

Problem Set 2

Michael Fryer

2025-11-15

1 Multiple Regression & Causal Models

The `foxes` dataset contains data on urban fox populations.

```
# First, load the foxes dataset
data(foxes)
d <- foxes
# You must set random seed to 390
set.seed(390)
```

Consider the following hypothesized causal relationship between **territory size** and **body weight** in foxes.

Daggy DAG

where A , F and W represent random variables `area` (territory size), `avgfood`, and `weight`, respectively.

If this DAG correctly describes the causal relationships, it makes specific predictions about what we should observe in the data. Your task is to test whether the observed patterns match these predictions.

- Territory size (A) has a **direct** effect on weight (W) : $A \rightarrow W$
- Food availability (F) has a **direct** effect on weight (W) : $F \rightarrow W$
- Territory size (A) has an **indirect** effect on weight (W) through food (F) : $A \rightarrow F \rightarrow W$

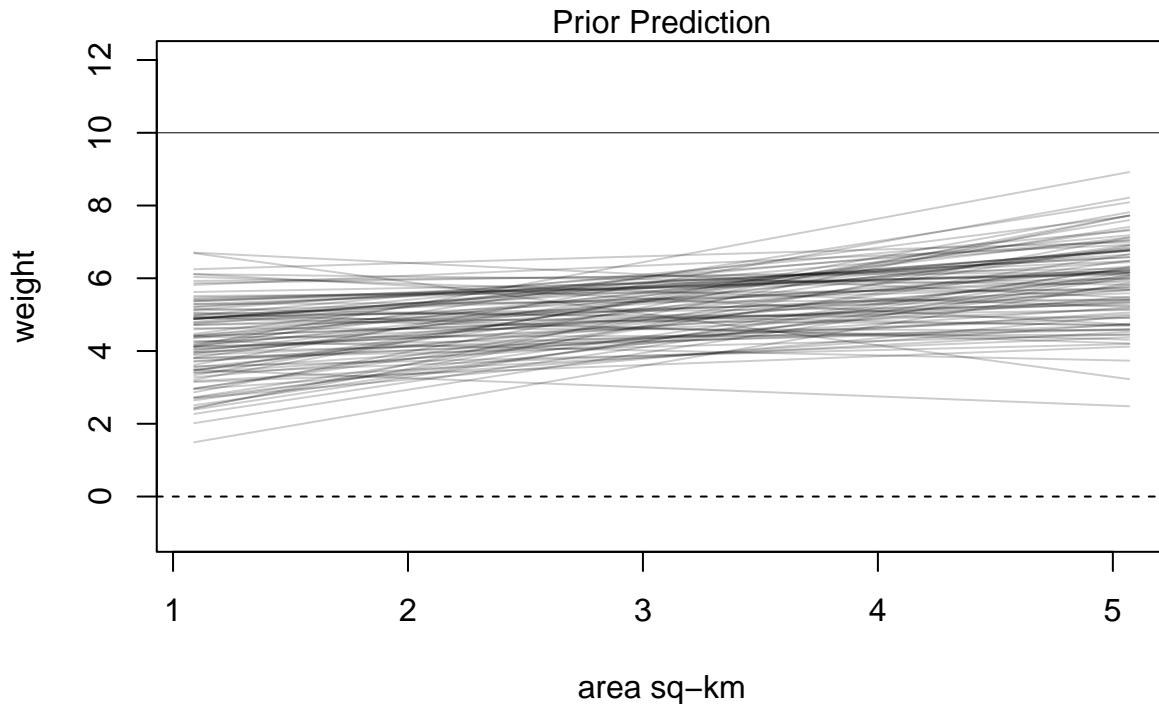
a) According to the DAG, territory size effects weight through two paths:

- Direct path: $A \rightarrow W$
- Indirect path: $A \rightarrow F \rightarrow W$

If we regress weight on territory size without including food, the coefficient should capture both pathways, the “total association” between A and W . Construct a linear regression (`m1a`) using `quap`. Urban foxes in this population have an average weight of 5kg. Use prior predictive simulation to assess the implications of your priors. Standardize the prediction variable.

1.1 Prior Predictive Simulation

```
N <- 100
# a is the intercept. This means a fox with territory size 0 has this weight.
# Let's assume that value is within 15% of the average weight of a fox
a <- rnorm(N, 5, 0.75)
# b represents the rate of change of a fox's weight with respect to territory
# size. Let's assume the mean for this value is 0.4 which represents an
# addition of 2 kg for every 5 sq-km
b <- rnorm(N, 0.4, 0.4)
xbar <- mean(d$area)
```



1.2 Linear Regression

```
m <- quap(
  alist(
    weight ~ dnorm(mu, sigma),
    mu <- a + b * (area - xbar),
    # Priors from earlier
    a ~ dnorm(5, 0.75),
    b ~ dnorm(0.4, 0.4),
    sigma ~ dunif(0, 5)
  ),
  data=d
)

##           mean         sd      5.5%     94.5%
## a     4.53944153 0.10833219 4.3663058 4.7125773
## b     0.05510892 0.11361874 -0.1264758 0.2366936
## sigma 1.17905755 0.07745406  1.0552710 1.3028441
```

Question: What association do you observe? What does your analysis suggest about how territory size relates to weight?

We observe a b value of 0.06 on average. This suggests that territory size has a very slightly positive effect on weight.

b) Regress weight on food availability. That is, construct a quap linear regression (`m1b`) to estimate the association of food availability and fox weight. *Before fitting the model, standardize both avgfood and weight to have mean 0 and standard deviation 1.*

Hint: *With standardized variables, regression slopes represent standardized effect sizes. A slope of 1.0 would indicate a perfect positive relationship, while slopes >2 would be implausibly large for most ecological relationships.*

Use prior predictive simulation to assess the implication of your priors. Write 1-2 sentences to justify your priors.

c) Now regress weight on *both* territory size and weight. Construct a quap model (`m1c`) that includes both predictors. Use the standardized variables. Explain your findings with 3-4 sentences and appropriate plots.

2 AI Declaration