

Prediction Competition 1

15/01/2024

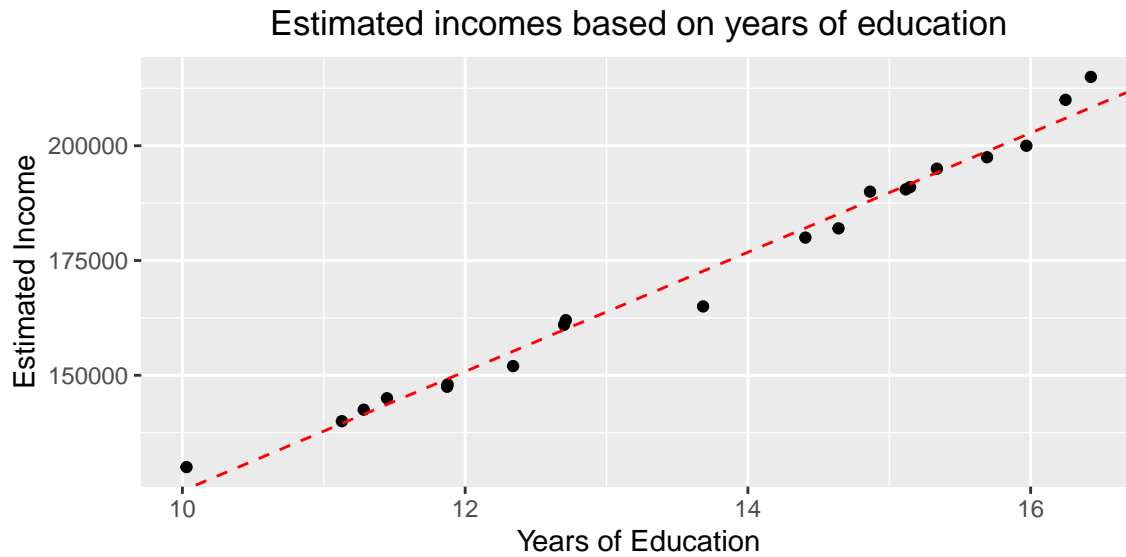
Anonymized name: Sukuna

Question 2a

```
model = lm(income_est ~ education)
model_summary = summary(model)
b0 = model_summary$coefficients[1]
b1 = model_summary$coefficients[2]
print(model_summary)
```



```
##
## Call:
## lm(formula = income_est ~ education)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7685  -1771    -98    1505   6685
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5067.8     5239.9  -0.967   0.346
## education    12990.7     380.2   34.165 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3271 on 18 degrees of freedom
## Multiple R-squared:  0.9848, Adjusted R-squared:  0.984
## F-statistic: 1167 on 1 and 18 DF, p-value: < 2.2e-16
```



```
y1 = round(b0 + b1*(12))
y2 = round(b0 + b1*(13))

income_diff = round(b1 * (13 - 12))
```

\$12991.

In this economy, I estimate Jimmys income with 12 years of education to be \$150820 and his income with 13 years of education to be \$163811. The difference in expected lifetime earnings for 13 years of education compared to 12 years is \$12991. I arrived at this answer using my linear regression model.

Question 2b

```
current_avg_income = mean(income_est)
future_avg_income = b0 + b1*(17)

avg_income_diff = future_avg_income - current_avg_income
```

\$43574

In this economy I estimate the average lifetime earnings increases by \$43574 if the government assigned everyone 17 years of education. I arrived at this answer through computing the average of my predictions incomes, computing my linear models average at 17 years, then taking the difference of the two averages.

Question 3

- The measured college outcomes are first-year cumulative GPA, percentage of grades that are A or A- and an indicator for academic struggle defined as having any grade of C+ or lower in the first year.
- The variables that are the most important predictors of college outcomes a student's standardized test score. The least important predictor of college outcomes are a student's high school GPA.
- In my opinion, the student's standardized test score is the most important variable. I believe this because all students are meant to take the same standardized exam and as stated in the paper, a persons socioeconomic background has little effect on the SAT and how well the individual does in their first year of college.

Question 4

I asked chat gpt:

```
can you turn these two columns into two variables in R script [11.28278 & 1425000 \ 13.68312 & 165000\
14.40669 & 180000\ 15.96969 & 200000\ 14.86386 & 190000\ 15.33683 & 195000\ 12.69982 & 161000\ 11.447
& 145000\ 11.87571 & 148000\ 11.87264 & 147500\ 15.14703 & 191000\ 11.12808 & 140000\ 15.11679 &
190500\ 12.33917 & 152000\ 16.24666 & 210000\ 16.42583 & 215000\ 15.69144 & 197500\ 12.7121 & 162000\
14.64092 & 182000\ 10.02925 & 130000\]
```

Chatgpt replied:

Certainly! In R, you can create two variables from the given data using the `c()` function. Here's an example of how you can do it:

```
data <- c(11.28278, 13.68312, 14.40669, 15.96969, 14.86386, 15.33683, 12.69982, 11.447, 11.87571, 11.87264,
15.14703, 11.12808, 15.11679, 12.33917, 16.24666, 16.42583, 15.69144, 12.7121, 14.64092, 10.02925) labels
<- c(1425000, 165000, 180000, 200000, 190000, 195000, 161000, 145000, 148000, 147500, 191000, 140000,
190500, 152000, 210000, 215000, 197500, 162000, 182000, 130000)
```

```
variable1 <- data variable2 <- labels
```

```
print(variable1) print(variable2)
```