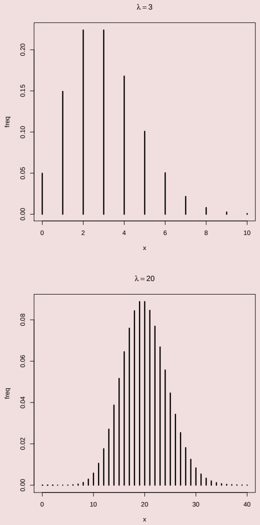
Chapter 7,9,10

$$
\begin{aligned}
y &\sim Po(\lambda), \\
\mathbb{E}[y] &= \lambda, \\
\mathbb{V}[y] &= \lambda.
\end{aligned}
$$



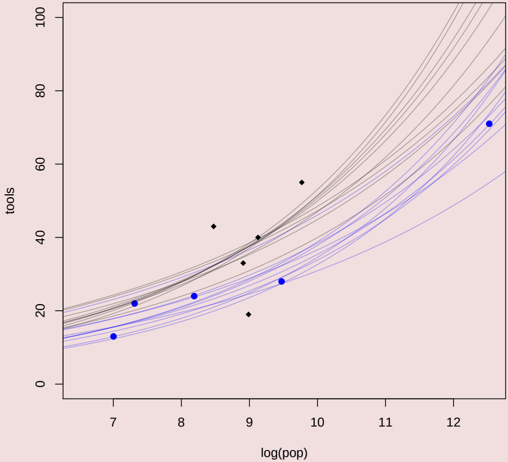
- Counts (non negative integers) without upper limit
- One parameter, λ.
- Variance equal to mean.
- For large λ close to normal.

	culture	population	contact	total_tools
1	Malekula	1100	low	13
2	Tikopia	1500	low	22
3	Santa Cruz	3600	low	24
4	Yap	4791	high	43
5	Lau Fiji	7400	high	33
6	Trobriand	8000	high	19
7	Chuuk	9200	high	40
8	Manus	13000	low	28
9	Tonga	17500	high	55
10	Hawaii	275000	low	71

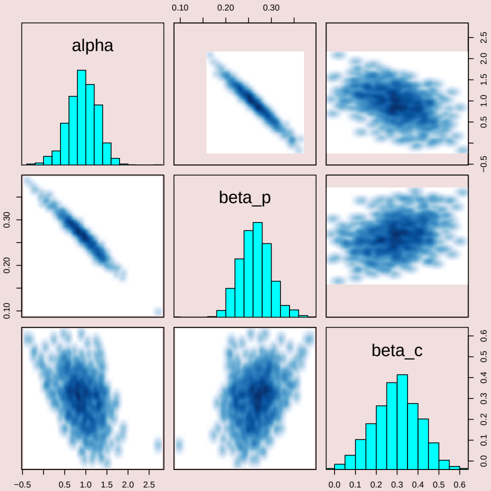
Poisson model

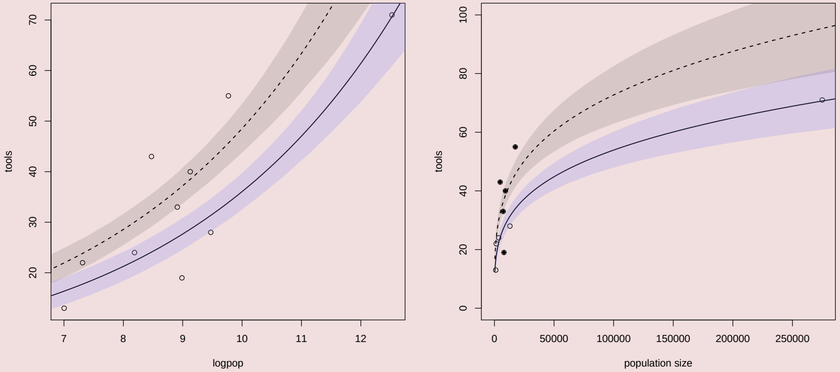
$$
\begin{aligned}
tools_i &\sim Po(\lambda_i) \\
g(\lambda_i) &= \alpha + \log(population_i)\beta_p + contact_i\beta_c \\
\alpha &\sim N(0, 10) \\
\beta_p &\sim N(0, 10) \\
\beta_c &\sim N(0, 10)
\end{aligned}
$$

samples of function

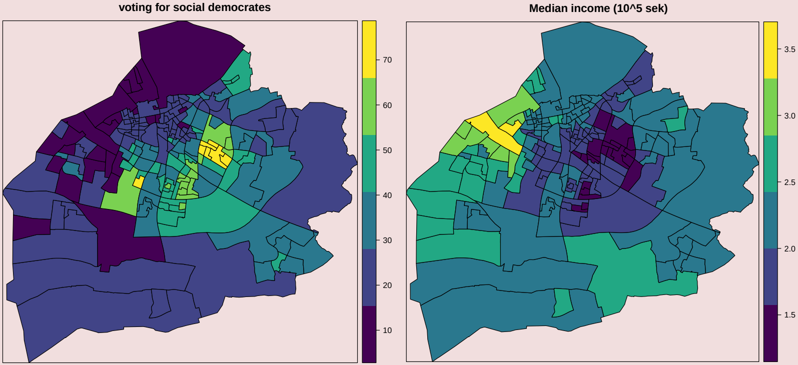


posterior fit

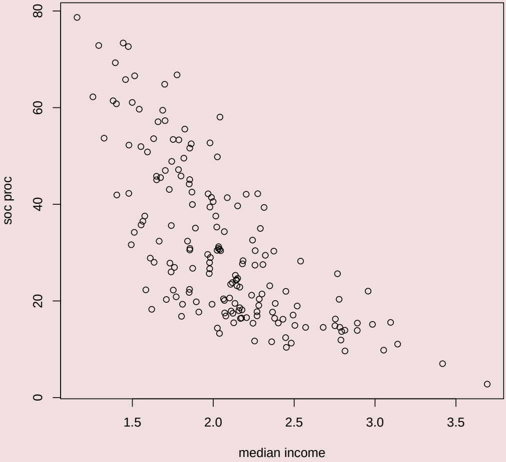



tools
70
60
50
40
30
20
7
8
9
10
11
12
logpop
tools
100
80
60
40
20
0
0
50000
100000
150000
200000
250000
population size

Voting in Malmö, data



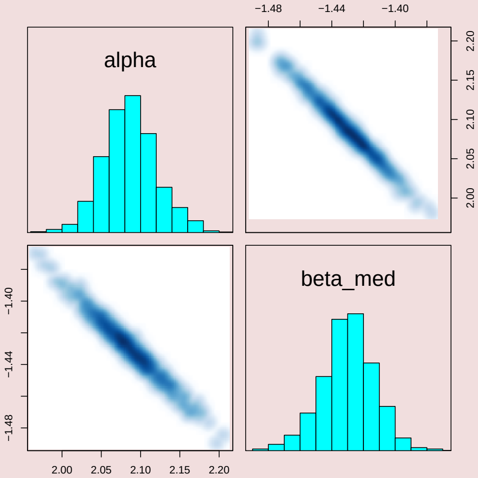
Voting in Malmö, data



Voting in Malmö, model

$$
\begin{aligned}
s_i &\sim bin(n_i, p_i), \\
g(p_i) &= \alpha + med_i \beta, \\
\alpha &\sim N(0, 10) \\
\beta &\sim N(0, 10)
\end{aligned}
$$

Posterior parameter

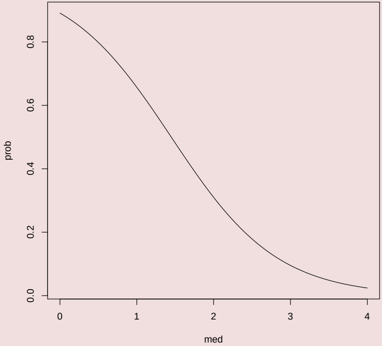


What would be the effect of an increase of income with $10K$ sek.

Name	Median income
Bellevue	3.7
Örtagården V	1.5

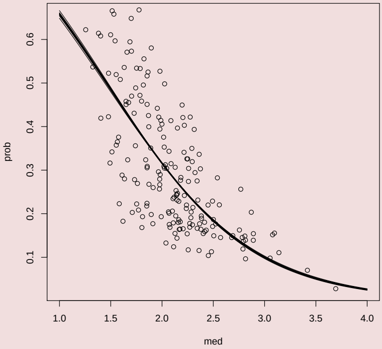
Two districts

What would be the effect of an increase of income with $10K$ sek.

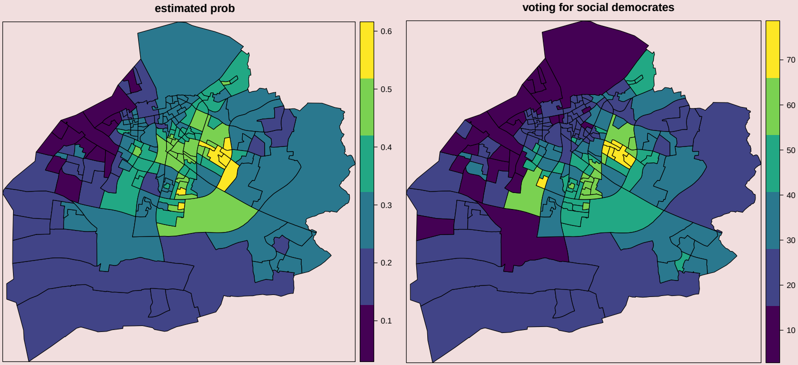


Name	Median income
Bellevue	3.7
Örtagården V	1.5

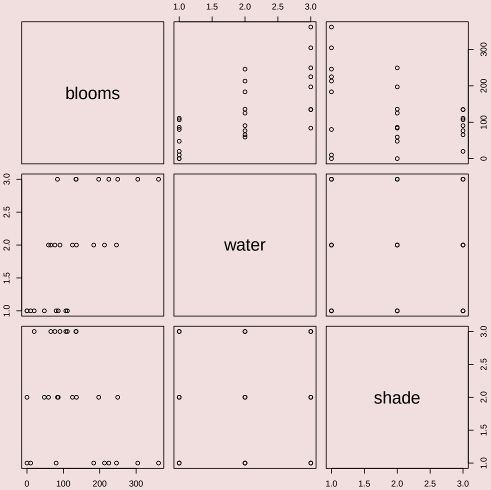
Examining posterior



Posterior p_i



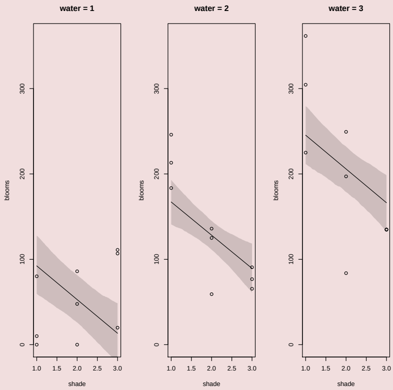
Tulips



Regular model

$$
\begin{aligned}
bloom_i &\sim N(\mu_i, \sigma), \\
\mu_i &= \alpha + water_i \beta_w + shade_i \beta_s
\end{aligned}
$$

Tulips



Tulips model

Regular model

$$
\begin{aligned}
bloom_i &\sim N(\mu_i, \sigma), \\
\mu_i &= \alpha + water_i \beta_w + shade_i \beta_s + water_i shade_i \beta_{ws}
\end{aligned}
$$

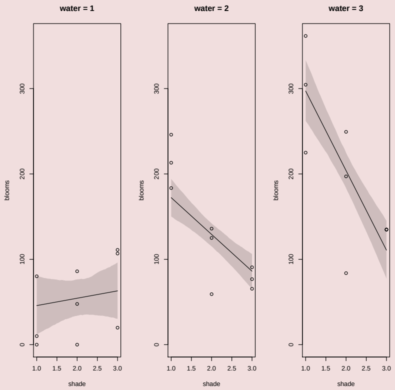
Model without interactions:

	Mean	StdDev	5.5%	94.5%
alpah	53.74	36.76	−5.01	112.49
beta_s	−39.03	12.99	−59.80	−18.26
beta_w	76.34	12.98	55.60	97.08
sigma	57.38	7.82	44.88	69.88

Model with interactions:

	Mean	StdDev	5.5%	94.5%
alpha	−139.13	60.10	−235.19	−43.07
beta_s	58.98	27.86	14.46	103.51
beta_w	176.32	27.86	131.79	220.85
beta_sw	−50.58	12.91	−71.22	−29.94
sigma	45.28	6.17	35.41	55.14

Tulips interacting



A common method to simplify interpretation is to standardize the covariates:

$$\tilde{X} = X - \bar{X}.$$

```
tulips$water <- tulips$water -mean(tulips$water)
tulips$shade <- tulips$shade -mean(tulips$shade)
```

tulips standardized

Model without interactions:

	Mean	StdDev	5.5%	94.5%
alpha	127.44	10.98	109.89	144.99
beta_s	−40.85	13.40	−62.27	−19.43
beta_w	74.43	13.41	53.00	95.85
sigma	57.37	7.82	44.88	69.87

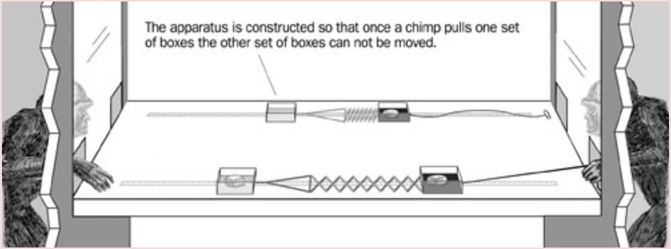
Model with interactions:

	Mean	StdDev	5.5%	94.5%
alpha	128.95	8.70	115.04	142.86
beta_s	−41.57	10.66	−58.60	−24.54
beta_w	75.74	10.66	58.71	92.77
beta_sw	−52.71	13.05	−73.56	−31.85
sigma	45.23	6.15	35.39	55.06

second example binomial example

Are Chimps altruistic?

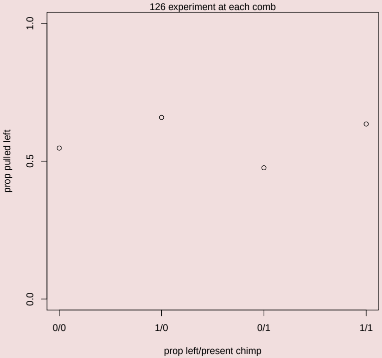
Chimpanzees data



Data , 504 observations, containing:

- Prop left: yes ($left = 1$), no ($left = 0$).
- Other Chimpanzee present: yes($C = 1$), no($C = 0$).
- Chimpanzee pulled left: yes($prop = 1$), no($prop = 0$)

Chimpanzees data



Data , 504 observations, containing:

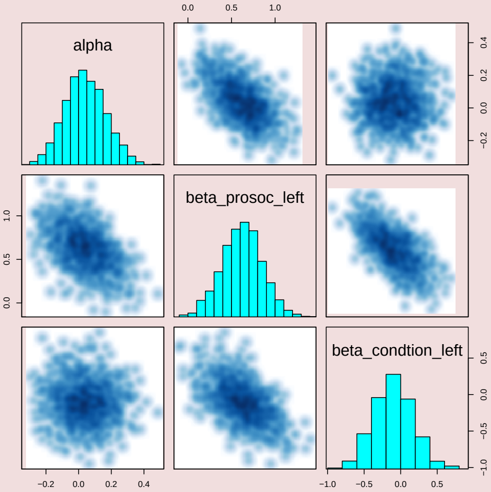
- Prop left: yes ($left = 1$), no ($left = 0$).
- Other Chimpanzee present: yes($C = 1$), no($C = 0$).
- Chimpanzee pulled left: yes($prop = 1$), no($prop = 0$)

Build a model checking if presence of other Chimpanzee matters for pulling left.

Build a model checking if presence of other Chimpanzee matters for pulling left.

$$
\begin{aligned}
left_i &\sim Bin(126, p_i) \\
g^{-1}(p_i) &\sim \alpha + prop_i(\beta_I + C_i \beta_C) \\
\alpha &\sim N(0, 10) \\
\beta_I &\sim N(0, 10) \\
\beta_C &\sim N(0, 10)
\end{aligned}
$$

Chimpanzees parameters



Vol 437|27 October 2005|**doi:10.1038/nature04243**

nature

LETTERS

Chimpanzees are indifferent to the welfare of unrelated group members

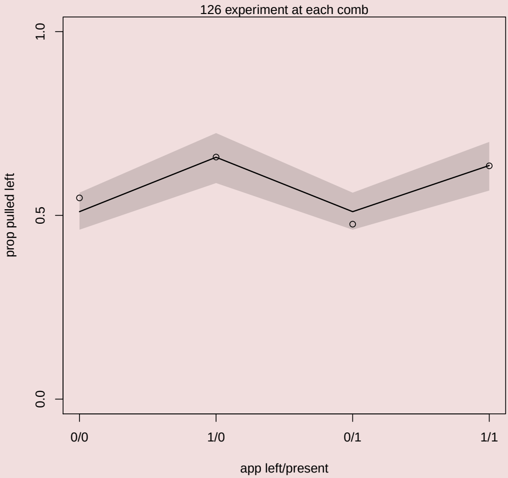
Joan B. Silk[1], Sarah F. Brosnan[2,3], Jennifer Vonk[4], Joseph Henrich[2], Daniel J. Povinelli[4], Amanda S. Richardson[3], Susan P. Lambeth[3], Jenny Mascaro[3] & Steven J. Schapiro[3]

Humans are an unusually prosocial species—we vote, give blood, recycle, give tithes and punish violators of social norms. Experimental evidence indicates that people willingly incur costs to help strangers in anonymous one-shot interactions[1,2], and that altruistic behaviour is motivated, at least in part, by empathy and concern for the welfare of others (hereafter referred to as other-regarding preferences)[1–3]. In contrast, cooperative behaviour in non-human primates is mainly limited to kin and reciprocating partners, and is virtually never extended to unfamiliar individuals[4]. Here we present experimental tests of the existence of other-regarding preferences in non-human primates, and show

This experimental setup maximizes the likelihood of observing other-regarding behaviour in two ways. First, actors can provide benefits to others at no cost to themselves, so other-regarding sentiments do not compete with selfish motives to obtain rewards. Second, actors interact with familiar group members. Prosocial responses in this experiment might occur because chimpanzees favour those that they cooperate with outside the context of this experiment, even if they lack other-regarding sentiments. However, the absence of prosocial behaviour in this experimental situation would provide strong evidence for the lack of other-regarding sentiments.

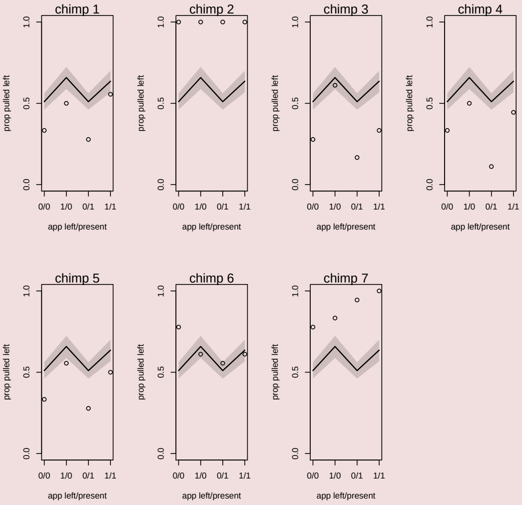
Chimpanzees fit

Does the model fit the data?



Chimpanzee individual fit

Does the model fit the data?



Chimpanzees model 2

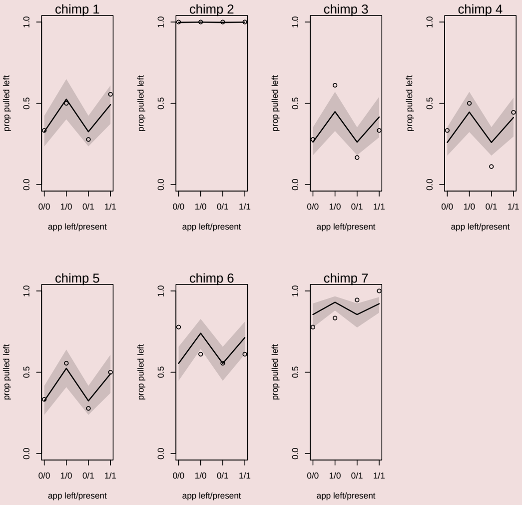
How do adjust our model for individual fit?

Chimpanzees model 2

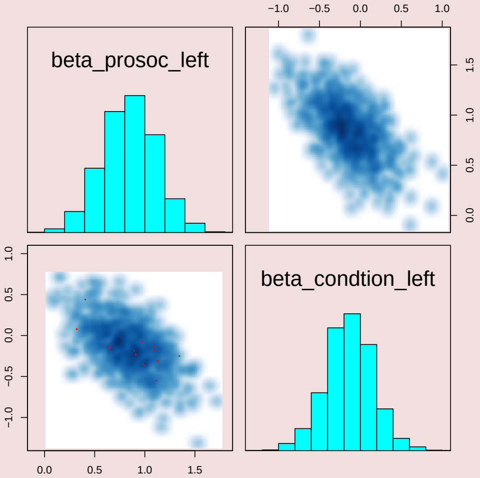
How do adjust our model for individual fit?

$$
\begin{aligned}
left_i &\sim Bin(18, p_i) \\
g^{-1}(p_i) &\sim \alpha_{j(i)} + prop_i(\beta_I + C_i\beta_C) \\
\alpha_j &\sim N(0, 10), j = 1, 2, \ldots, 7 \\
\beta_I &\sim N(0, 10) \\
\beta_C &\sim N(0, 10)
\end{aligned}
$$

Does the model fit the data?

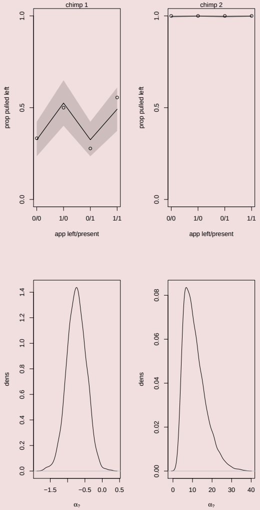


Chimpanzee parameters fit 2



Find the monkey

Which α_j belongs to which chimp, why the shape of the distribution?



Third example: admissions data

- A classical data studying the Berkeley admission data for gender bias.
- The university was sued for gender discrimination in the PhD application processes.

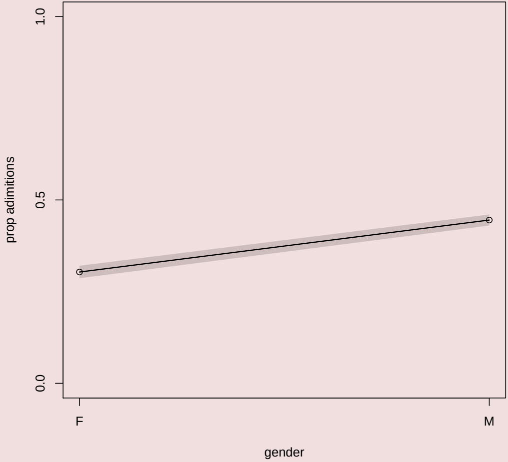
The pure data

	male	female
admitted	1198	557
rejected	1493	1278

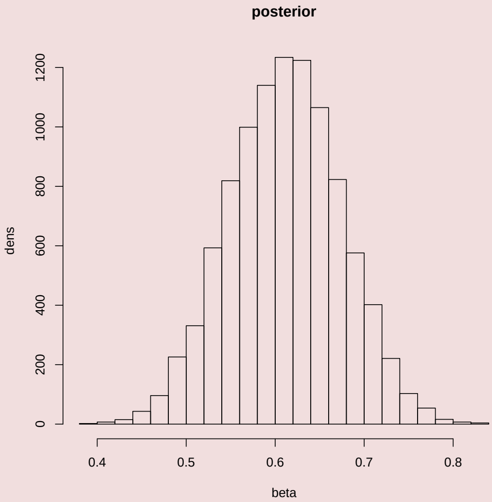
Simple model

$$
\begin{aligned}
adm_i &\sim Bin(n_i, p_i), \\
g(p_i) &= \alpha + male_i\beta, \\
\alpha &\sim N(0, 10) \\
\beta &\sim N(0, 10)
\end{aligned}
$$

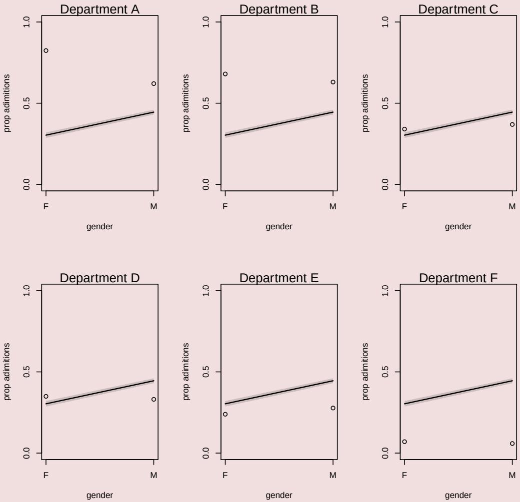
Does the model fit the data?



posterior difference between gender



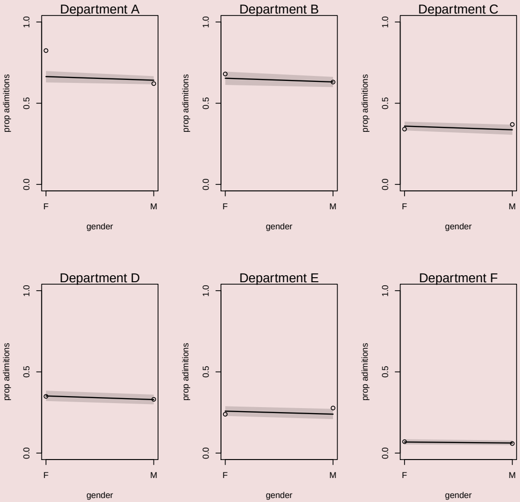
Does the model fit the data?



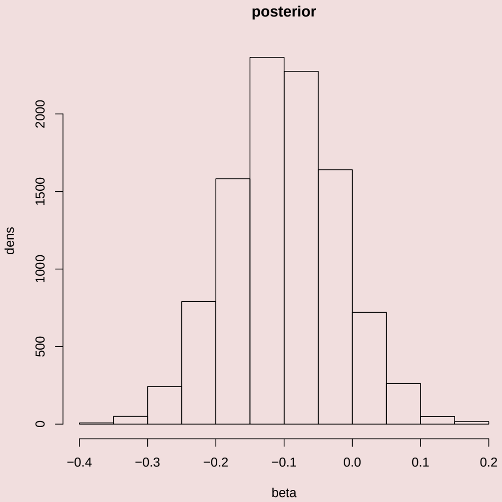
Adjusting for department

$$
\begin{aligned}
adm_i &\sim Bin(n_i, p_i),\\
g(p_i) &= \alpha_{i(j)} + male_i\beta,\\
\alpha_j &\sim N(0, 10)\\
\beta &\sim N(0, 10)
\end{aligned}
$$

Does the model fit the data?



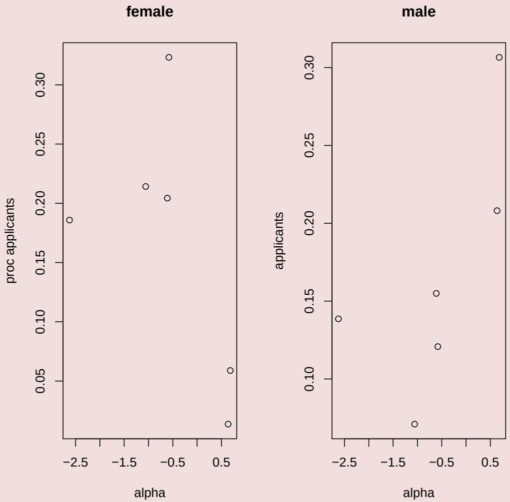
posterior difference between gender



What happened?

- Why is there such big difference between the two estimates?
- Why was there no little in the Chimpanzee data?
- Is the original wrong?

Who applies to what department



Article

Sex Bias in Graduate Admissions: Data from Berkeley

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

P. J. Bickel, E. A. Hammel, J. W. O'Connell

Determining whether discrimination because of sex or ethnic identity is being practiced against persons seeking passage from one social status or locus to another is an important problem in our society today. It is legally impor- deceision to admit or to deny admission. The question we wish to pursue is whether the decision to admit or to deny was influenced by the sex of the applicant. We cannot know with any certainty the influences on the evaluators in the

Figure: 1975, Science, 187 (4175), 398-404

- We will now use Rstan, to sample from the posterior distribution.

- We will now use Rstan, to sample from the posterior distribution.
- Named after Stanislaw Ulam, inventor of the Monte Carlo method (more next week).

Rstan

- We will now use Rstan, to sample from the posterior distribution.
- Named after Stanislaw Ulam, inventor of the Monte Carlo method (more next week).
- Will be tricky to install (if you use Windows or MAC) READ the INSTRUCTIONS

Rstan

- Why is hard to install? Requires a c++ compiler. This compiler needs to talk to R.
-
- I will put up Rmarkdown code before Friday on the homepage.
- Good idea to work through Rstudio, and files ending with .stan.

Height again again again again

$$h_i \sim N(\mu_i, \sigma),$$
$$\mu_i = \alpha + w_i\beta,$$
$$\alpha \sim N(0, 100)$$
$$\beta \sim N(0, 10)$$
$$\sigma \sim U[0, 50]$$

File model.stan contains:

```
data{
        int<lower=1> n;        // number of observations
        vector[n] w;           // weights
        real<lower=0> h[n];    // heights
}
parameters {
        real alpha;
        real beta;
        real<lower=0,upper=50> sigma;  // U[0,50]
}
model {
        vector[n] mu = alpha + w * beta;
        alpha  ~ normal(0, 100);
        beta   ~ normal(0,10);
        h      ~ normal(mu, sigma);
}
```

Simple Binomial

$$left_i \sim Bin(18, p_i)$$
$$g^{-1}(p_i) \sim \alpha + prop_i(\beta_l + C_i\beta_C)$$
$$\alpha \sim N(0, 10)$$
$$\beta_l \sim N(0, 10)$$
$$\beta_C \sim N(0, 10)$$

File model.stan contains:

```
data{
        int<lower=1> N; // number of observations
        int<lower = 0> pulled_left[N]; //number of le
        vector[N] prosoc_left; //prop on the left
        vector[N] C_prop;
}
parameters {
        real alpha;
        real beta_l;
        real beta_C;
}
model {
        belta_l ~  normal(0,10);
        belta_C ~  normal(0,10);
        alpha   ~  normal(0,10);
        pulled_left     ~ binomial_logit(18, alpha +
}
```

Setting up basic data

```
data("chimpanzees")

data.agg <- aggregate(x = list(pulled_left = chimpanzees$pulled_left),
                                              by=list(actor
= chimpanzees$actor,
 prosoc_left = chimpanzees$prosoc_left,
 condition   = chimpanzees$condition),
FUN = sum)

data.agg<-as.list(data.agg)
data.agg$N <- length(data.agg$actor)
data.agg$C_prop = data.agg$prosoc_left  *data.agg$condition
```

sampling posterior distribution using stan:

```
simple_fit = stan(file = "stan_example.stan",
                        data=as.list(data.agg),
                        iter = 2000,
                        warmup = 1000,
                        chains = 2)
```

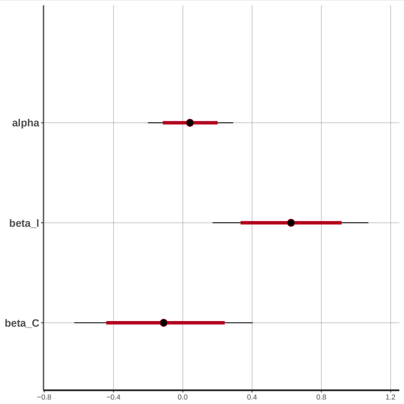
Summary

```
summary(simple_fit, pars=c("alpha","beta_l"))$summary
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%
alpha	0.05	0.00	0.13	−0.20	−0.03	0.05	0.14	0.30
beta_l	0.61	0.01	0.23	0.18	0.46	0.61	0.76	1.08

Plotting cred interval

```
plot(simple_fit)
```



Getting posterior samples

```
samples <-extract(simple_fit)
head(samples$beta_l)
```

```
[1] 0.5500067 0.4557822 0.4985765
[4] 0.1840083 0.8021953 0.5937610
```

$$left_i \sim Bin(18, p_i)$$
$$g^{-1}(p_i) \sim \alpha_{j(i)} + prop_i(\beta_I + C_i\beta_C)$$
$$\alpha_j \sim N(0, 10), j = 1, .., 7$$
$$\beta_I \sim N(0, 10)$$
$$\beta_C \sim N(0, 10)$$

File model.stan contains:

```
data {
        int<lower = 1> N;
        int<lower = 1> Nactor;
        int<lower = 0> pulled_left[N];
        int<lower = 1> actor[N]; // which chimp
        vector[N] prosoc_left;
        vector[N] C_prop;
}

parameters{
        vector[Nactor] alpha;
        real beta_prosoc_left;
        real beta_condtion_left;
}
model{
        real alphas[N];
        alpha                ~ normal(0, 10);
        beta_prosoc_left     ~ normal(0, 10);
        beta_condtion_left ~ normal(0, 10);
        for(i in 1:N)
        alphas[i] = alpha[actor[i]] + prosoc_left[i]
                    + cond_left[i]* beta_condtion_le

        pulled_left ~ binomial_logit(18, alphas );
}
```