

# Chapter 7 Questions

Melissa Van Busse

June 29, 2018

Chapter 7 points: 2 + 4 + 4 = 10 points total.

## Easy (2 points total)

### 7E1

- a) There's an interaction between yeast and sugar.
- b) A third variable would be choice of degree. Some education doesn't lead to higher income.
- c) Gasoline makes a car go... conditional on the car being turned on.

### 7E2

Only the first one is an interaction since it's the only one that can be reworded using "conditional".

### 7E3

a)

$$\mu_i = \alpha + \beta_H \text{heat} + \beta_D \text{dry} + \beta_{HD} \text{heat} \cdot \text{dry}$$

b)

$$\mu_i = \alpha + \beta_c \text{cylinder} + \beta_f \text{fuel injector}$$

c)

$$\mu_i = \alpha + \beta_P \text{parents} + \beta_F \text{friends}$$

d)

$$\mu_i = \alpha + \beta_S \text{social} + \beta_A \text{appendages}$$

## Medium (4 points total)

### 7M1 (2 points)

This is because there is a three-way interaction. This means that in order for tulips to bloom, temperature, water and shade must all interact, and hot temperatures do not support tulips.

### 7M2 (2 points)

Just multiply by an indicator variable. It's hot out, let's say that  $H = 0$ . Then the regression equation would be  $\mu = LH$ , where  $L$  is the linear equation that we already established in the chapter.

## Hard (4 points total)

### 7H1 (2 points)

#### The question

Return to the data (tulips) example in the chapter. Now include the *bed* variable as a predictor in the interaction model. Don't interact *bed* with the other predictors; just include it as a main effect. Note that *bed* is categorical. So to use it properly, you will need to either construct dummy variables or rather an index variable, as explained in Chapter 6.

#### The solution

First, load in the data:

```
library(rethinking)
library(knitr)
data(tulips)
d <- tulips
kable(head(tulips))
```

bed	water	shade	blooms
a	1	1	0.00
a	1	2	0.00
a	1	3	111.04
a	2	1	183.47
a	2	2	59.16
a	2	3	76.75

We can see that the *bed* variable is a factor variable, so let's see how many different factors it has:

```
d$bed
## [1] a a a a a a a a b b b b b b b b c c c c c c c c
## Levels: a b c
```

It has 3 levels. The question tells us we can either create dummy variables, or use an index variable. Let's create 2 dummy variables: one for b and one for c.

```
d$bedb <- rep(0, length(d$blooms))
d$bedc <- rep(0, length(d$blooms))
for (i in 1:length(d$bed)) {
  if (d$bed[i] == "b") {
    d$bedb[i] <- 1
  }
  if (d$bed[i] == "c") {
    d$bedc[i] <- 1
  }
}
```

We should also centre the non-factor variables, as in the chapter. These variables are *water* and *shade*. Note: even though it might not LOOK like these variables aren't factor variables, looking at the structure of the data frame reveals that the only factor variable the author of the dataset put in was *bed*:

```
str(d)
```

```
## 'data.frame': 27 obs. of 6 variables:
## $ bed : Factor w/ 3 levels "a","b","c": 1 1 1 1 1 1 1 1 1 2 ...
## $ water : int 1 1 1 2 2 2 3 3 3 1 ...
## $ shade : int 1 2 3 1 2 3 1 2 3 1 ...
## $ blooms: num 0 0 111 183.5 59.2 ...
## $ bedb : num 0 0 0 0 0 0 0 0 0 1 ...
## $ bedc : num 0 0 0 0 0 0 0 0 0 0 ...
```

So now we centre *water* and *shade*:

```
d$water <- d$water - mean(d$water)
d$shade <- d$shade - mean(d$shade)
```

Let's take a look at how the *blooms* variable is distributed, so we can figure out what priors to use later

```
mean(d$blooms)
```

```
## [1] 128.9937
```

```
sd(d$blooms)
```

```
## [1] 92.68392
```

Now we can fit the model using **map**. Remember, the question tells us to model all the variables, but to ignore the the interaction between *bed* and the other variables. Thus the model will be as follows:

$$\begin{aligned} \text{Blooms} &\sim \text{Normal}(\mu, \sigma) \\ \mu &= \alpha + \beta_w \text{water} + \beta_s \text{shade} + \beta_{ws} \text{water,shade} + \beta_b \text{bed}_b + \beta_c \text{bed}_c \\ \alpha &\sim \text{Normal}(129, 20) \\ \beta_w &\sim \text{Normal}(0, 100) \\ \beta_s &\sim \text{Normal}(0, 100) \\ \beta_{ws} &\sim \text{Normal}(0, 100) \\ \beta_b &\sim \text{Normal}(0, 100) \\ \beta_c &\sim \text{Normal}(0, 100) \\ \sigma &\sim \text{Uniform}(0, 100) \end{aligned}$$

```
lmod7H1 <- map(
  alist(
    blooms ~ dnorm(mu, sigma) ,
    mu <- alpha + beta_W*water + beta_S*shade + beta_WS*water*shade +
      beta_B*bedb + beta_C*bedc,
    alpha ~ dnorm(129, 20
  ),
  c(beta_W, beta_S, beta_WS, beta_B, beta_C) ~ dnorm(0, 100),
  sigma ~ dunif(0, 100)
),
data = d,
start = list(alpha = 129, beta_W = 0, beta_S = 0, beta_WS = 0,
  beta_B = 0, beta_C = 0, sigma = 93)
)
precis(lmod7H1)
```

```
##           Mean StdDev   5.5%  94.5%
```

```
## alpha    107.83  11.09  90.11 125.55
## beta_W    75.18   9.31  60.30  90.06
## beta_S   -41.26   9.31 -56.14 -26.38
## beta_WS  -52.17  11.38 -70.36 -33.98
## beta_B    34.13  17.07   6.85  61.40
## beta_C    38.71  17.07  11.43  65.99
## sigma    39.67   5.49  30.89  48.46
```

## 7H2 (2 points)

### The Question

Use WAIC to compare the model from 7H1 to a model that omits *bed*. What do you infer from this comparison?

```
# The model from the textbook (changed the variable names though)
m3 <- map(
  alist(
    blooms ~ dnorm(mu,sigma),
    mu <- alpha + beta_W*water + beta_S*shade + beta_WS*water*shade,
    alpha ~ dnorm(130, 100),
    c(beta_W,beta_S,beta_WS) ~ dnorm(0, 100),
    sigma ~ dunif(0, 100)
  ),
  start = list(alpha = 130, beta_W = 0, beta_S = 0, beta_WS = 0, sigma = 90),
  data = d)

compare(lmod7H1, m3)
```

```
##           WAIC pWAIC dWAIC weight    SE dSE
## lmod7H1 293.5    8.7   0.0   0.77  9.46  NA
## m3      295.8    6.5   2.4   0.23 10.34 5.52
```

We can see that the model from question 1 has a slightly better value of WAIC, but not by much. Thus, we can conclude that including the variable *bed* in the model helps, but not enough for it to really matter – the treatments (*water* and *shade*) mattered much more.