20240904_lab0

Yerim Oh

Beer example

In this lab, "we examine data from sixteen student volunteers at Ohio State University who each drank a randomly assigned number of cans of beer. These students were evenly divided between men and women, and they differed in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood."[1,2]

References:

(1) Diez, David M., Christopher D. Barr, and Mine Cetinkaya-Rundel. OpenIntro statistics. Boston, MA, USA. OpenIntro, 2012.
(2) J. Malkevitch and L.M. Lesser. For All Practical Purposes: Mathematical Literacy in Today's World. WH Freeman & Co, 2008.

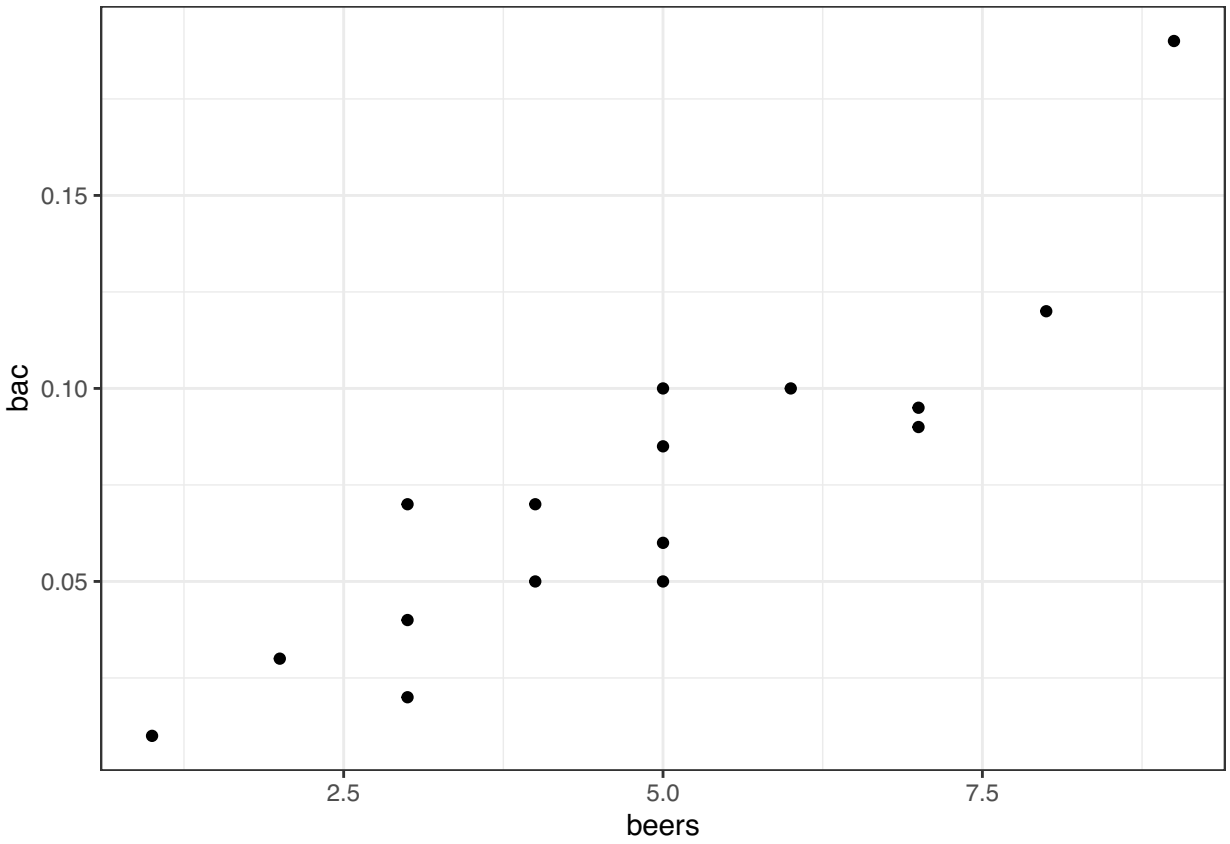
a) What are the response and explanatory variables?

response variable: blood alcohol content (BAC) in grams of alcohol per deciliter of blood

explanatory variable: number of cans of beer

b) Plot the data and describe the relationship between the variables.

```
ggplot(data = beer, mapping = aes(x = beers, y = bac)) +
  geom_point() +
  theme_bw()
```



c) Which function f could describe the relationship between X and Y?

A linear relationship seems appropriate.

Estimated equation of the line: $f(\text{beers}) = \beta_0 + \beta_1 \times$ beers

linear relationship → bac = f(beer) = $\beta_0 + \beta_1 \times$ beers

d) How could we estimate this function (name the method or code it in the following R chunk)?

Method of Least Squares

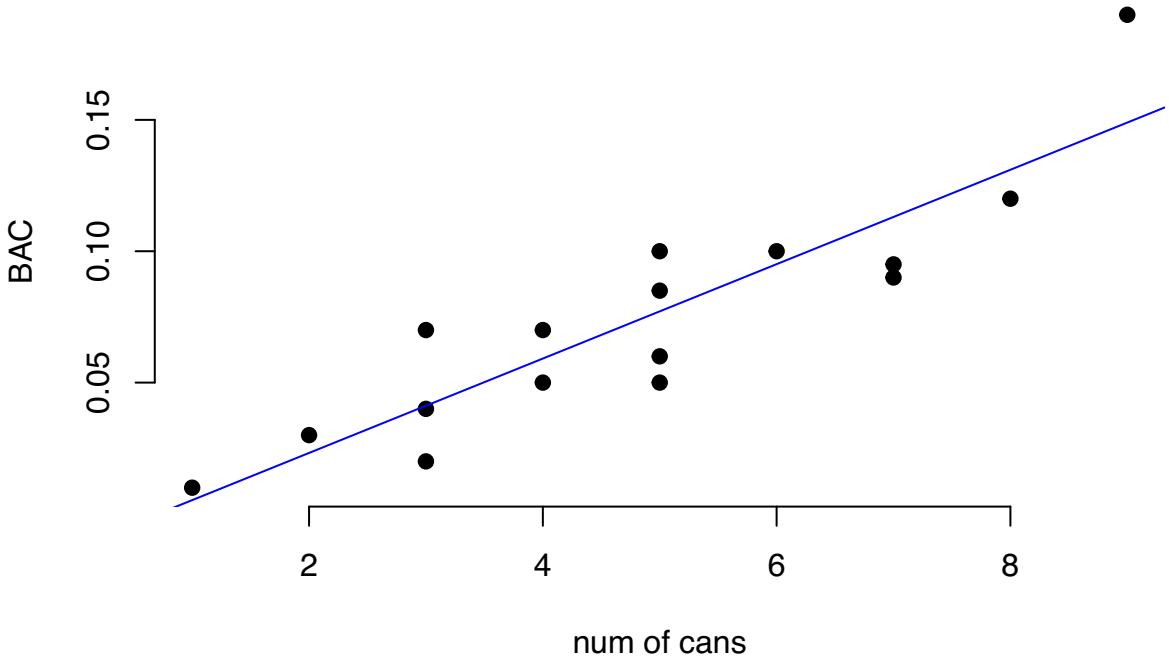
$\hat{f}(\text{beers}) = -0.0127 + 0.01796 \times \text{beers}$

```
# fit the model
beer_model <- lm(bac ~ beers, data = beer)
summary(beer_model)
```

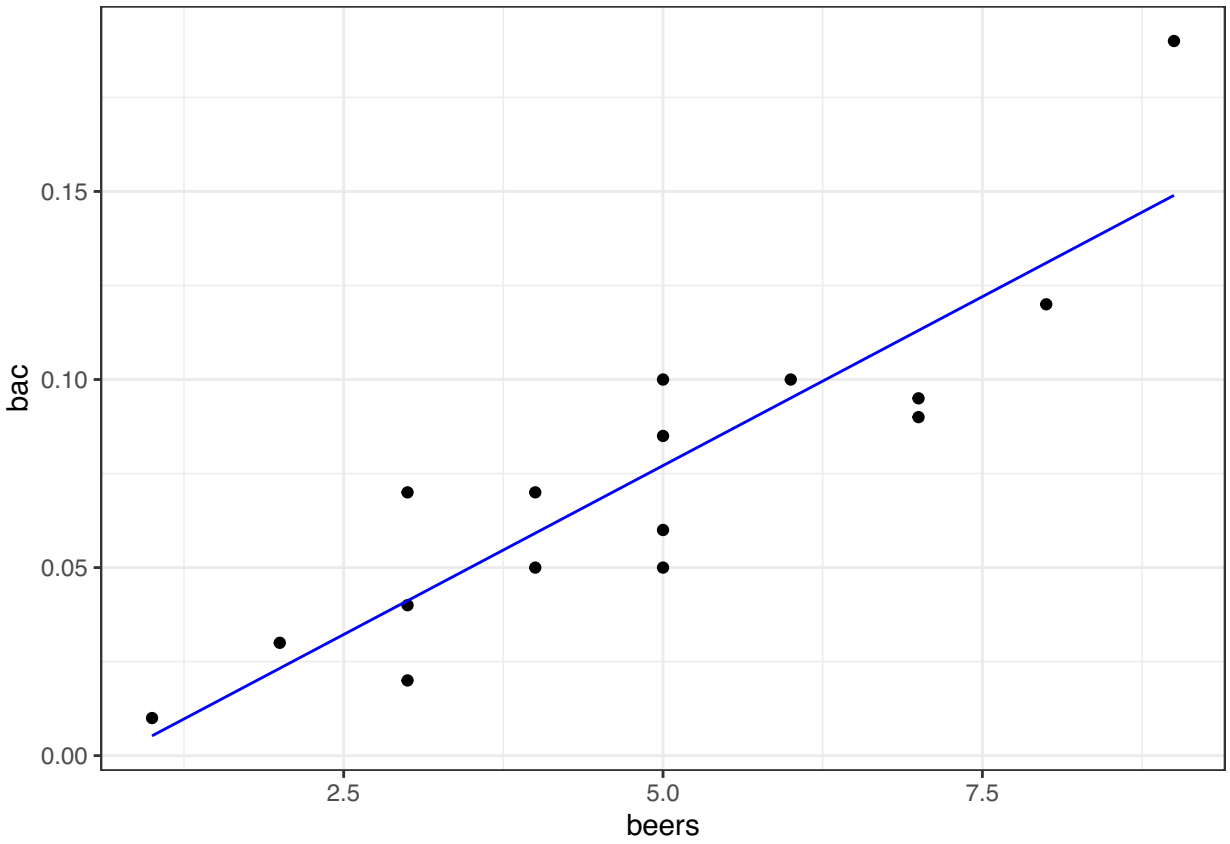
```
##
## Call:
## lm(formula = bac ~ beers, data = beer)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.027118 -0.017350  0.001773  0.008623  0.041027
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.012701   0.012638  -1.005    0.332
## beers        0.017964   0.002402   7.480 2.97e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02044 on 14 degrees of freedom
## Multiple R-squared:  0.7998, Adjusted R-squared:  0.7855
## F-statistic: 55.94 on 1 and 14 DF,  p-value: 2.969e-06
```

```r
# plot the linear function: 2 ways
plot(beer$beers, beer$bac,
     xlab = "num of cans", ylab = "BAC",
     pch = 19, frame = FALSE)
abline(beer_model, col = "blue")
```




```r
beer <- beer %>%
  mutate(fitted = predict(beer_model))
ggplot() +
  geom_point(data = beer, mapping = aes(x = beers, y = bac)) +
  geom_line(data = beer, mapping = aes(x = beers, y = fitted), color = "blue") +
  theme_bw()
```


0.15
0.10
0.05
0.00
bac
2.5
5.0
7.5
beers