Geog 509: Bayesian Data Analysis
Chapter 5 Problem Set

Jesse Piburn

2020-02-28

5E1 - 5E4

5E1

Models 2,3, and 4 are multiple regressions

5E2

$l_i = latitude$, $A_i = Animal Diversity$, and $P_i = Plant Diversity$

$$l_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_A A_i + \beta_P P_i$$
$$\alpha \sim \text{Normal}(0, 50)$$
$$\beta_A \sim \text{Normal}(0, 20)$$
$$\beta_P \sim \text{Normal}(0, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

5E3

$t_i = $ Time to PhD, $F_i = $ amount of funding, and $S_i = $ laboratory size

$$t_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_F F_i + \beta_S S_i$$
$$\alpha \sim \text{Normal}(0, 50)$$
$$\beta_F \sim \text{Normal}(0, 20)$$
$$\beta_S \sim \text{Normal}(0, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

Assuming "positively associated" in the context of the question means recieving the degree faster is a positive outcome, that would mean as funding and lab size increase, time to degree would decrease. This would result in both funding and lab size having negative slopes.

5E4

Models 1, 3, 4 and 5 are inferentially equivalent

5M3

High divorce rate can cause a higher marriage rate in the sense that a marriage rate counts the number of marriages per some fixed amount of people. If people are getting divorced at a higher rate, some of those people are likely to get married again, therefore inreasing the number of marriages while the count of people remains the same. One potential way to evaluate this relationship is the regress marriage rate on both divorce rate and some indicator of remarriage rate.

5M5

In this model we could include indicators for each of the causes in the hypotheses. For example, higher gas prices could lead to less driving and therefore more exercise, means that we would want to include measures for gas price, driving (e.g. miles driven per week), and exercise (e.g. time exercising per week). Including the claims in the second hypothesis we would want to add a measure of how often people eat out at restaurants (e.g. times per week) and also a measure of how much they eat at the restaurant (e.g. calories per meal. . . assuming you could actually measure that).

5H1-5H3

Things we'll need for all the questions

```
library(brms)
library(tidyverse)
library(tidybayes)
library(bayesplot)
library(patchwork)


data(foxes, package = "rethinking")
d <- foxes


rm(foxes)
```

note: All of code for for generating the charts was omitted from this report as it was distracting and took up too much space. If interested, please see the source Rmd here

5H1

weight ~ area

```
# weight ~ area
h1.area <-
  brm(data = d, family = gaussian,
      weight ~ 1 + area,
      prior = c(
        prior(normal(0, 10), class = Intercept),
        prior(normal(0, 5), class = b),
        prior(cauchy(0, 1), class = sigma)
        ),
      iter = 2000, warmup = 500, chains = 4, cores = 4,
      seed = 1988)

h1.area
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: weight ~ 1 + area
##    Data: d (Number of observations: 116)
## Samples: 4 chains, each with iter = 2000; warmup = 500; thin = 1;
##          total post-warmup samples = 6000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     4.45      0.40     3.67     5.22 1.00     6328     4603
## area          0.03      0.12    -0.21     0.26 1.00     6004     4046
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     1.20      0.08     1.05     1.36 1.00     5153     3717
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

weight ~ groupsize
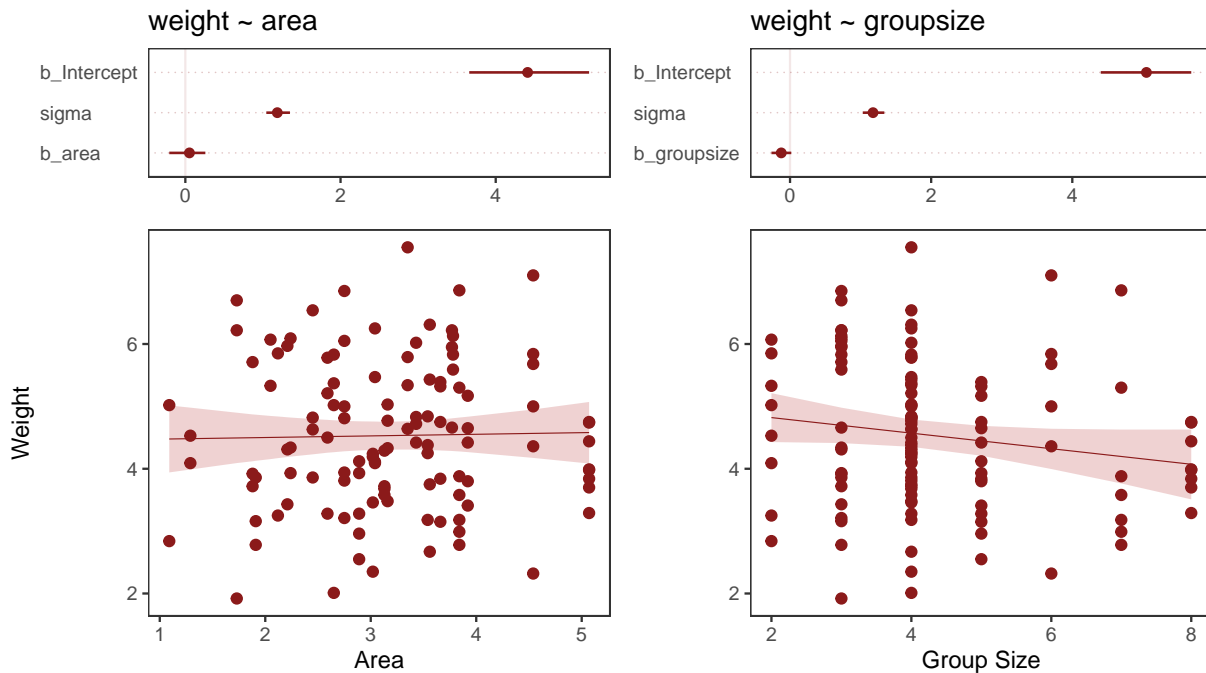
```r
# weight ~ groupsize
h1.groupsize <-
  brm(data = d, family = gaussian,
      weight ~ 1 + groupsize,
      prior = c(
        prior(normal(0, 10), class = Intercept),
        prior(normal(0, 5), class = b),
        prior(cauchy(0, 1), class = sigma)
        ),
      iter = 2000, warmup = 500, chains = 4, cores = 4,
      seed = 1988)

h1.groupsize
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: weight ~ 1 + groupsize
##    Data: d (Number of observations: 116)
## Samples: 4 chains, each with iter = 2000; warmup = 500; thin = 1;
##          total post-warmup samples = 6000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     5.07      0.33     4.42     5.71 1.00     5277     4288
## groupsize    -0.12      0.07    -0.26     0.02 1.00     5233     3984
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     1.18      0.08     1.04     1.35 1.00     5015     3993
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
```

```
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

When considered individually, groupsize has a small negative relationship with weight
while area has little to none

5H2

weight ~ area + groupsize

```r
# weight ~ area + groupsize
h2 <-
  brm(data = d, family = gaussian,
      weight ~ 1 + area + groupsize,
      prior = c(
        prior(normal(0, 10), class = Intercept),
        prior(normal(0, 5), class = b),
        prior(cauchy(0, 1), class = sigma)
        ),
      iter = 2000, warmup = 500, chains = 4, cores = 4,
      seed = 1988)
h2
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: weight ~ 1 + area + groupsize
##    Data: d (Number of observations: 116)
## Samples: 4 chains, each with iter = 2000; warmup = 500; thin = 1;
##          total post-warmup samples = 6000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     4.45      0.37     3.72     5.18 1.00     6627     4884
```

```
## area             0.61        0.20        0.23      1.01 1.00      3225      3955
## groupsize       -0.43        0.12       -0.67     -0.19 1.00      3183      3874
##
## Family Specific Parameters:
##        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      1.14      0.08     1.00     1.30 1.00     5411     3948
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

When considered together however, the effect of both are intensified

Area now has a very strong positive effect when accounting for groupsize
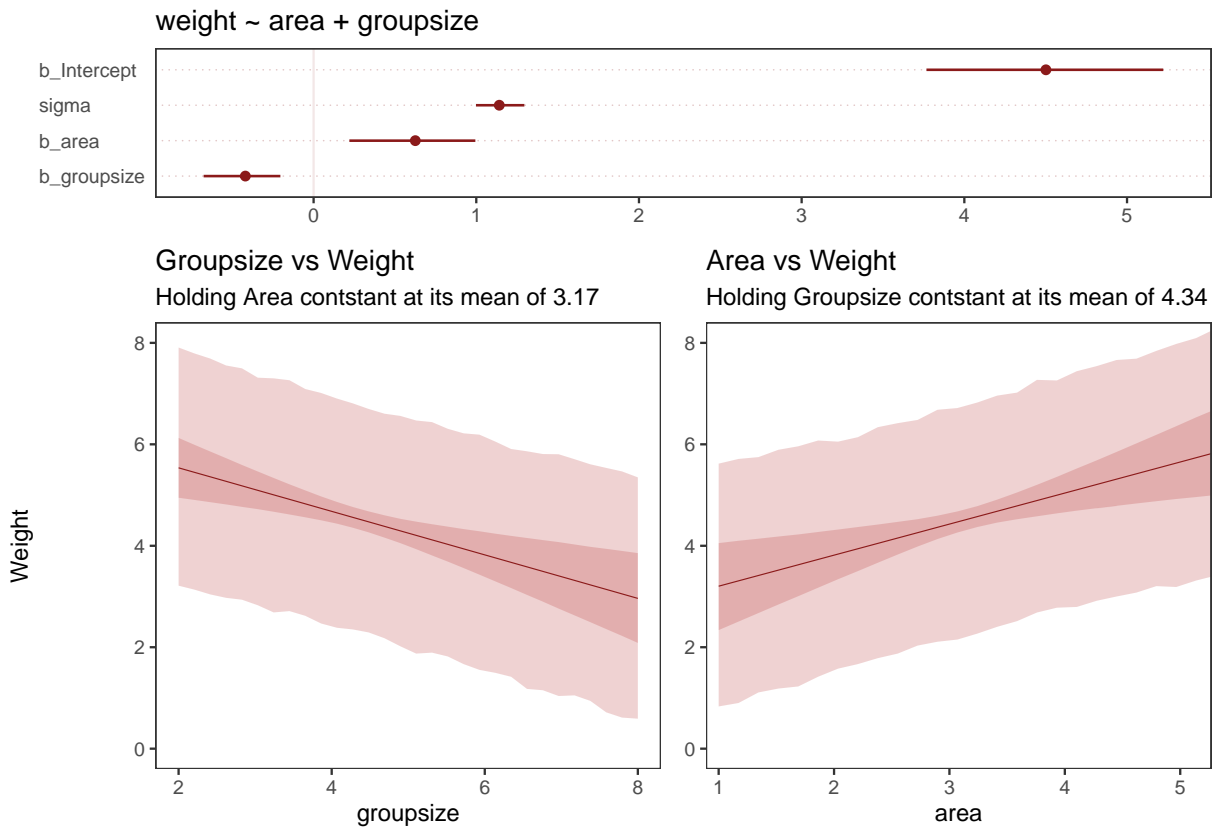
The results here are different than when considering them invididually because we are asking the model a different question and when we ask a different question we shouldn't be surprised when we get different answers. When considering only `weight ~ area`, for example. We are asking, what is the predictive power of area for predicting weight? However, when we include groupsize as well, the question becomes what is the predictive power of area, once I already know groupsize.

5H3

weight ~ avgfood + groupsize

```r
# weight ~ avgfood + groupsize
h3.ag <-
  brm(data = d, family = gaussian,
```

```
    weight ~ 1 + avgfood + groupsize,
    prior = c(
      prior(normal(0, 10), class = Intercept),
      prior(normal(0, 5), class = b),
      prior(cauchy(0, 1), class = sigma)
      ),
    iter = 2000, warmup = 500, chains = 4, cores = 4,
    seed = 1988)
```

```
h3.ag
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: weight ~ 1 + avgfood + groupsize
##    Data: d (Number of observations: 116)
## Samples: 4 chains, each with iter = 2000; warmup = 500; thin = 1;
##          total post-warmup samples = 6000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     4.17      0.43     3.31     5.01 1.00     5360     4879
## avgfood       3.62      1.16     1.26     5.83 1.00     3941     3072
## groupsize    -0.54      0.15    -0.83    -0.23 1.00     4050     3262
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     1.14      0.08     1.00     1.30 1.00     4621     3803
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

weight ~ avgfood + groupsize + area

```
# weight ~ avgfood + groupsize + area
h3.aga <-
  brm(data = d, family = gaussian,
      weight ~ 1 + avgfood + groupsize + area,
      prior = c(
        prior(normal(0, 10), class = Intercept),
        prior(normal(0, 5), class = b),
        prior(cauchy(0, 1), class = sigma)
        ),
      iter = 2000, warmup = 500, chains = 4, cores = 4,
      seed = 1988)
```

```
h3.aga
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: weight ~ 1 + avgfood + groupsize + area
##    Data: d (Number of observations: 116)
## Samples: 4 chains, each with iter = 2000; warmup = 500; thin = 1;
##          total post-warmup samples = 6000
```
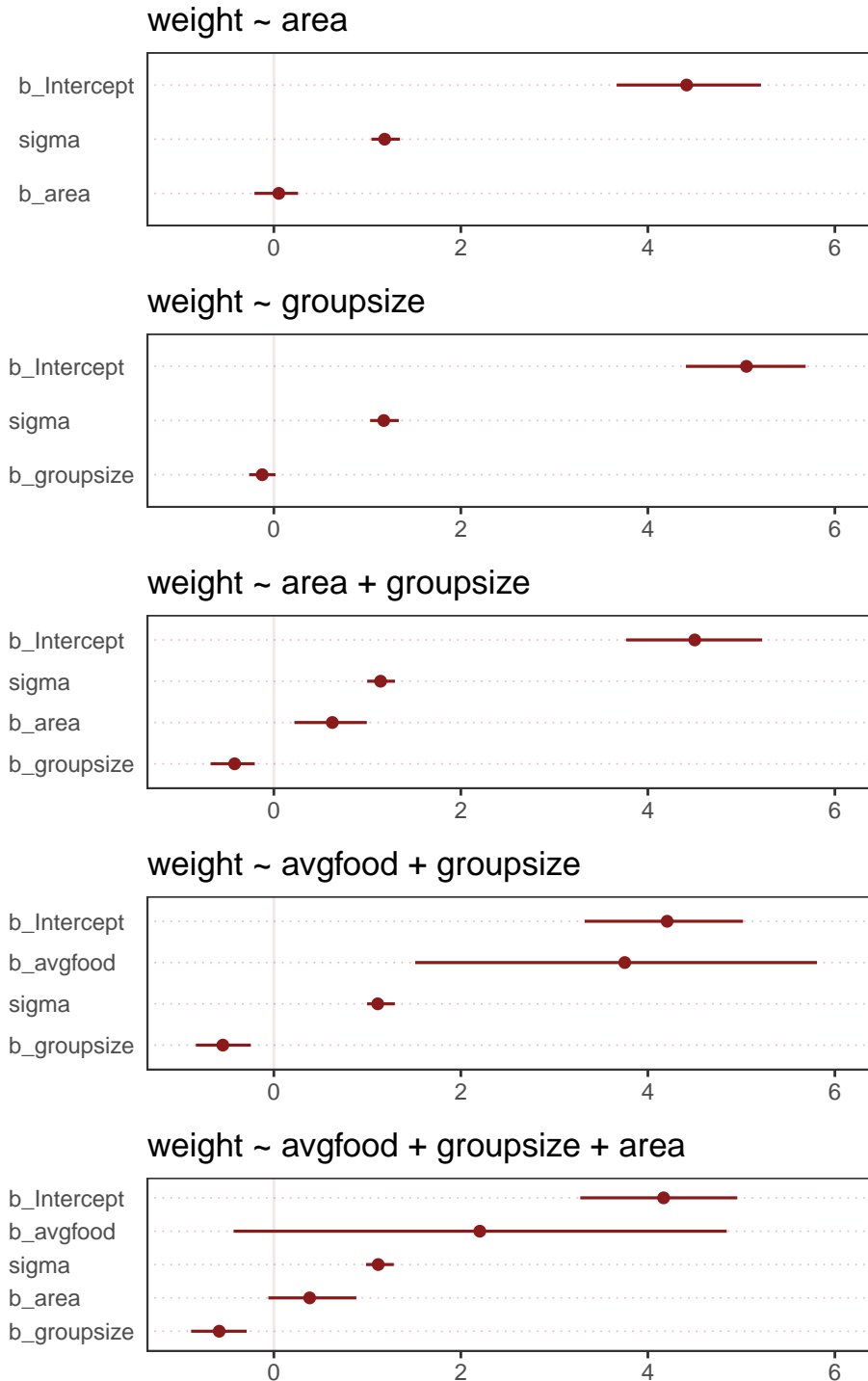
```
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     4.11      0.43     3.27     4.95 1.00     5730     5071
## avgfood       2.25      1.41    -0.44     5.07 1.00     4048     3770
## groupsize    -0.59      0.16    -0.91    -0.29 1.00     4546     3647
## area          0.41      0.24    -0.06     0.88 1.00     4977     4036
##
## Family Specific Parameters:
##        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      1.13      0.08     0.99     1.30 1.00     5934     3225
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Comparison of weight prediction models

weight ~ area

weight ~ groupsize

weight ~ area + groupsize

weight ~ avgfood + groupsize

weight ~ avgfood + groupsize + area

(a): `area` is a better predictor of `weight` than `avgfood` for two reasons. 1) The amount of food available is a consequence of the area and 2) avgfood has very large credible intervals even when area is not included in the model because it has a very high correlation with both groupsize and area as seen below
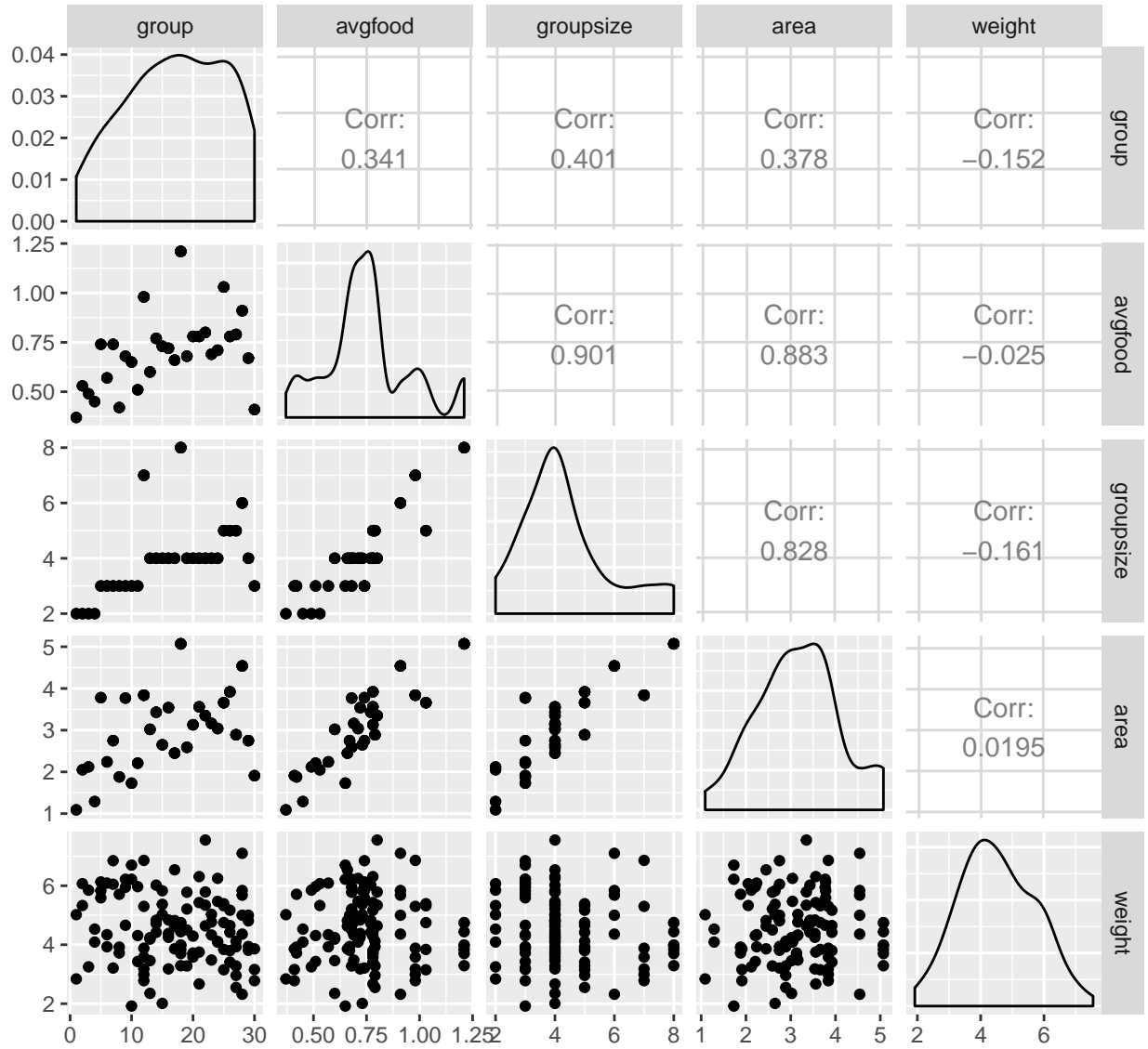
(b): This occurs because of weak identifiability in the model caused by multicollinearity between area and avgfood