5.2 interpreting logistic regresssion

Chris Parrish

July 3, 2016

Contents

5.2 interpreting logistic regresssion

references:
- ARM chapter 05, github
- arsenic, Wikipedia
- arsenic, World Health Organization
- arsenic, American Cancer Society

```
library(rstan)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
library(ggplot2)
```

nes

National Election Study, 1992

data

```
### Data
# nesYYYY_vote.data.R, where YYYY == seq(1952, 2000, 4)
```

model

nes_logit.stan

```
data {
  int<lower=0> N;
  vector[N] income;
  int<lower=0,upper=1> vote[N];
}
parameters {
  vector[2] beta;
}
model {
  vote ~ bernoulli_logit(beta[1] + beta[2] * income);
}
```

fit

```
### Logistic model: vote ~ income
## Sampling over 13 years & building a data frame for Figure 5.4
nes_vote.ggdf <- data.frame(c(), c(), c(), c()) # empty data frame
years <- seq(1952, 2000, 4)
for (i in years) {
    source(paste("nes", i, "_vote.data.R", sep = ""))
    data.list <- c("N", "vote", "income")
    sf <- stan(file='nes_logit.stan', data=data.list, iter=1000, chains=4)
    beta.post <- extract(sf, "beta")$beta
    beta.mean <- colMeans(beta.post)
    beta.sd   <- apply(beta.post, 2, sd)
    nes_vote.ggdf <- rbind(nes_vote.ggdf, data.frame(Year = i,
                                          coef = beta.mean[2],
                                          lower = beta.mean[2] - beta.sd[2],
                                          upper = beta.mean[2] + beta.sd[2]))
}
```

```
### Figure 5.4
p <- ggplot(nes_vote.ggdf) +
    geom_pointrange(aes(x = Year, y = coef, ymin = lower, ymax = upper),
                    color = "steelblue") +
    geom_hline(yintercept = 0, linetype = "dashed", size = 0.5) +
    scale_y_continuous("Coefficient of income")
print(p)
```