SR Ch 11 Hw

Emilio Horner

2023-01-27

1e.

.35

```
.35/(1-.35)
```

```
## [1] 0.5384615
```

```
log(.538)
```

```
## [1] -0.6198967
```

-.619

2e.

.32 is the log odds

```
(exp(.32))/(1+exp(.32))
```

```
## [1] 0.5793243
```

58% probability

3e.

One can get proportional odds by exponentiating the coefficient

```
exp(1.7)
```

```
## [1] 5.473947
```

This value is 5.5

4e.

The offset in Poisson distribution is a way of standardizing the units to account for difference.The book gives the example of time. The offset makes the rate of events consistent in different times. An example could be a admissions rate in which some data was in people admitted per year vs admitted per month.

1m.

Aggregated data includes a multiplicative term because there are different ways one could get to the number of counts in the disaggregated form. The multiplicative impacts the likelihood but not the inferences because it doesn't impact the actual posterior distribution.

2m.

I think this is the same as the question 1e. We could exponentiate it to 5.5.

3m.

We need to use the logit link not the identity link to standarize the x axis and get probability on the y axis. Things in the real world aren't linear in the sense of y = mx + b , and are more complicated. Essentially the point is to convert covariate values onto a scale of probability.

4m.

The Poisson distribution is used to see how the model measured up to the data.It is used when the model only predicts something in the positive.

5m.

It would restrict the value between 0 and 1. The Poisson distribution should be used where discrete events are occuring in a continuous manner.

6m.

The binomial and Poisson would be the same when the trials are large in number. the constraints would be binary outcomes. this means that the outcomes could only have two options.

7m.

```
library(rethinking)
```

```
## Loading required package: rstan
```

```
## Loading required package: StanHeaders
```

```
## Loading required package: ggplot2
```

```
## rstan (Version 2.21.7, GitRev: 2e1f913d3ca3)
```

```
## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
```

```
## Do not specify '-march=native' in 'LOCAL_CPPFLAGS' or a Makevars file
```

```
## Loading required package: parallel
```

```
## Loading required package: dagitty
```

```
## rethinking (Version 2.01)
```

```
##
## Attaching package: 'rethinking'
```

```
## The following object is masked from 'package:stats':
##
##     rstudent
```

```
data("chimpanzees")
d <- chimpanzees
```

```
QuapChimp <- quap(
    alist(
        pulled_left ~ dbinom( 1 , p ) ,
        logit(p) <- a ,
        a ~ dnorm( 0 , 1 )
) , data=d)
```

```
precis (QuapChimp)
```

```
##        mean          sd      5.5%      94.5%
## a 0.3175824 0.08984774 0.1739884 0.4611764
```

I got this from pg 330 in the book, I'm not sure about the questions he's asking though.

8m.

```
library(dplyr )
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

library(rethinking) data(Kline) d <- Kline d %>% filter(!row_number(10)) d$P <- standardize(log(d$population)) d$contact_id d <- ifelse(d$contact=="high" , 2L , 1L) dat2 <- list(T = d$total_tools, P = d$population, cid = d$contact_id) m12.2 <- ulam(alist(T ~ dgampois(lambda , phi), lambda <- exp(a[cid])*P^b[cid] / g, a[cid] ~ dnorm(1,1), b[cid] ~ dexp(1), g ~ dexp(1), phi ~ dexp(1)), data=dat2 , chains=4 , log_lik=TRUE)

Hawaii is the outlier, so it makes sense to drop it.