chp5

Vasco Brazão

15/02/2021

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

5E1

(4) would be the standard way to write a multiple linear regression. I Suppose (2) could be valid? If we force the intercept to be 0. (3) seems plausible, but from the lack of an index on beta I would think you're forcing the beta for x to be equal to -1 * the beta for z, which is.. strange?

Van Bussel agrees https://github.com/castels/StatisticalRethinking/blob/master/Chapter%205/VanBussel_Chapter5_Questions.pdf

5E2

mu_latitude_i = alpha + beta_adiv * adiv_i + beta_pdiv * pdiv_i

5E3

$$time_i \sim \text{Normal}(\mu, \sigma)$$

$$\mu_i = \alpha + \beta_f f_i + \beta_s s_i$$

Both slope parameters should be positive.

Van Bussel agrees!

But I still can't make a stupid latex document. One day.

5E4

1, 3, 4 would be my guesses.

Van Bussel disagrees - 4 is not correct. But I still think it works?

And a latex document thing was created! I cannot believe my eyes. What fresh hell awaits me now? We shall see.. we shall see.

5M1

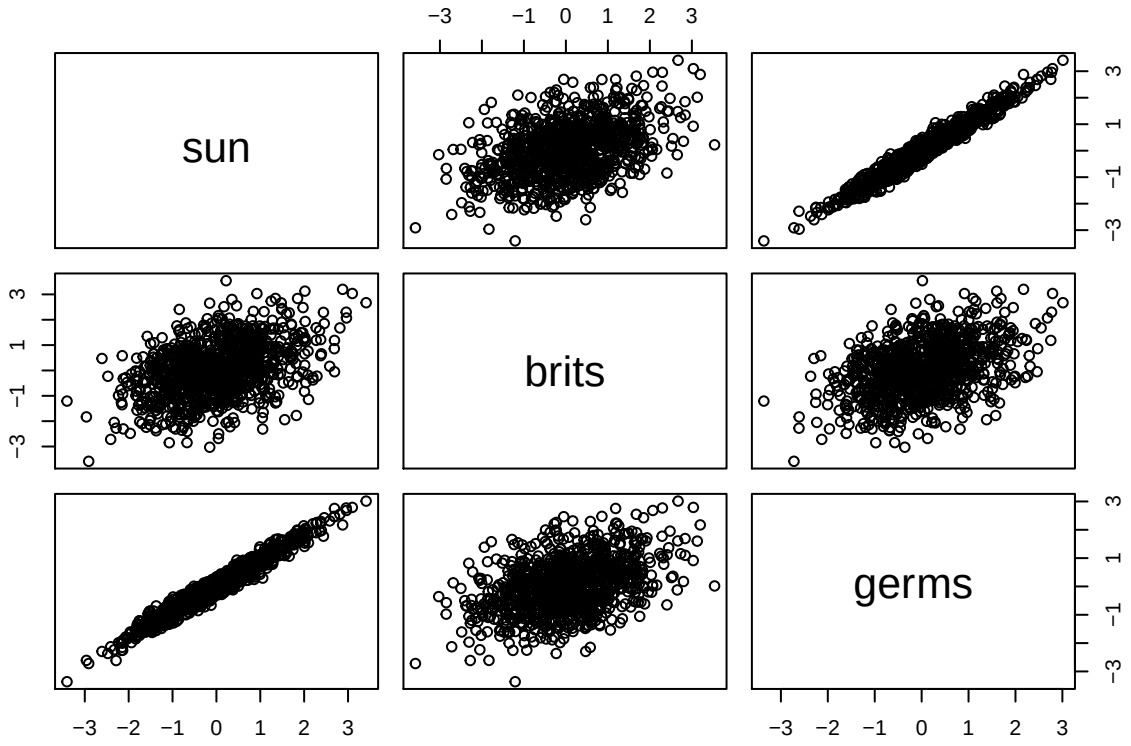
Inventing a spurious correlation.

Let's say that both British and German tourists like going to sunny places. For a given year, they base their decision of whether to go Portugal for vacation on the number of sunlight days of the previous year. Germans respond more strongly to this parameter because they now have an easier time going to Portugal than Brits, and they also stray less from the line because they are German.

```
n <- 1000

df <- tibble(
  sun = rnorm(n, 0, 1),
  brits = 0.5*sun + rnorm(n, 0, 1),
  germs = 0.9*sun + rnorm(n, 0, 0.2)
)
```

```
pairs(df)
```



There appears to be a positive relationship betweet amount of brits and amount of germans.

```
m1 <- lm(df$brits ~ df$germs)
m2 <- lm(df$brits ~ df$germs + df$sun)

summary(m1)
```

```
##
## Call:
## lm(formula = df$brits ~ df$germs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8695 -0.6682  0.0048  0.6669  3.5218
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01078    0.03077    0.35    0.726
## df$germs     0.47405    0.03298   14.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9731 on 998 degrees of freedom
## Multiple R-squared:  0.1715, Adjusted R-squared:  0.1707
## F-statistic: 206.6 on 1 and 998 DF,  p-value: < 2.2e-16
```

```
summary(m2)
```

```
##
## Call:
## lm(formula = df$brits ~ df$germs + df$sun)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9548 -0.6398 -0.0055  0.6604  3.4460
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01274    0.03070   0.415   0.6782
## df$germs     0.08730    0.15430   0.566   0.5717
## df$sun       0.36345    0.14167   2.565   0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9704 on 997 degrees of freedom
## Multiple R-squared:  0.1769, Adjusted R-squared:  0.1753
## F-statistic: 107.1 on 2 and 997 DF,  p-value: < 2.2e-16
```

Et voila. when sun is in the model, knowing how many germans went to Portugal does not really tell us much more about how many brits went.