milk3

Chris Parrish

June 20, 2016

Contents

milk3

reference:
- McElreath, Statistical Rethinking, chap 5, pp.119-164

milk

```
library(rethinking)
library(ggplot2)
library(cowplot)
theme_set(theme_gray())
```

data

```
## R code 5.48
data(milk)
d <- milk
str(d)
```

```
## 'data.frame':    29 obs. of  8 variables:
##  $ clade         : Factor w/ 4 levels "Ape","New World Monkey",..: 4 4 4 4 4 2 2 2 2 2 ...
##  $ species       : Factor w/ 29 levels "A palliata","Alouatta seniculus",..: 11 8 9 10 16 2 1 6 28 2
##  $ kcal.per.g    : num  0.49 0.51 0.46 0.48 0.6 0.47 0.56 0.89 0.91 0.92 ...
##  $ perc.fat      : num  16.6 19.3 14.1 14.9 27.3 ...
##  $ perc.protein  : num  15.4 16.9 16.9 13.2 19.5 ...
##  $ perc.lactose  : num  68 63.8 69 71.9 53.2 ...
##  $ mass          : num  1.95 2.09 2.51 1.62 2.19 5.25 5.37 2.51 0.71 0.68 ...
##  $ neocortex.perc: num  55.2 NA NA NA NA ...
```
```
unique(d$clade)
```

```
## [1] Strepsirrhine    New World Monkey Old World Monkey Ape
## Levels: Ape New World Monkey Old World Monkey Strepsirrhine
```

exploratory data analysis

```
ggplot(d, aes(x = clade, y = kcal.per.g, fill = clade)) +
  geom_boxplot() +
  scale_fill_manual(values = c("wheat", "paleturquoise", "thistle", "tomato"))
```
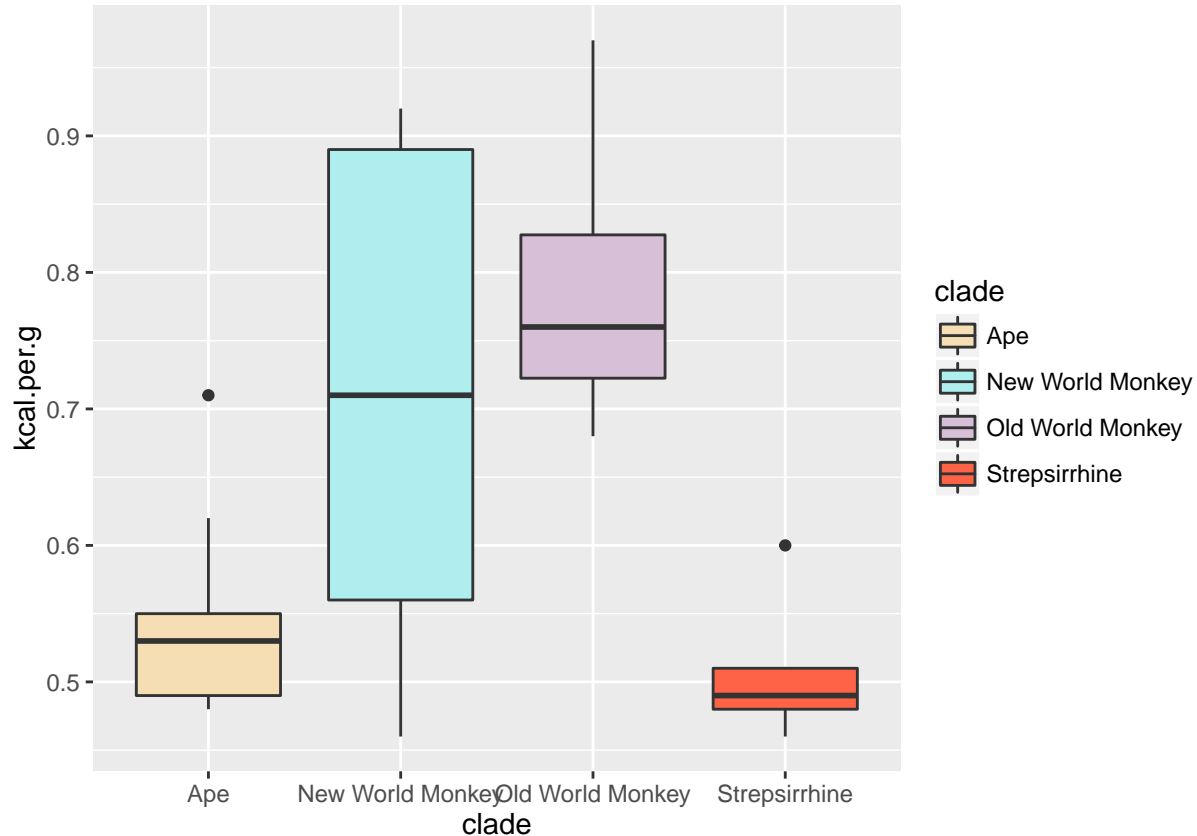

Indicator variables

```
## R code 5.49
( d$clade.NWM <- ifelse( d$clade=="New World Monkey" , 1 , 0 ) )
```

```
##   [1] 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
## R code 5.50
d$clade.OWM <- ifelse( d$clade=="Old World Monkey" , 1 , 0 )
d$clade.S <- ifelse( d$clade=="Strepsirrhine" , 1 , 0 )
```

map

```
## R code 5.51
m5.16 <- map(
    alist(
        kcal.per.g ~ dnorm( mu , sigma ) ,
        mu <- a + b.NWM*clade.NWM + b.OWM*clade.OWM + b.S*clade.S ,
        a ~ dnorm( 0.6 , 10 ) ,
        b.NWM ~ dnorm( 0 , 1 ) ,
        b.OWM ~ dnorm( 0 , 1 ) ,
```

```
        b.S ~ dnorm( 0 , 1 ) ,
        sigma ~ dunif( 0 , 10 )
    ) ,
    data=d )
precis(m5.16)
```

```
##        Mean StdDev  5.5% 94.5%
## a      0.55   0.04  0.49  0.61
## b.NWM  0.17   0.05  0.08  0.25
## b.OWM  0.24   0.06  0.15  0.34
## b.S   -0.04   0.06 -0.14  0.06
## sigma  0.11   0.02  0.09  0.14
```

analysis

```
## R code 5.52
# sample posterior
post <- extract.samples(m5.16)

# compute averages for each category
mu.ape <- post$a
mu.NWM <- post$a + post$b.NWM
mu.OWM <- post$a + post$b.OWM
mu.S <- post$a + post$b.S

# summarize using precis
precis( data.frame(mu.ape,mu.NWM,mu.OWM,mu.S) )
```

```
##         Mean StdDev |0.89 0.89|
## mu.ape 0.55   0.04  0.49  0.61
## mu.NWM 0.71   0.04  0.66  0.78
## mu.OWM 0.79   0.05  0.72  0.86
## mu.S   0.51   0.05  0.43  0.59
## R code 5.53
diff.NWM.OWM <- mu.NWM - mu.OWM
quantile( diff.NWM.OWM , probs=c(0.025,0.5,0.975) )
```

```
##        2.5%        50%       97.5%
## -0.19340951 -0.07352363  0.04585199
```

unique intercepts

```
## R code 5.54
( d$clade_id <- coerce_index(d$clade) )
```

```
##  [1] 4 4 4 4 4 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1
```

```
str(d)
```

```
## 'data.frame':    29 obs. of  12 variables:
##  $ clade        : Factor w/ 4 levels "Ape","New World Monkey",..: 4 4 4 4 4 2 2 2 2 2 ...
##  $ species      : Factor w/ 29 levels "A palliata","Alouatta seniculus",..: 11 8 9 10 16 2 1 6 28 27
```

```
##  $ kcal.per.g    : num  0.49 0.51 0.46 0.48 0.6 0.47 0.56 0.89 0.91 0.92 ...
##  $ perc.fat      : num  16.6 19.3 14.1 14.9 27.3 ...
##  $ perc.protein  : num  15.4 16.9 16.9 13.2 19.5 ...
##  $ perc.lactose  : num  68 63.8 69 71.9 53.2 ...
##  $ mass          : num  1.95 2.09 2.51 1.62 2.19 5.25 5.37 2.51 0.71 0.68 ...
##  $ neocortex.perc: num  55.2 NA NA NA NA ...
##  $ clade.NWM     : num  0 0 0 0 0 1 1 1 1 1 ...
##  $ clade.OWM     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ clade.S       : num  1 1 1 1 1 0 0 0 0 0 ...
##  $ clade_id      : int  4 4 4 4 4 2 2 2 2 2 ...
```

```r
## R code 5.55
m5.16_alt <- map(
    alist(
        kcal.per.g ~ dnorm( mu , sigma ) ,
        mu <- a[clade_id] ,
        a[clade_id] ~ dnorm( 0.6 , 10 ) ,
        sigma ~ dunif( 0 , 10 )
    ) ,
    data=d )
precis( m5.16_alt , depth=2 )
```

```
##        Mean StdDev 5.5% 94.5%
## a[1]   0.55   0.04 0.48  0.61
## a[2]   0.71   0.04 0.65  0.78
## a[3]   0.79   0.05 0.71  0.86
## a[4]   0.51   0.05 0.43  0.59
## sigma 0.11   0.02 0.09  0.14
```

parameter distributions

```r
# extract samples
post <- extract.samples( m5.16_alt )
str(post)
```

```
## List of 2
##  $ sigma: num [1:10000] 0.1311 0.1193 0.0961 0.1303 0.1174 ...
##  $ a    : num [1:10000, 1:4] 0.561 0.532 0.592 0.494 0.53 ...
```

```r
# plot each parameter distribution
a1.plot <- parameter.dist(parameter = "a1", values = post$a[ , 1])
a2.plot <- parameter.dist(parameter = "a2", values = post$a[ , 2])
a3.plot <- parameter.dist(parameter = "a3", values = post$a[ , 3])
a4.plot <- parameter.dist(parameter = "a4", values = post$a[ , 4])
# display parameter distributions
plot_grid(a1.plot, a2.plot, a3.plot, a4.plot,
          labels=c("a[1]", "a[2]", "a[3]", "a[4]"), ncol = 2, nrow = 2)
```

a[1]

density

a[2]

density

0.482 0.545 0.604
HPDI   mean   HPDI

0.654 0.714 0.776
HPDI   mean   HPDI

a[3]

density

a[4]

density

0.712 0.788 0.859
HPDI   mean   HPDI

0.427 0.509 0.591
HPDI   mean   HPDI

5