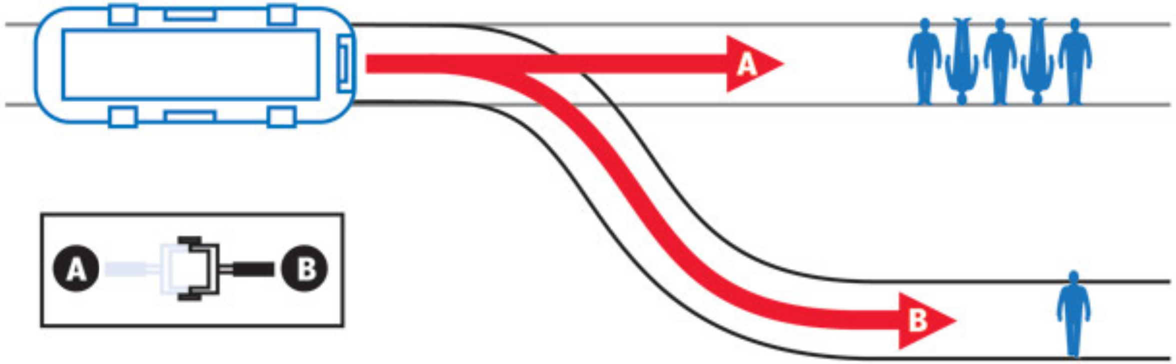

A
B
A
B




A
B
A
B







How morally permissible is it
to not pull the lever?

never 1 2 3 4 5 6 7 always

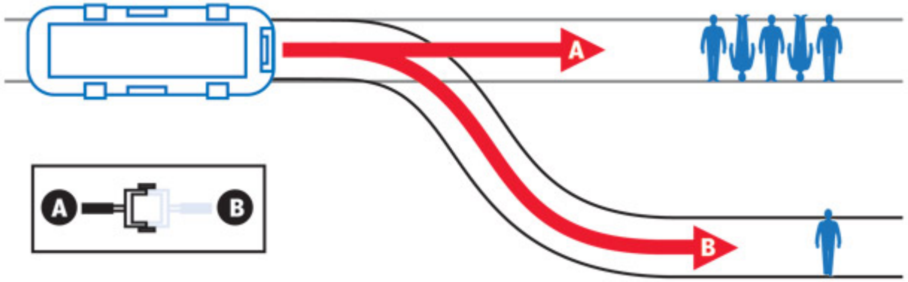
Three principles

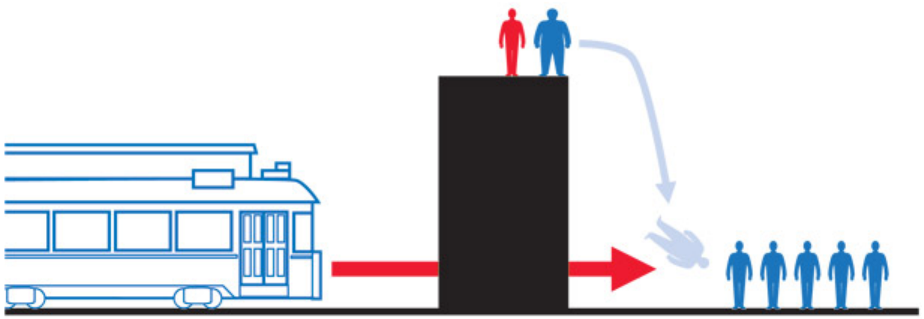
- ***Action***: Harm caused by action is morally worse than same harm caused by inaction.

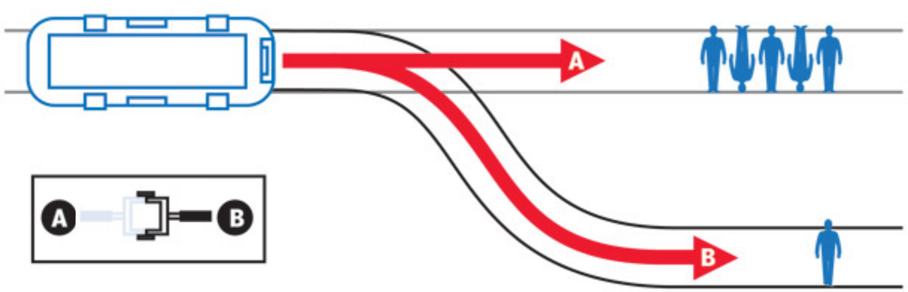
- ***Intention***: Harm intended as means to goal worse than same harm foreseen as a side effect of goal.

- ***Contact***: Harm caused by physical contact worse than same harm without physical contact.

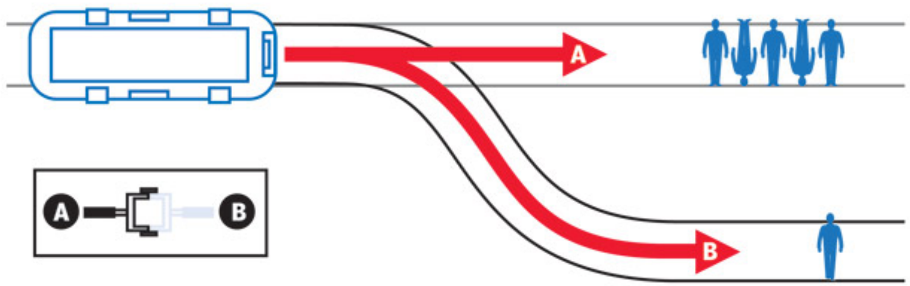
action intention contact

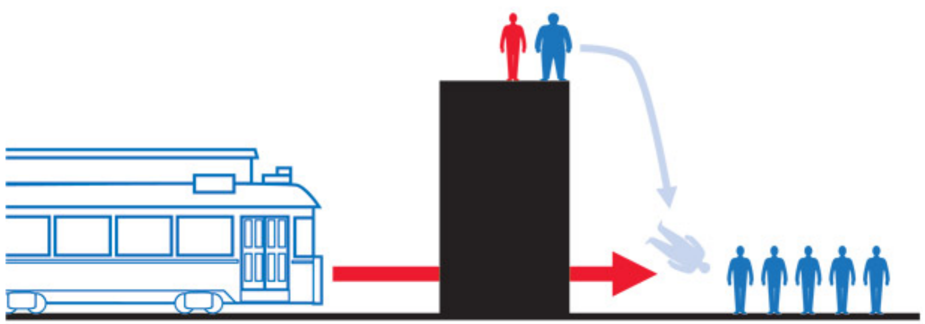


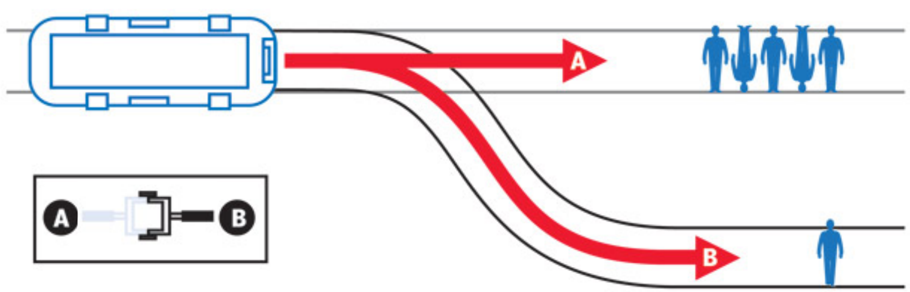


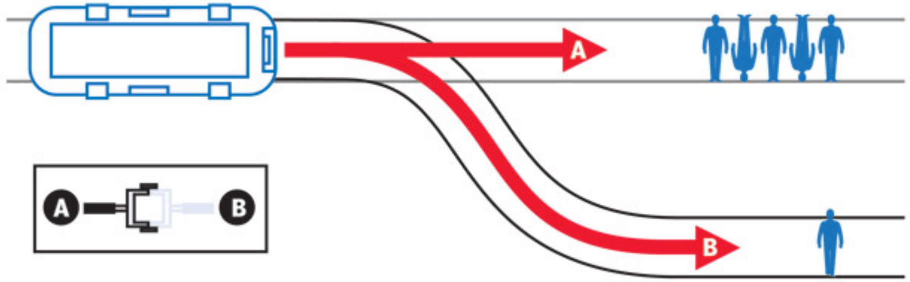
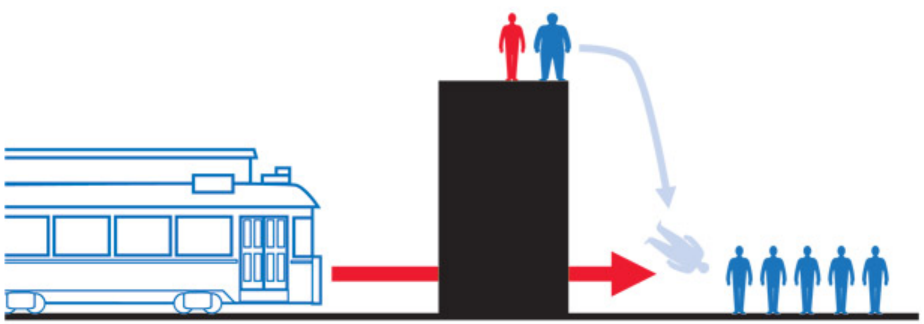
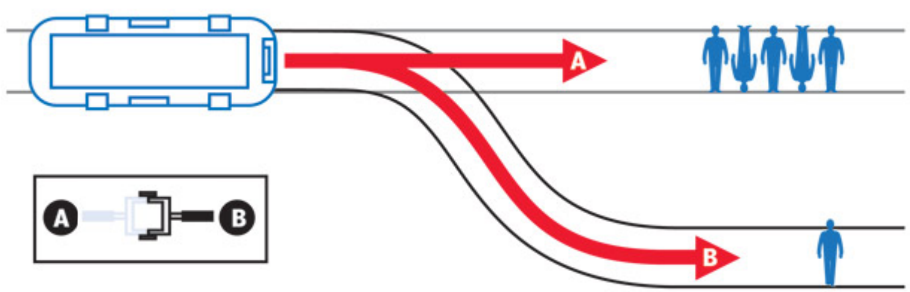


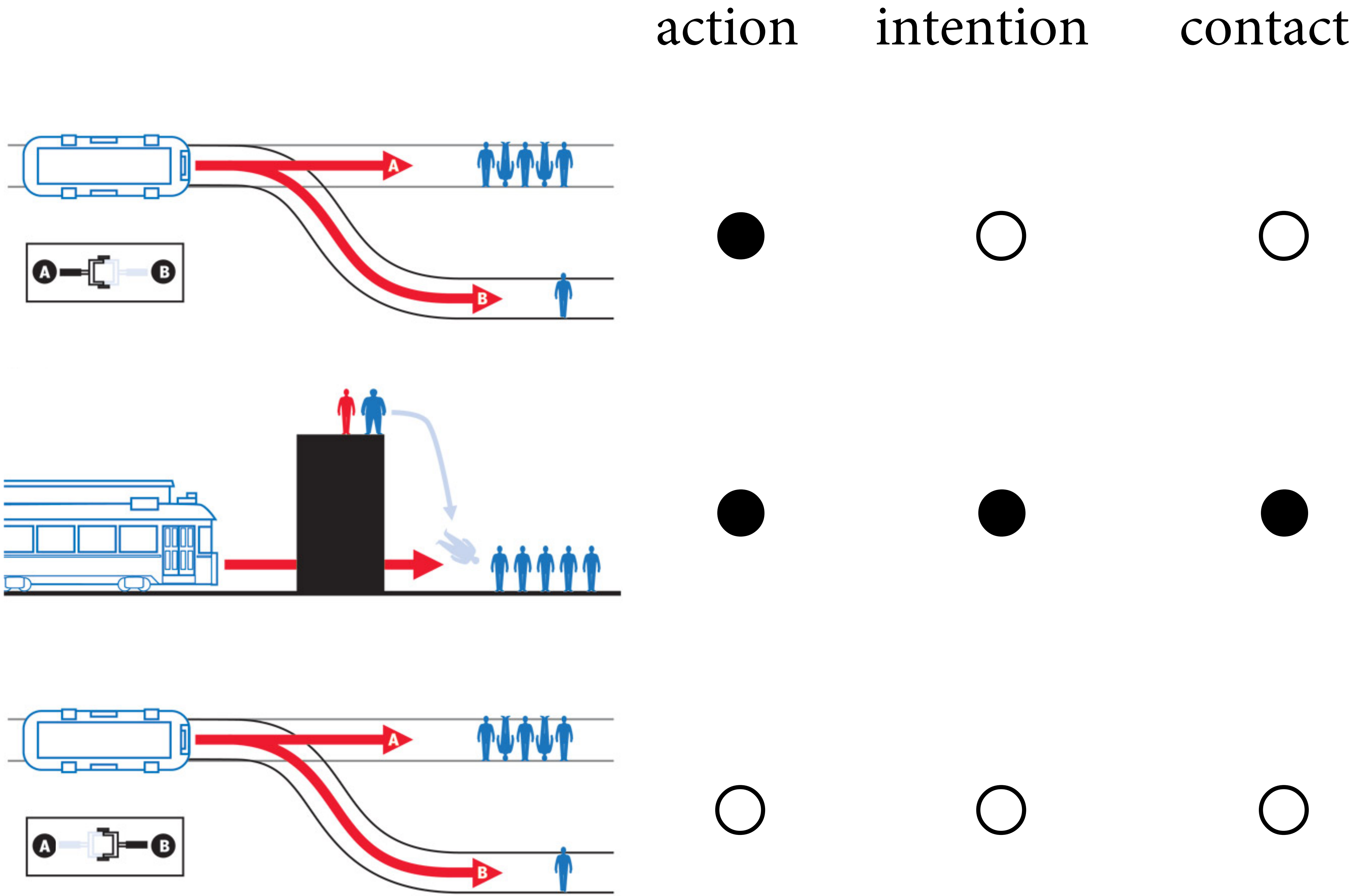
action	intention	contact
●	○	○





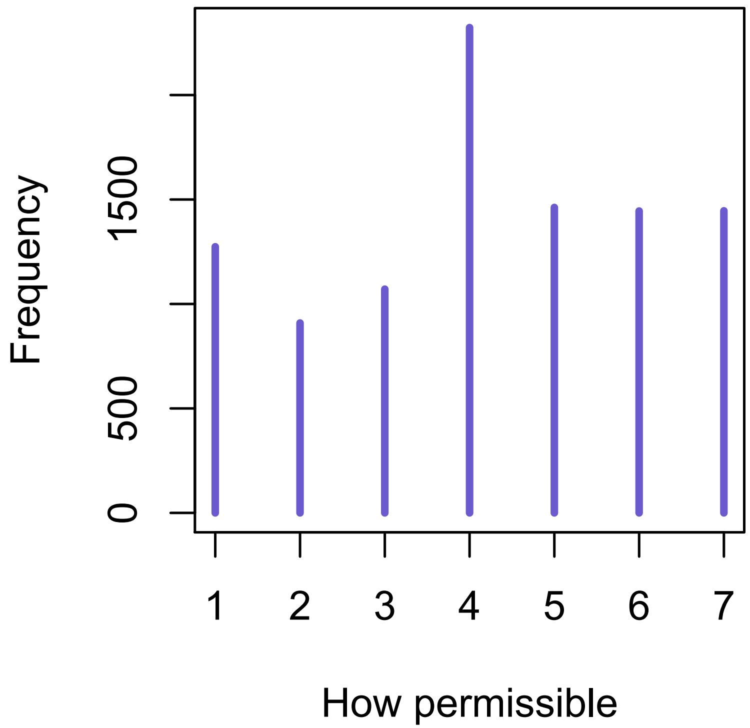


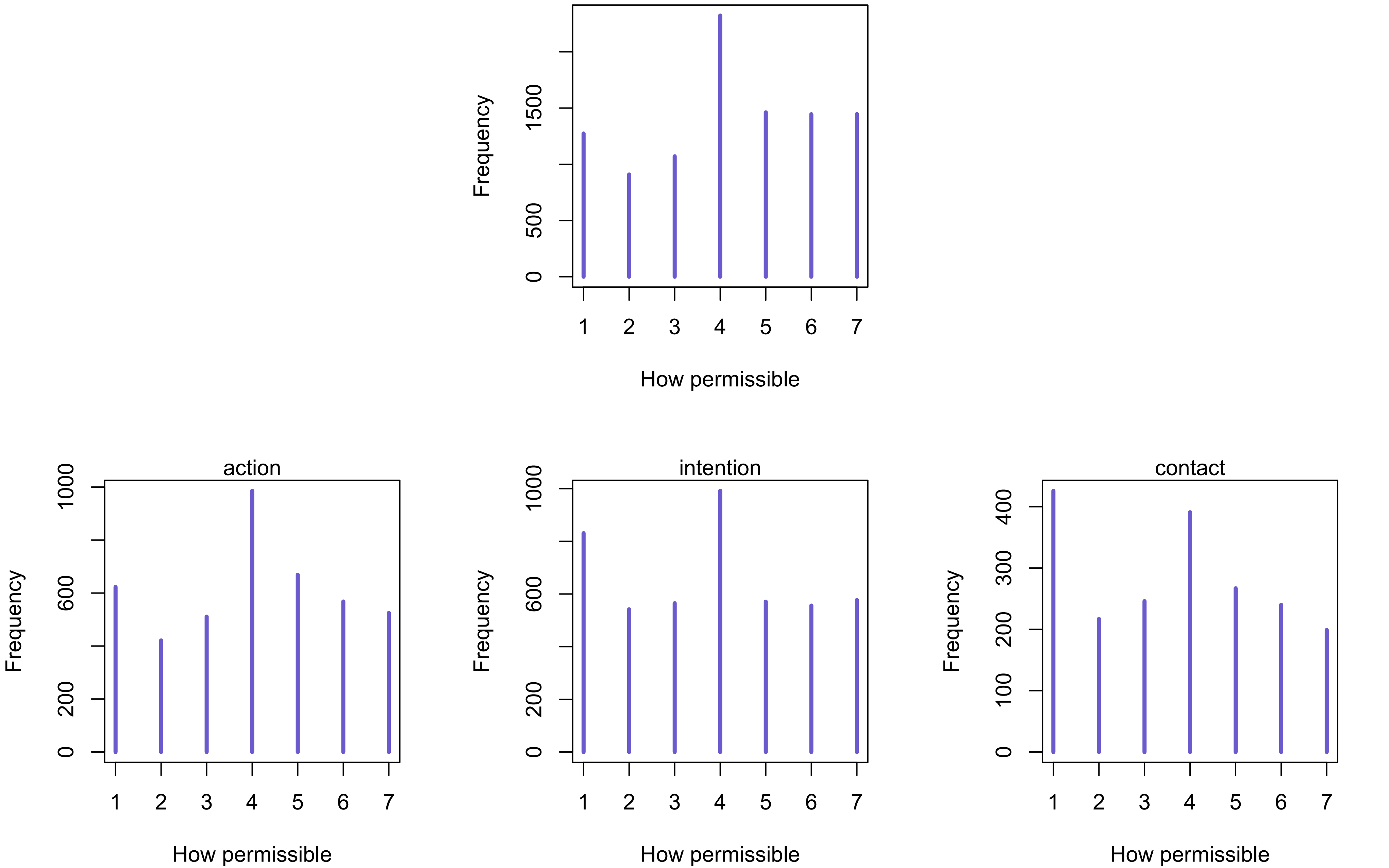
	action	intention	contact
	●	○	○
	●	●	●
			


action
intention
contact
A
B
A
B
A
B
A
B

Moral intuitions

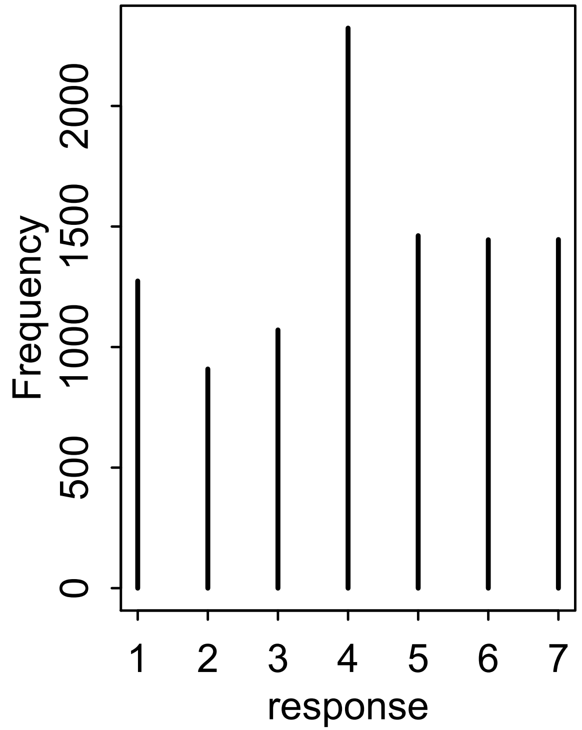
- Cushman et al. experiments
- 331 individuals, 30 scenarios, 9930 responses
- How do responses vary with action, intention, contact?
- Age, gender, individual?




Frequency
0
500
1500
1
2
3
4
5
6
7
How permissible
action
intention
contact
1000
600
200
400
300
100

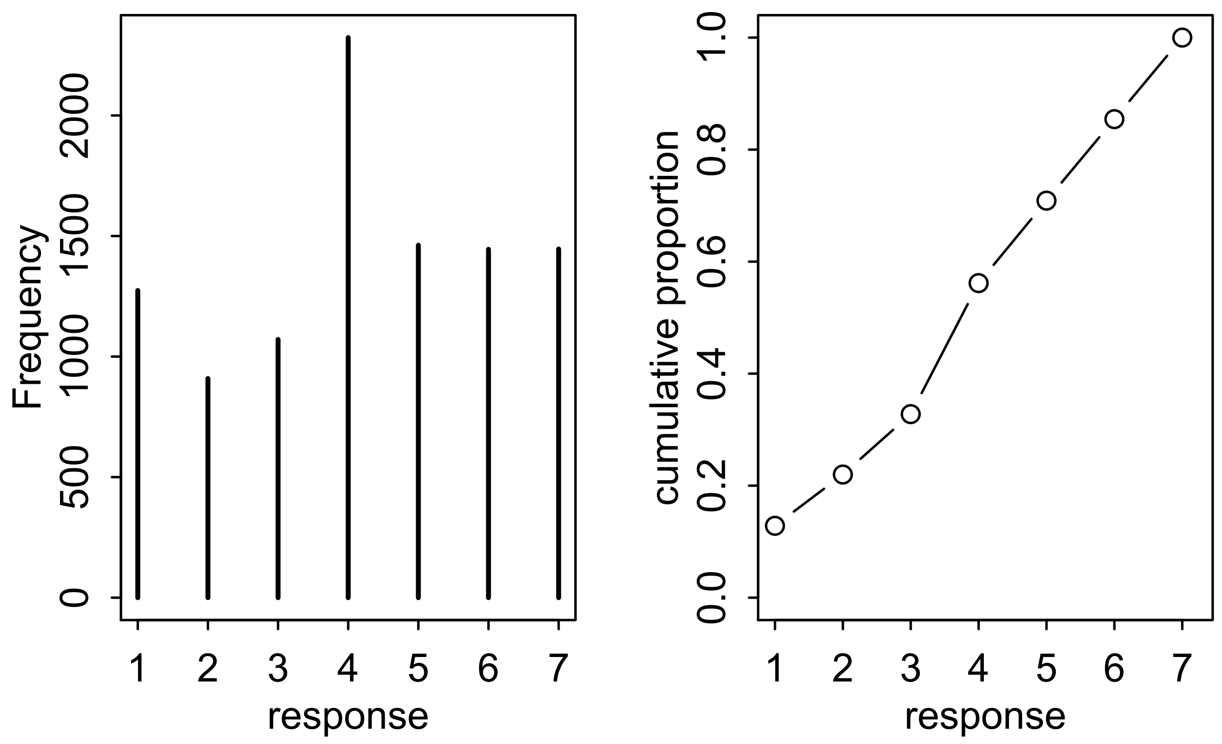
Ordered logit

- A log-cumulative-odds link probability model



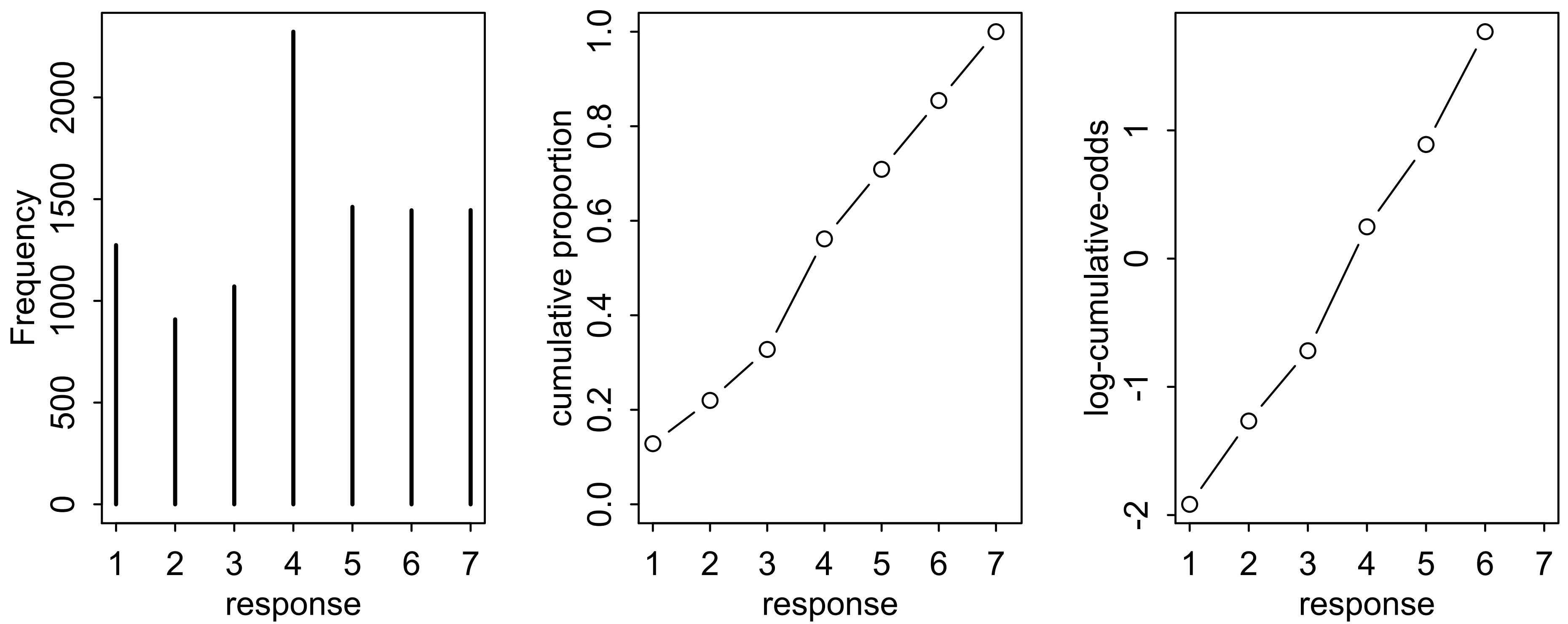
Ordered logit

- A log-cumulative-odds link probability model



Ordered logit

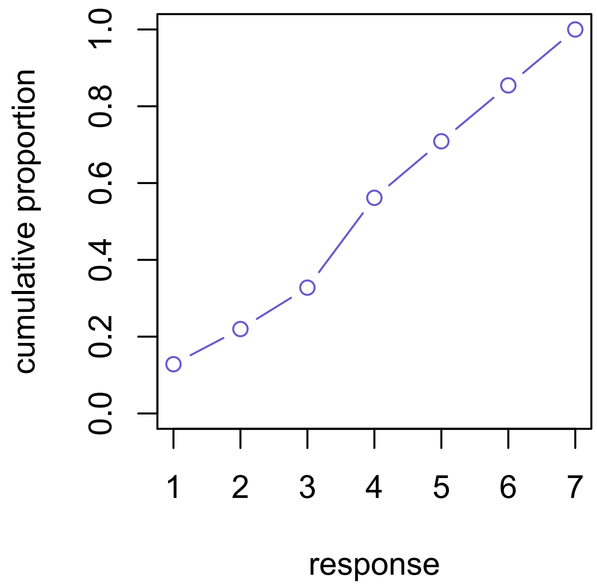
- A log-cumulative-odds link probability model



Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \le k)}{1 - \Pr(y_i \le k)} = \phi_k$$

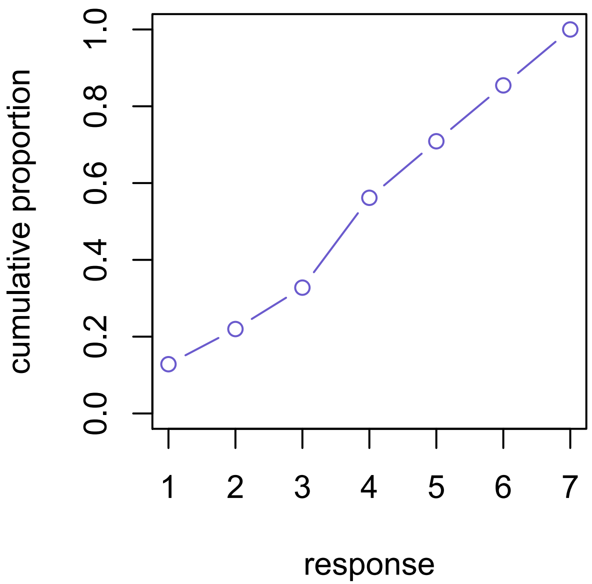


Ordered logit

- A log-cumulative-odds link probability model

cumulative log-odds

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$



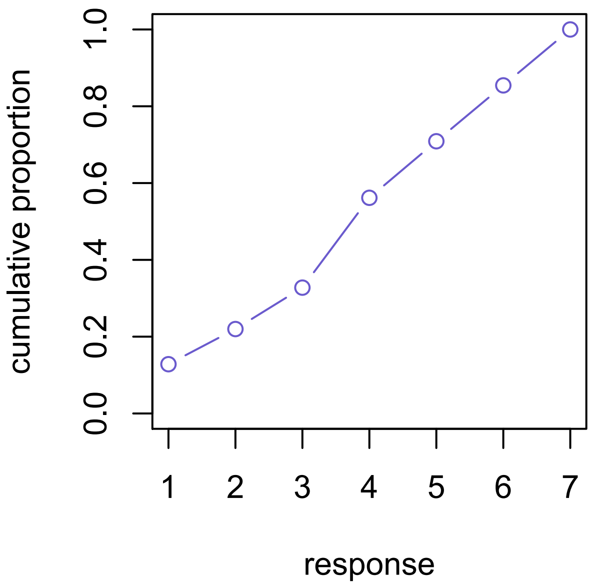
Ordered logit

- A log-cumulative-odds link probability model

response

cumulative log-odds

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$



Ordered logit

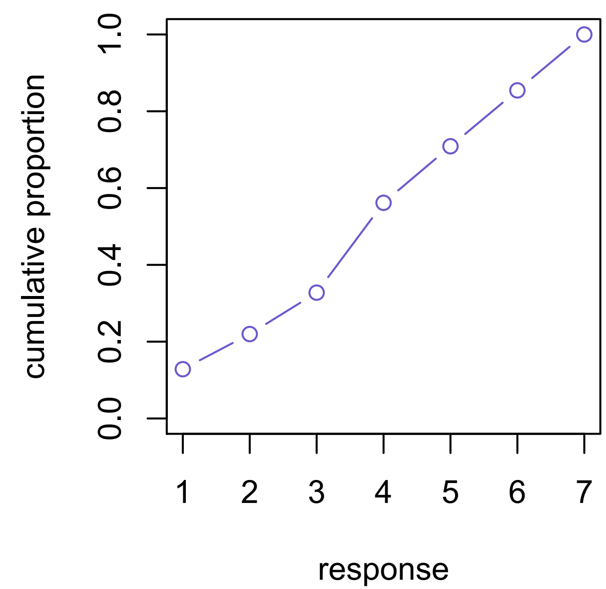
- A log-cumulative-odds link probability model

category

response

cumulative log-odds

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$



Ordered logit

- A log-cumulative-odds link probability model

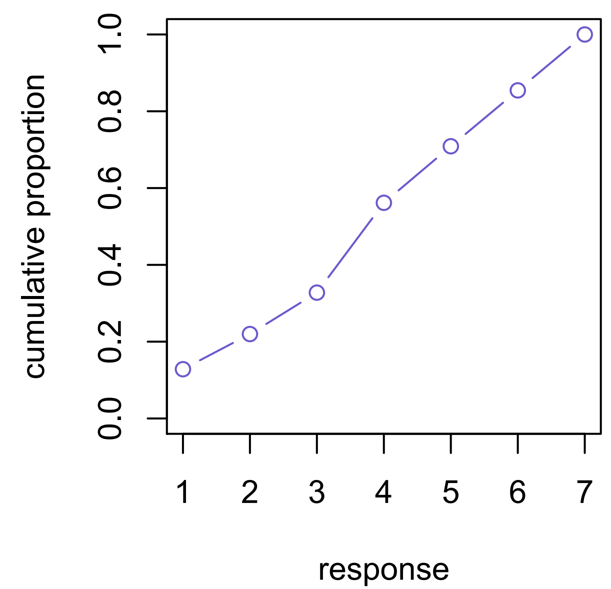
category

response

cumulative log-odds

$$\log \frac{\Pr(y_i \le k)}{1 - \Pr(y_i \le k)} = \phi_k$$

linear model

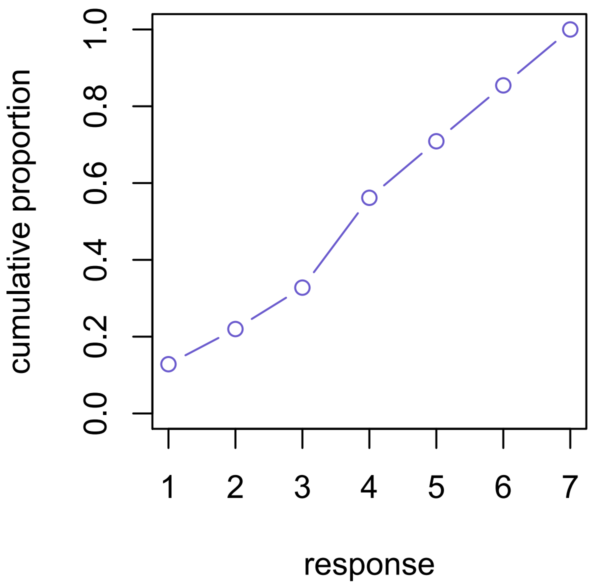


Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

$$\Pr(y_i \leq k) = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$

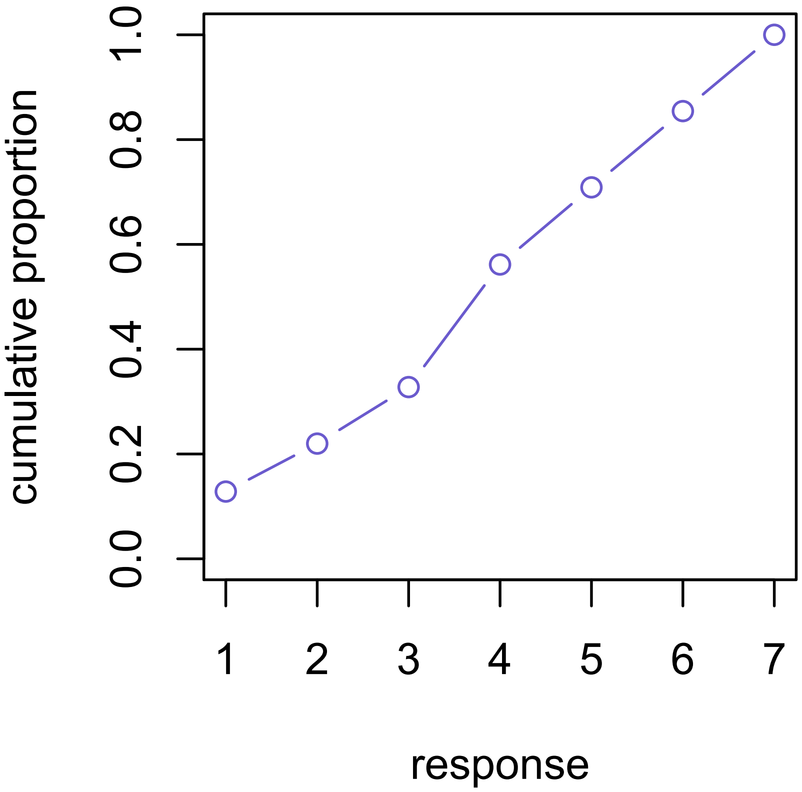


Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \le k)}{1 - \Pr(y_i \le k)} = \phi_k$$

$$\Pr(y_i \le k) = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$

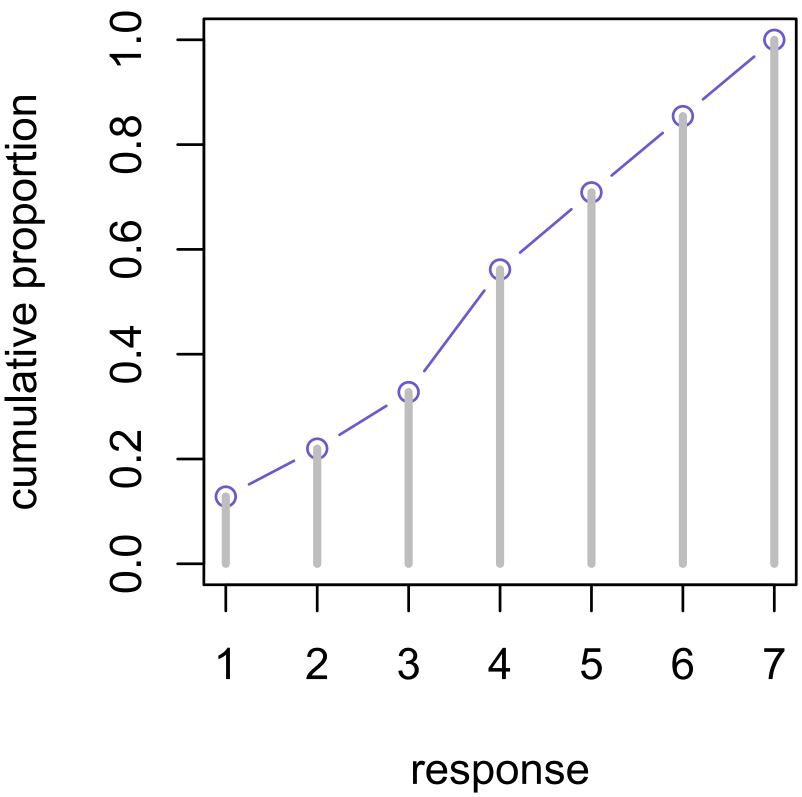


Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

$$\boxed{\Pr(y_i \leq k)} = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$



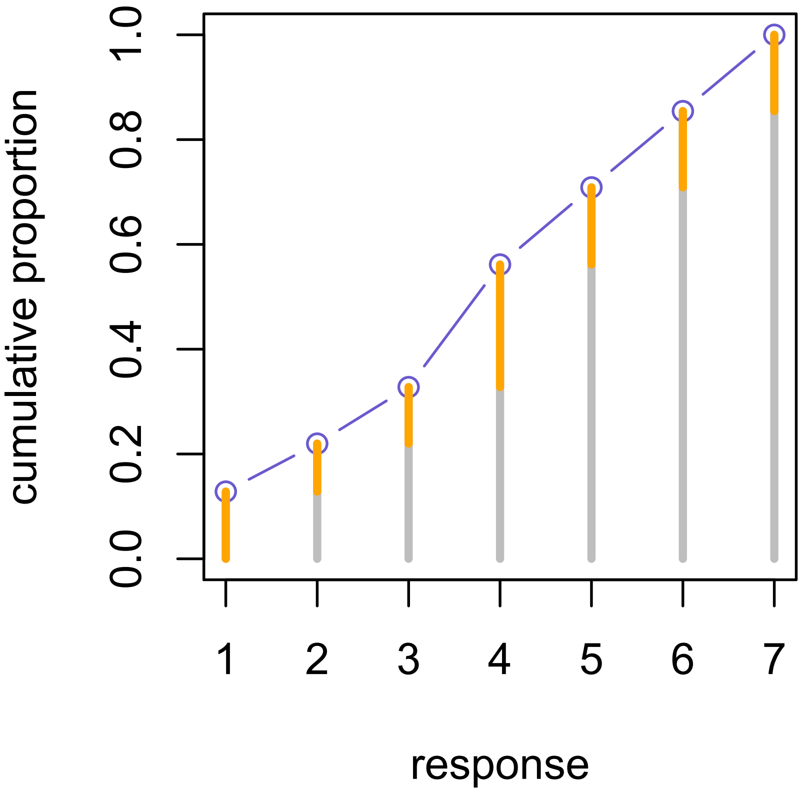
Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

$$\boxed{\Pr(y_i \leq k)} = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$

$$\boxed{\Pr(y_i = k)} = \Pr(y_i \leq k) - \Pr(y_i \leq k - 1)$$



Ordered logit

- Simplest model just uses an intercept for each category:

cumulative probabilities of each response

$$
\begin{aligned}
R_i &\sim \text{Ordered}(\mathbf{p}) \\
\log \frac{p_k}{1 - p_k} &= \alpha_k \\
\alpha_k &\sim \text{Normal}(0, 10)
\end{aligned}
$$

intercept unique to category

Ordered logit in Stan

R code 11.8

```
# note that data with name 'case' not allowed in Stan
# so will pass pruned data list
m11.1stan <- map2stan(
    alist(
        response ~ dordlogit( phi , cutpoints ),
        phi <- 0,
        cutpoints ~ dnorm(0,10)
    ) ,
    data=list(response=d$response),
    start=list(cutpoints=c(-2,-1,0,1,2,2.5)) ,
    chains=2 , cores=2 )

# need depth=2 to show vector of parameters
precis(m11.1stan,depth=2)
```

```
              Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
cutpoints[1] -1.92   0.03      -1.97      -1.87  1012    1
cutpoints[2] -1.27   0.02      -1.31      -1.23  1461    1
cutpoints[3] -0.72   0.02      -0.75      -0.68  1845    1
cutpoints[4]  0.25   0.02       0.22       0.28  2000    1
cutpoints[5]  0.89   0.02       0.85       0.92  2000    1
cutpoints[6]  1.77   0.03       1.72       1.81  1851    1
```

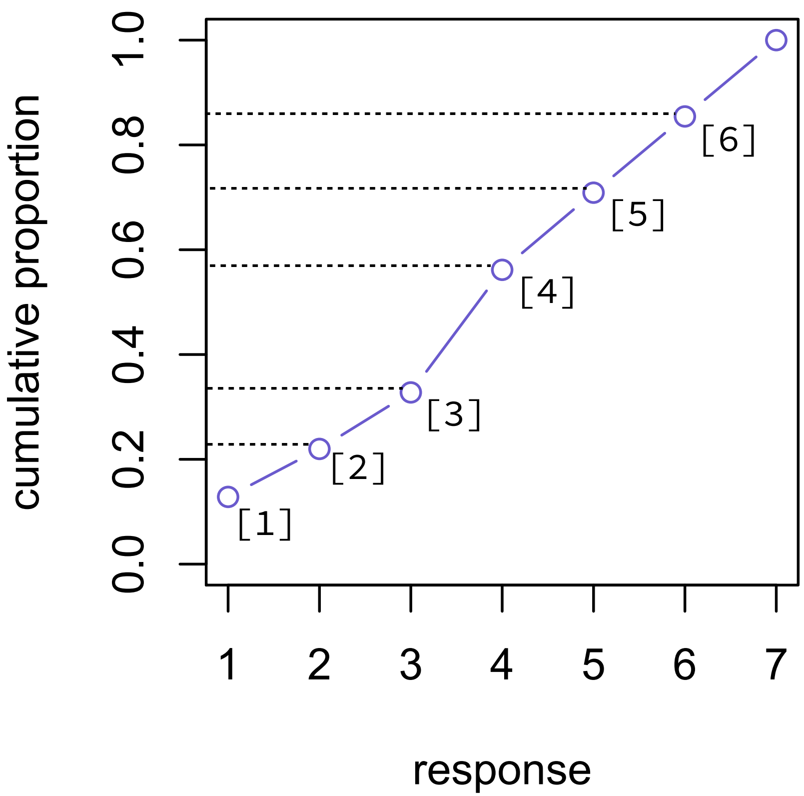
```
logistic( coef( m11.1stan ) )
```

```
cutpoints[1] cutpoints[2] cutpoints[3] cutpoints[4] cutpoints[5] cutpoints[6]
   0.1281697    0.2198018    0.3276686    0.5617471    0.7091352    0.8546406
```

7th cutpoint missing, because known to be infinity (on logit scale)

```
logistic( coef( m11.1stan ) )
```

```
cutpoints[1] cutpoints[2] cutpoints[3] cutpoints[4] cutpoints[5] cutpoints[6]
   0.1281697    0.2198018    0.3276686    0.5617471    0.7091352    0.8546406
```



Adding predictor variables

In general:

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \alpha_k - \phi_i$$

$$\phi_i = \beta x_i$$

Trolley data:

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \alpha_k - \phi_i$$

$$\phi_i = \beta_A A_i + \beta_I I_i + \beta_C C_i$$

NO INTERCEPT in *phi*!

Adding predictor variables

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \alpha_k - \phi_i$$

$$\phi_i = \boxed{\beta_A A_i + \beta_I I_i + \beta_C C_i}$$

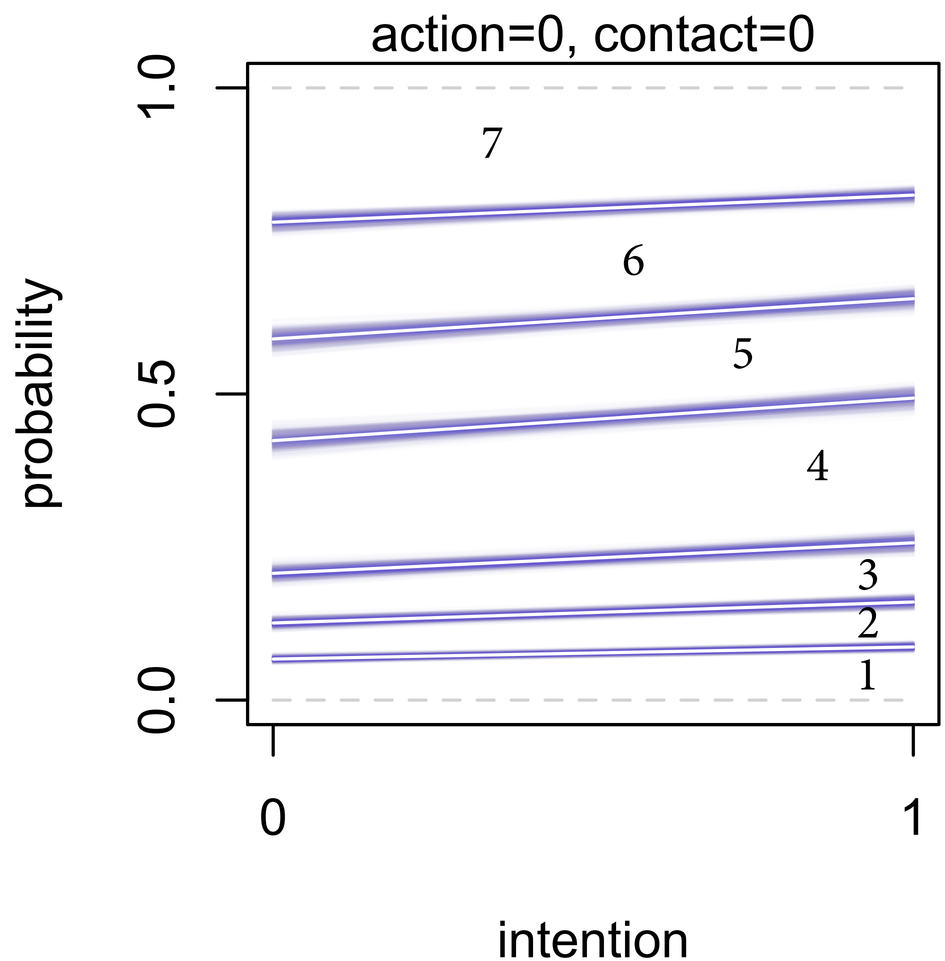
R code 11.13

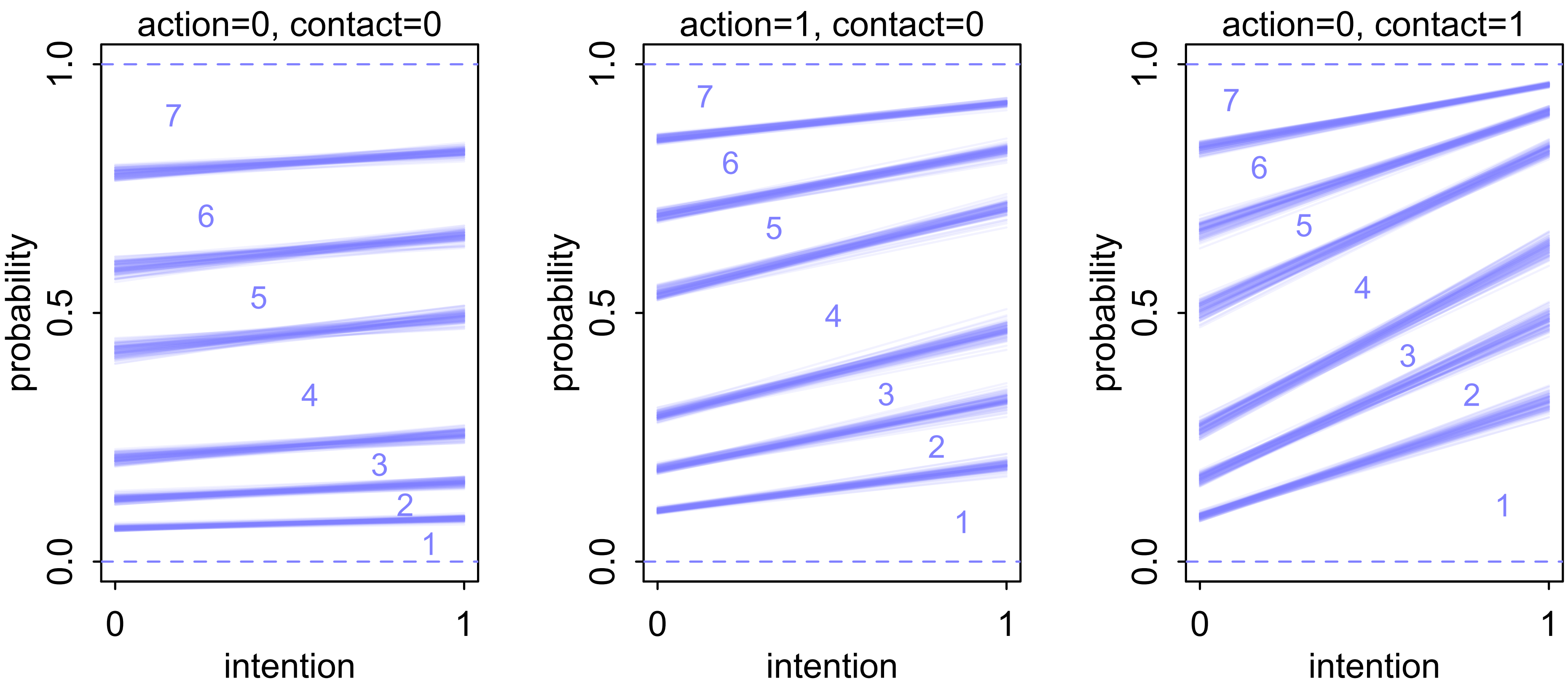
```
m11.2 <- map(
    alist(
        response ~ dordlogit( phi , c(a1,a2,a3,a4,a5,a6) ) ,
        phi <- bA*action + bI*intention + bC*contact,
        c(bA,bI,bC) ~ dnorm(0,10),
        c(a1,a2,a3,a4,a5,a6) ~ dnorm(0,10)
    ) ,
    data=d ,
    start=list(a1=-1.9,a2=-1.2,a3=-0.7,a4=0.2,a5=0.9,a6=1.8) )
```

Plotting ordered logits

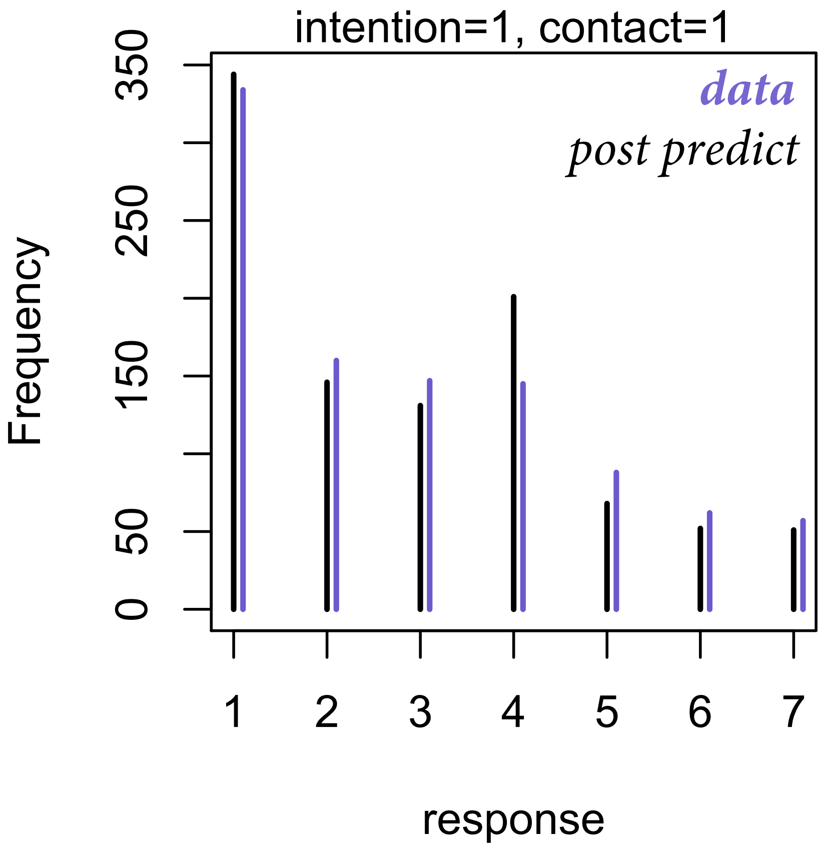
- Oh, bother: Posterior prediction a *vector* of probabilities, one for each level of outcome

- How to plot this?

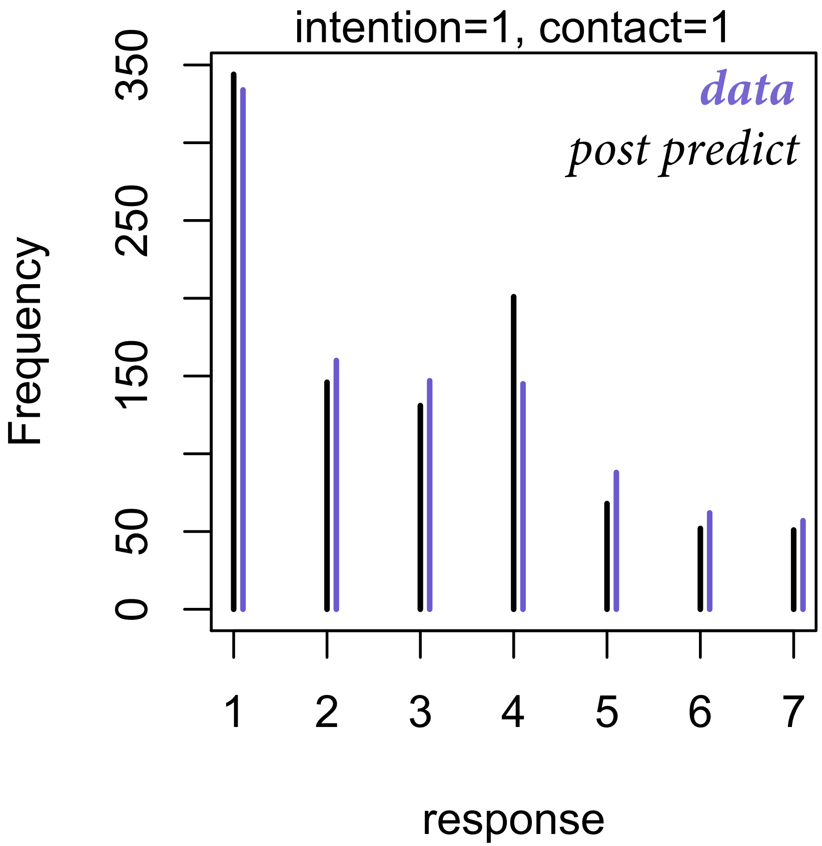

action=0, contact=0
probability
1.0
0.5
0.0
7
6
5
4
3
2
1
0
1
intention


action=0, contact=0
action=1, contact=0
action=0, contact=1
probability
intention
0.0
0.5
1.0
0
1
1
2
3
4
5
6
7

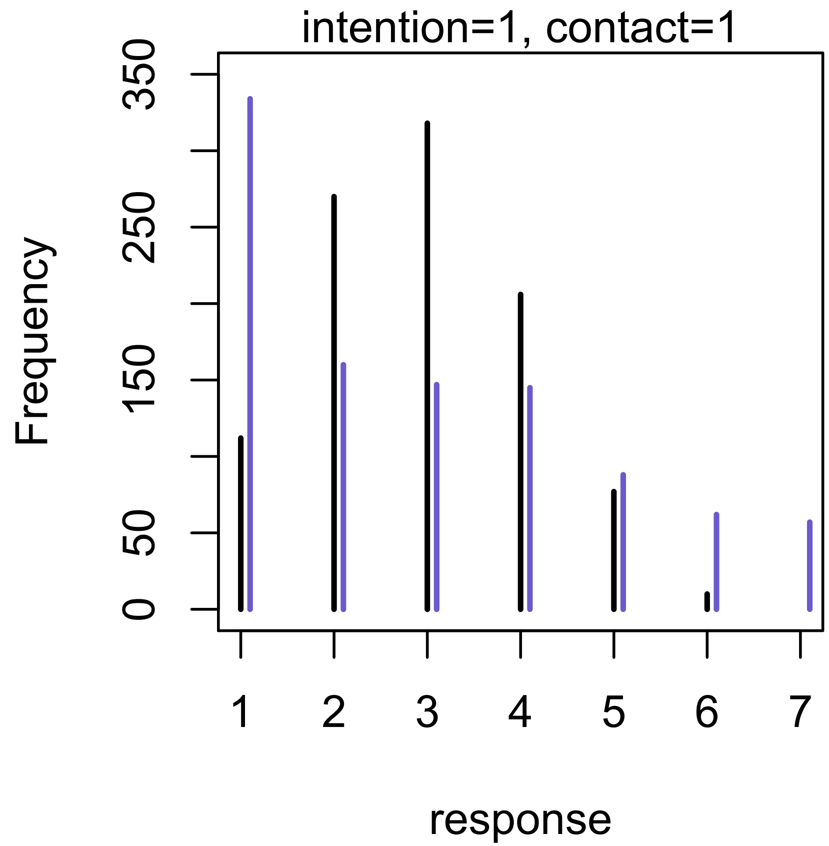
ordered logit



ordered logit



binomial



Ordered logit

- MAP estimation can be hard; choose good starting values. See notes for details.

- Stan handles these models fine. Will be slower than other outcome types.

- Also *ordered probit*; uses cumulative normal link

Mixtures

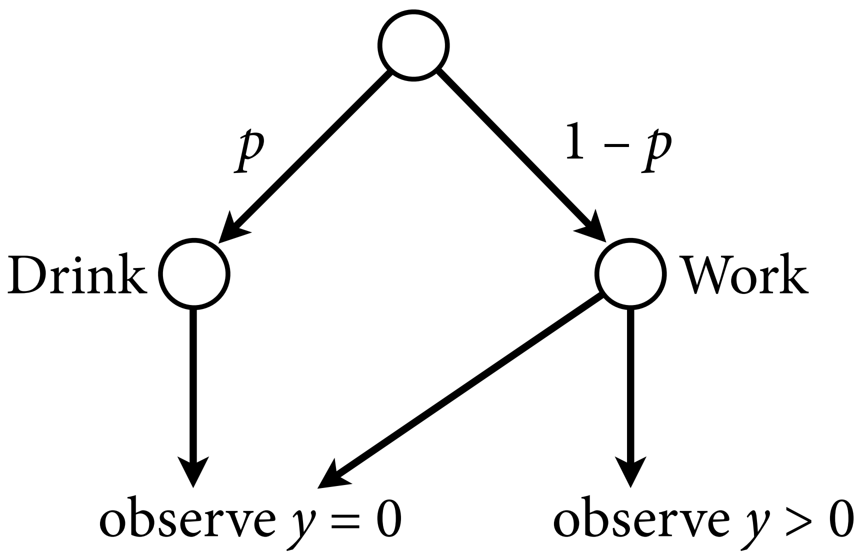
- Some outcomes mix different processes
  - replace parameter of likelihood with distribution of its own
- Over-dispersion: counts often more variable than expected, because probabilities/rates are variable
  - beta-binomial, gamma-Poisson (negative-binomial)
- Zero-inflated mixtures

Monastery Mystery

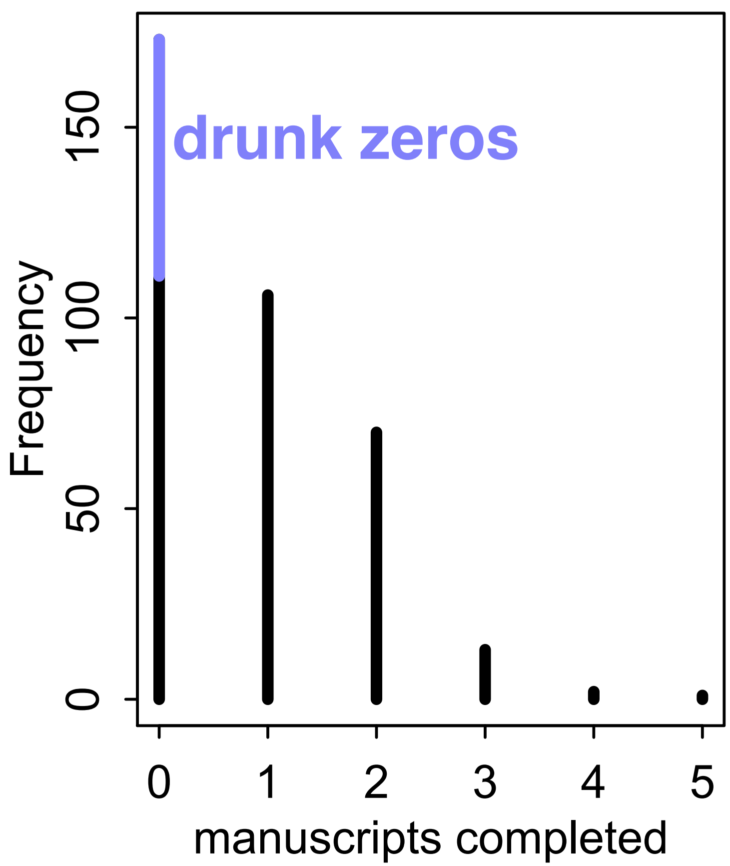
- Monks copy manuscripts
- They also get drunk
- Data: num manuscripts completed each day
- Can infer number of days they got drunk?

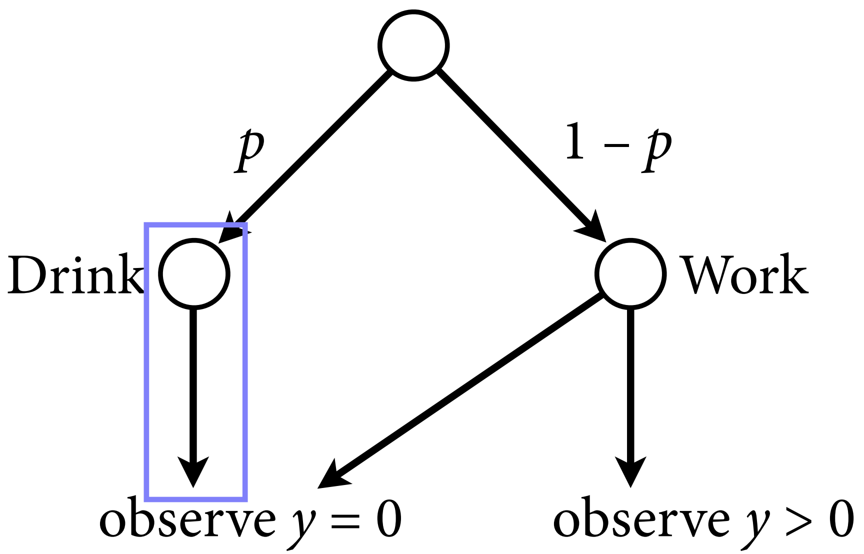
Analyze?

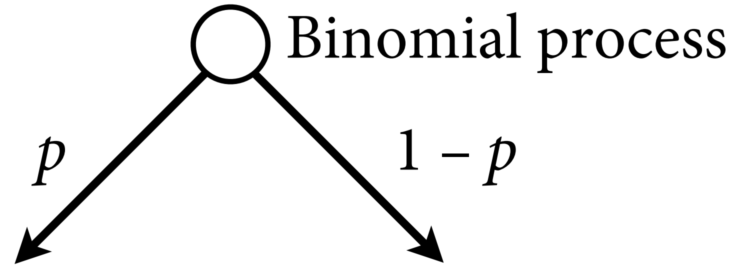
- Zero-inflated Poisson observations
- Hidden state: drunk or sober
- Can estimate probability of drinking and rate of production when sober
- Must build a new likelihood, a mixture of stochastic processes

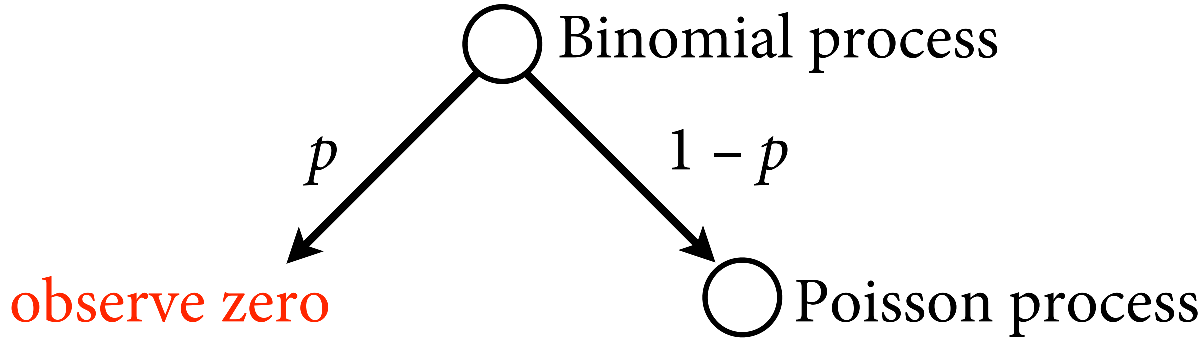


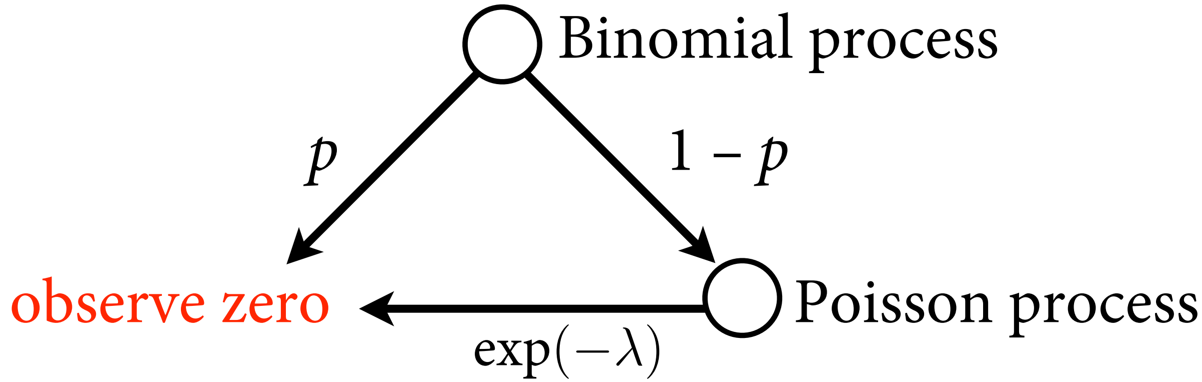
Analyze?

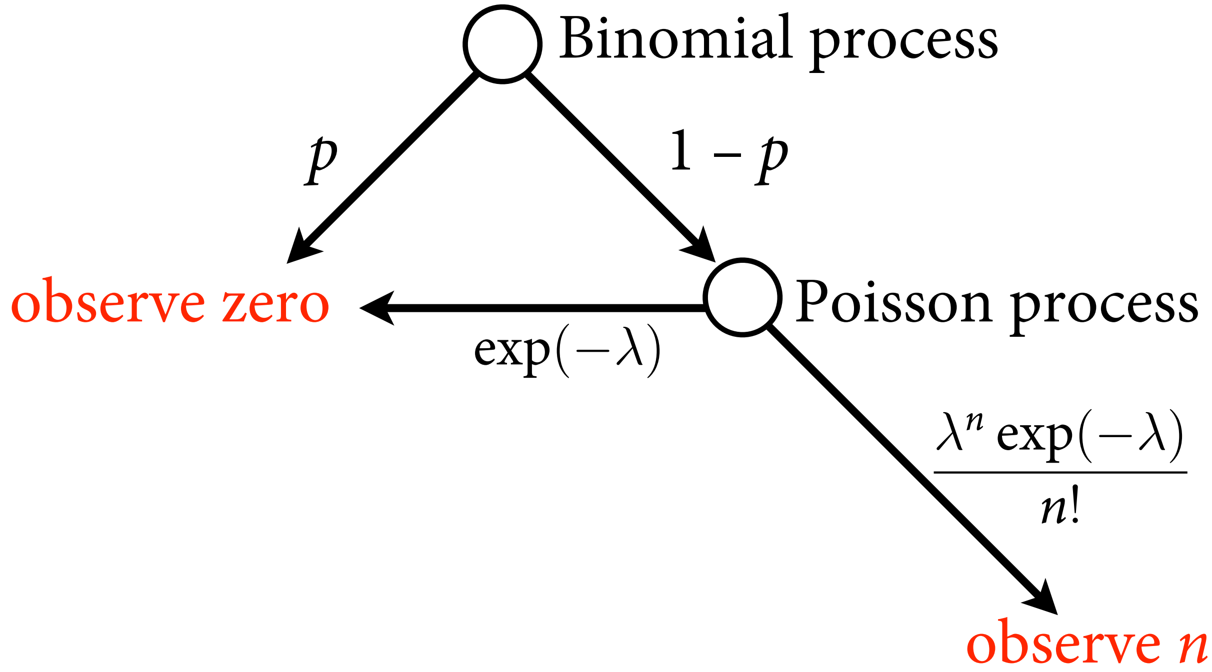

drunk zeros
Frequency
0
50
100
150
0 1 2 3 4 5
manuscripts completed

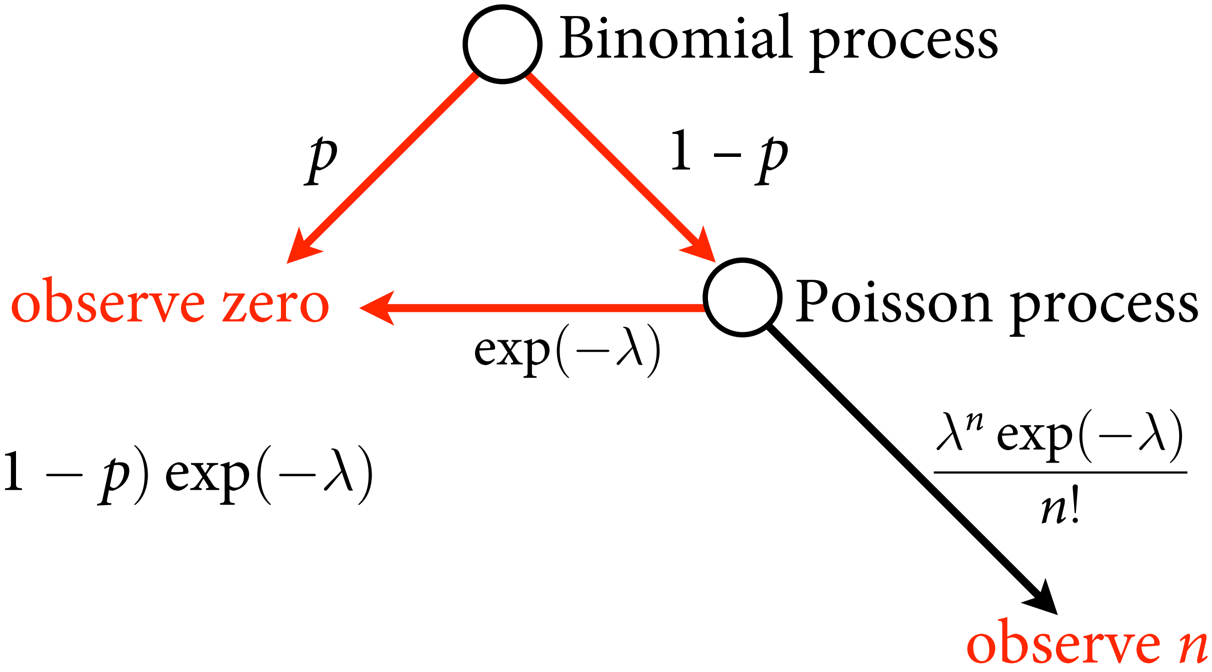

p
1 – p
Drink
Work
observe y = 0
observe y > 0


Binomial process
p
1 – p

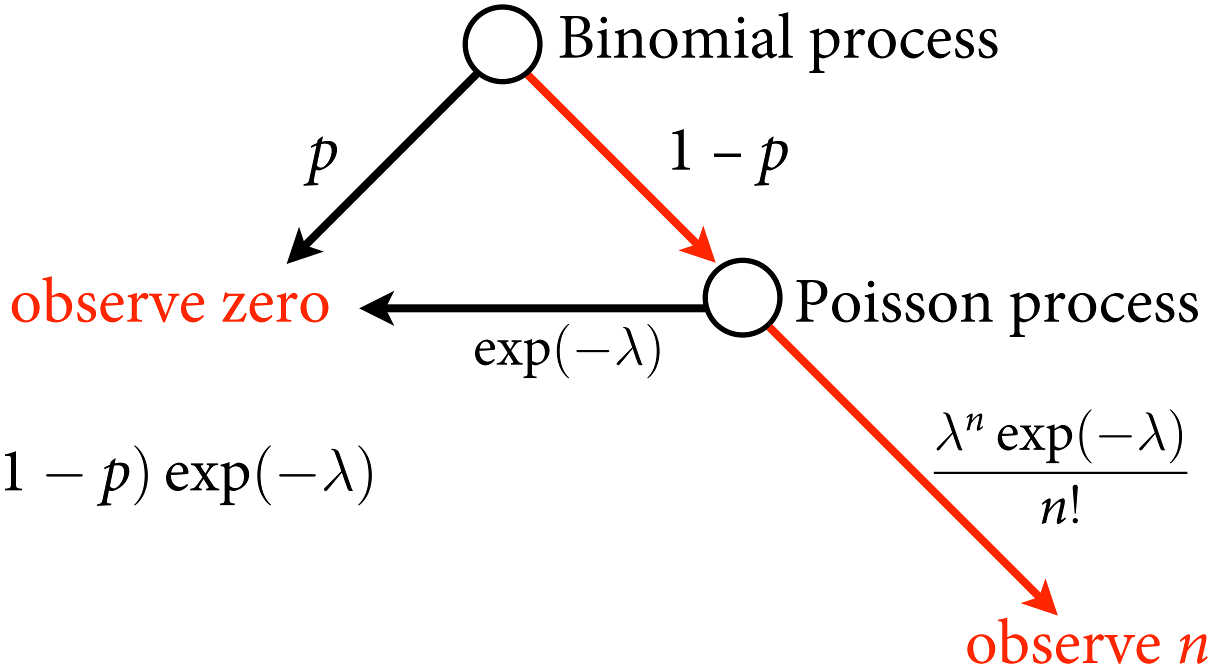

Binomial process
p
1 – p
observe zero
Poisson process


Binomial process
p
$1 - p$
observe zero
Poisson process
$\exp(-\lambda)$


Binomial process
p
$1 - p$
observe zero
Poisson process
$\exp(-\lambda)$
$\dfrac{\lambda^n \exp(-\lambda)}{n!}$
observe n



$$\Pr(0|p,\lambda) = p + (1-p)\exp(-\lambda)$$



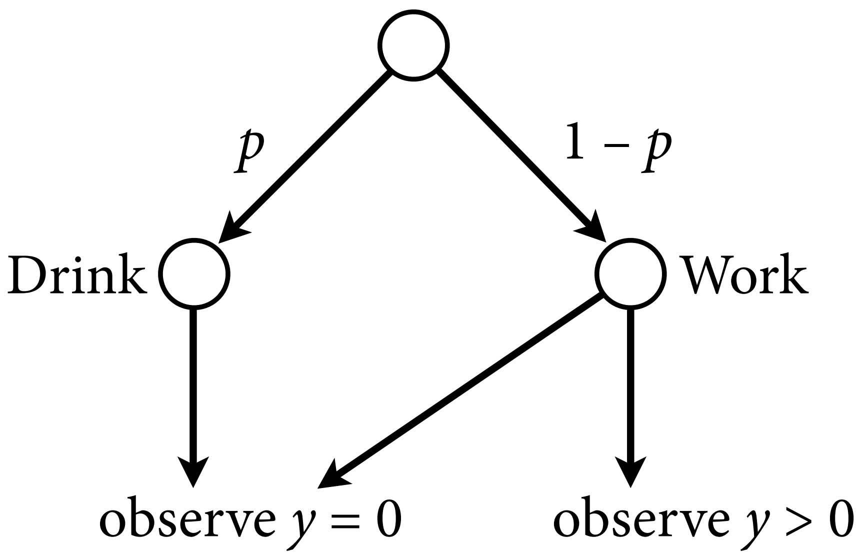
$$\Pr(0|p, \lambda) = p + (1 - p)\exp(-\lambda)$$

$$\Pr(n|p, \lambda) = (1 - p)\frac{\lambda^n \exp(-\lambda)}{n!}$$

Zero-inflated Poisson model

$$
\begin{aligned}
y_i &\sim \text{ZIPoisson}(p_i, \lambda_i) \\
\text{logit}(p_i) &= \alpha_p + \beta_p x_i \\
\log(\lambda_i) &= \alpha_\lambda + \beta_\lambda x_i
\end{aligned}
$$

Linear models are independent



Simulate, validate, cromulate

- As models get more complicated, no guarantees you can
  - specify model correctly
  - estimate actual process reliably
  - Bayes not magic, just logic
- Simulate "dummy data"
  - recover estimates
  - understand the model
- Try parameter combinations hostile to estimation, so you know limits of the golem

Simulated manuscripts

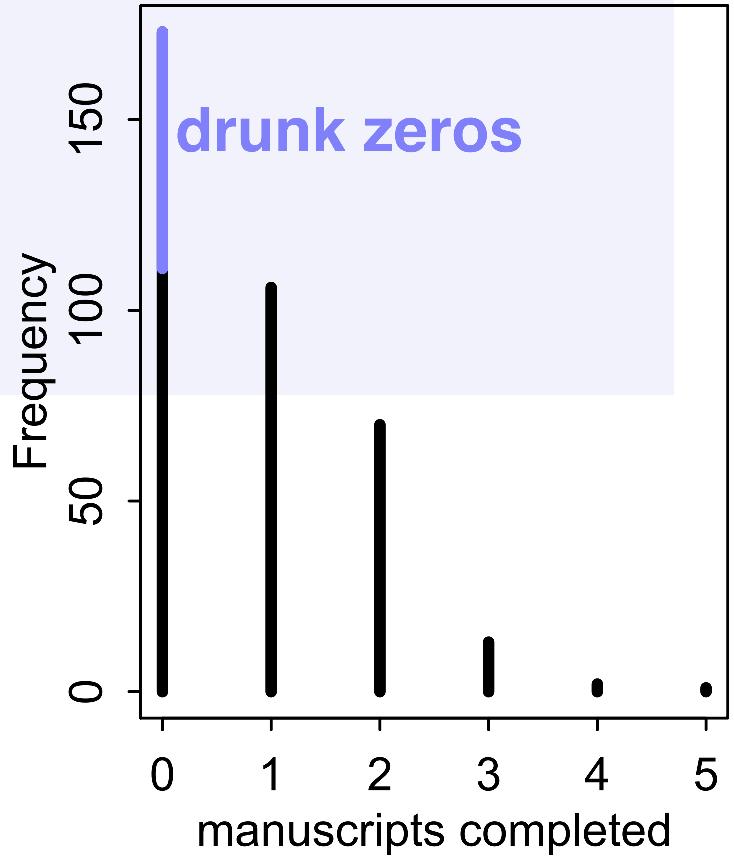
R code 11.20

```
# define parameters
prob_drink <- 0.2 # 20% of days
rate_work <- 1    # average 1 manuscript per day

# sample one year of production
N <- 365

# simulate days monks drink
drink <- rbinom( N , 1 , prob_drink )

# simulate manuscripts completed
y <- (1-drink)*rpois( N , rate_work )
```



Fit model to dummy data

$$\begin{aligned} y_i &\sim \text{ZIPoisson}(p_i, \lambda_i) \\ \text{logit}(p_i) &= \alpha_p \\ \log(\lambda_i) &= \alpha_\lambda \end{aligned}$$

```
m11.4 <- map(
    alist(
        y ~ dzipois( p , lambda ),
        logit(p) <- ap,
        log(lambda) <- al,
        ap ~ dnorm(0,1),
        al ~ dnorm(0,10)
    ) ,
    data=list(y=y) )
```

R code
11.22

```
m11.4 <- map(
    alist(
        y ~ dzipois( p , lambda ),
        logit(p) <- ap,
        log(lambda) <- al,
        ap ~ dnorm(0,1),
        al ~ dnorm(0,10)
    ) ,
    data=list(y=y) )
precis(m11.4)
```

R code 11.22

```
    Mean StdDev 2.5% 97.5%
ap -1.39   0.31 -2.0 -0.78
al  0.05   0.08 -0.1  0.21
```

```
logistic(-1.39) # probability drink
exp(0.05)       # rate finish manuscripts, when not drinking
```

R code 11.23

```
[1] 0.1994078
[1] 1.051271
```

$$y_i \sim \text{ZIPoisson}(p_i, \lambda_i)$$
$$\text{logit}(p_i) = \alpha_p$$
$$\log(\lambda_i) = \alpha_\lambda$$

Other mixtures

- Can ZIBinomial, too
- Also “hurdle” models, aka zero-augmented
- Continuous mixtures for overdispersed counts
  - beta-binomial
  - gamma-Poisson
  - We’ll focus on multilevel models instead

Holiday Labor

- Homework: 10H3, 11H1
- Resume on January 3: Chapter 12, Multilevel Models