Homework_02

Emmenta Janneh

2024-01-27

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
cdi <- read.csv("https://dcgerard.github.io/stat_415_615/data/cdi.csv")
head(cdi)
```

```
##   id      county state area     pop percent_18_34 percent_65 physicians  beds
## 1  1 Los_Angeles    CA 4060 8863164          32.1        9.7      23677 27700
## 2  2        Cook    IL  946 5105067          29.2       12.4      15153 21550
## 3  3      Harris    TX 1729 2818199          31.3        7.1       7553 12449
## 4  4   San_Diego    CA 4205 2498016          33.5       10.9       5905  6179
## 5  5      Orange    CA  790 2410556          32.6        9.2       6062  6369
## 6  6       Kings    NY   71 2300664          28.3       12.4       4861  8942
##   crimes high_school bachelors poverty unemployment capita_income total_income
## 1 688936        70.0      22.3    11.6          8.0         20786       184230
## 2 436936        73.4      22.8    11.1          7.2         21729       110928
## 3 253526        74.9      25.4    12.5          5.7         19517        55003
## 4 173821        81.9      25.3     8.1          6.1         19588        48931
## 5 144524        81.2      27.8     5.2          4.8         24400        58818
## 6 680966        63.7      16.6    19.5          9.5         16803        38658
##   region
## 1      W
## 2     NC
## 3      S
## 4      W
## 5      W
## 6     NE
```

County Demographic Information

log(y) = 1.87 + 0.79log(x)

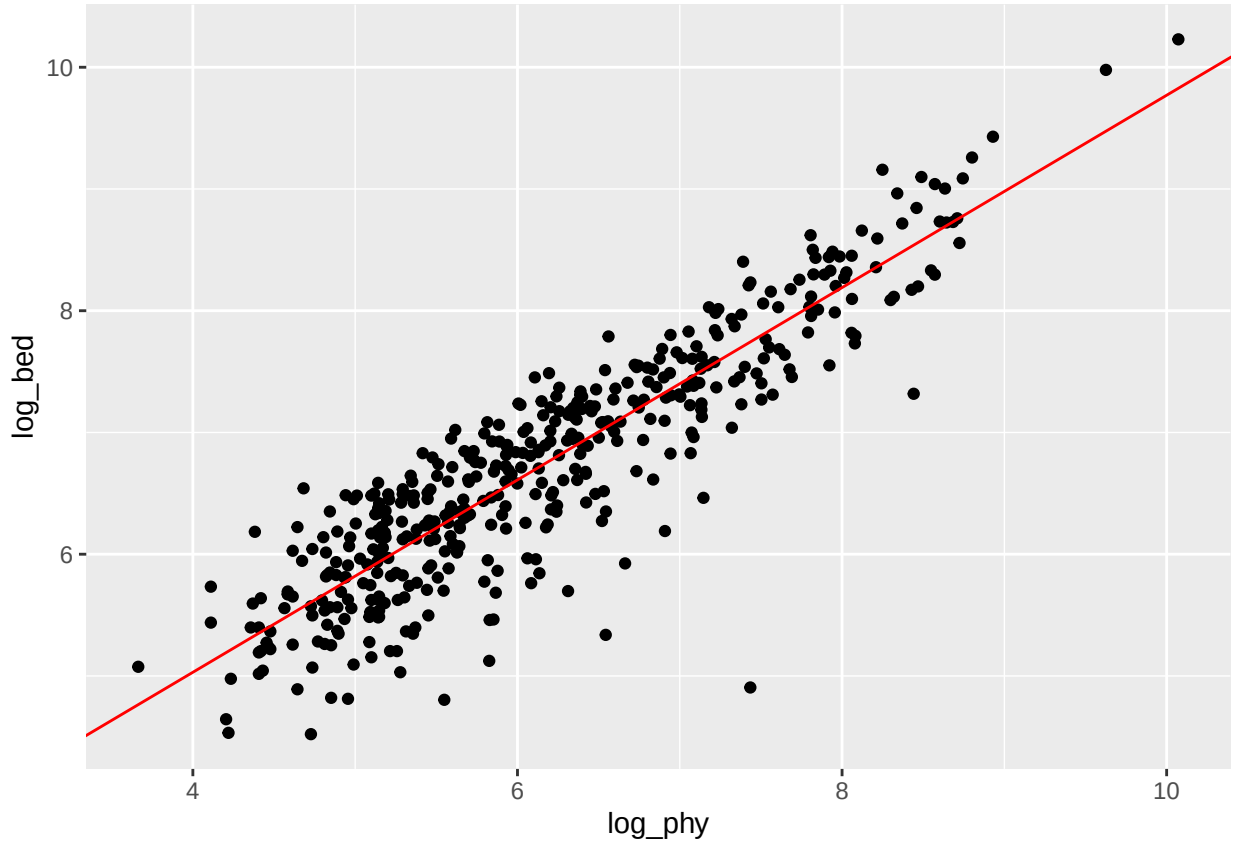
1. The given equation is an example of a **Power-law** relationship. In a power-law relationship, the dependent variable(y) is related to the independent variable(x) through the power function, expressed as: y = y.Cˆ B1 where C is a constant and B1 is the exponent or slope of the line.

2.

```
reg_table <- mutate(cdi, log_phy = log(physicians),
                         log_bed = log(beds)) %>%
               select(log_phy, log_bed)
head(reg_table)
```

```
##     log_phy   log_bed
## 1 10.072259 10.229188
## 2  9.625954  9.978131
## 3  8.929700  9.429396
## 4  8.683555  8.728912
## 5  8.709795  8.759198
## 6  8.488999  9.098515
```

```
ggplot(data = reg_table, aes(x = log_phy, y = log_bed)) +
  geom_point() +
  geom_abline(slope = 0.79, intercept = 1.87, color = "red")
```



3. log(y) = 1.87 + 0.79log(x)
   y = 1.87 . xˆ0.79
4. For every C number of beds in a hospital, there is Cˆ0.79 as many physicians, on average.

log(y) = 6.26 + 0.022x

5. The given equation is an example of a **Exponential growth/decline** relationship. In a exponential growth/decline relationship, the dependent variable(y) is related to the independent variable(x) through the exponential function, expressed as: y = y.exp(C*B1) where C is a constant and B1 is slope of the line.
6. For every unit higher in beds in the hospital, there is exp(0.022) times more beds, on average.

Rephrase: For each unit in beds, there is a 2.2% times more bachelors degree holders, on average.

7. **6.26** is the y-intercept of the regression line.

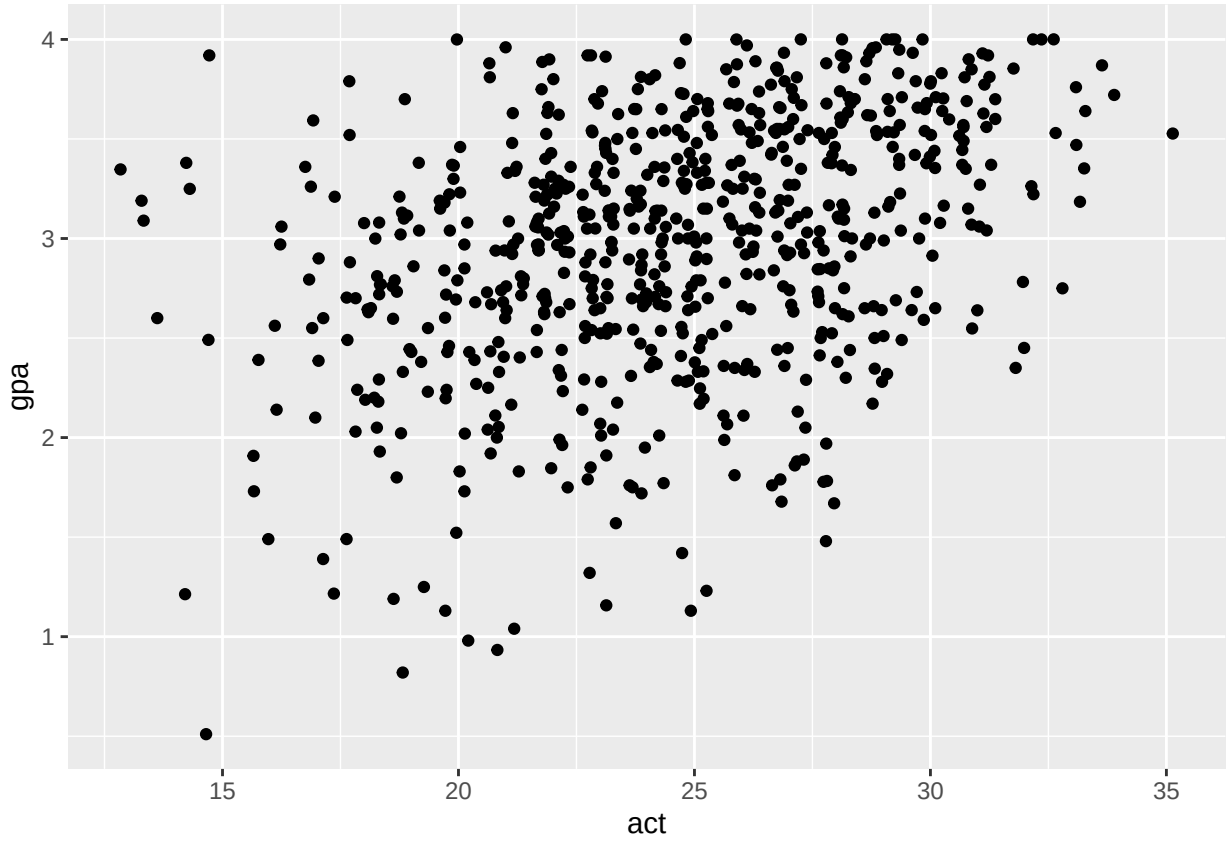
University Admissions Data

```
uad <- read.csv("https://dcgerard.github.io/stat_415_615/data/university.csv")
head(uad)
```

```
##   id  gpa rank act year
## 1  1 0.98   61  20 1996
## 2  2 1.13   84  20 1996
## 3  3 1.25   74  19 1996
## 4  4 1.32   95  23 1996
## 5  5 1.48   77  28 1996
## 6  6 1.57   47  23 1996
```

1.

```
ggplot(uad, aes(x = act, y = gpa)) +
  geom_jitter()
```



From the above plot, there is an association between gpa and act of students. Although the association is a little faint, the graph shows a positive association between the variables.

2. **0.058**: For every unit score in ACT(x), student's GPA is higher by 0.058, on average. **1.56**: This is the GPA score when ACT score is 0.

Miscellaneous

1. No I won't be happy, the economy is recovered an I am expecting to get the same value as my old salary or more. Decreasing my salary by 10% and later increasing it by another 10% will not get the same old value.