

# Plot height and weight of the "women" dataset. Make the title "Heights and Weights"

```
plot(women$weight, women$height, main="Heights and Weights")
```

# Make a contingency table of tobacco consumption and education

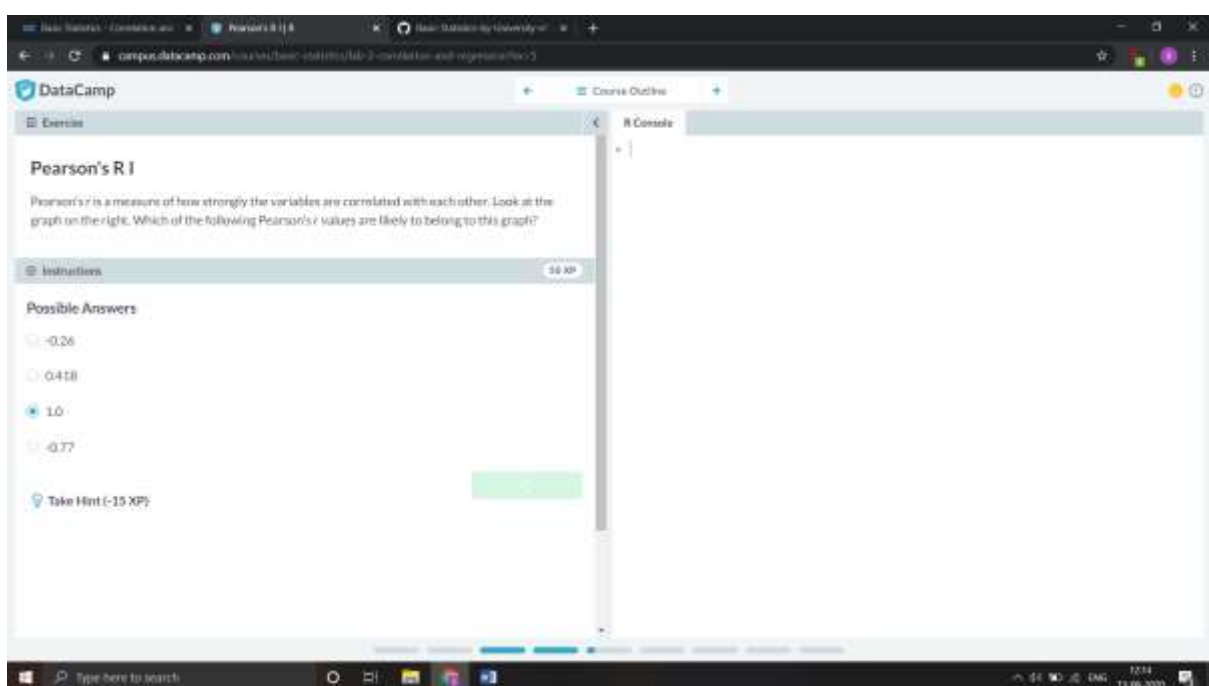
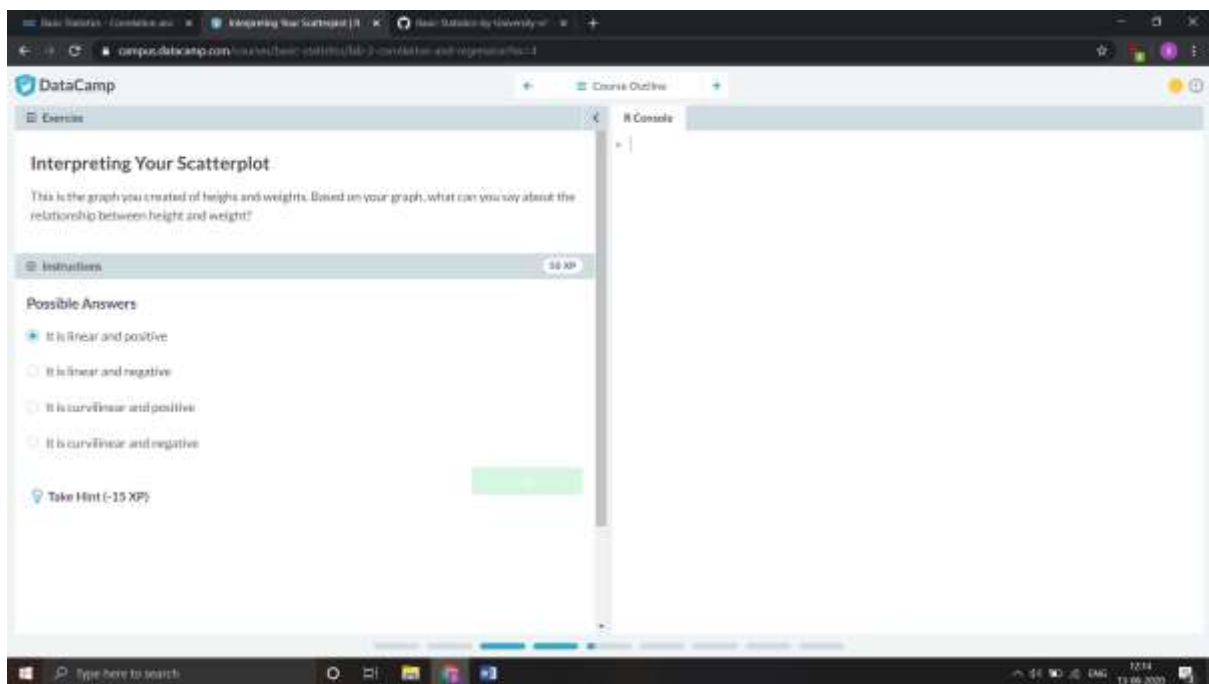
```
table(smoking$tobacco, smoking$student)
```

# What percentage of high school students smoke 0-9g of tobacco?

38.6

# Of the students who smoke the most, what percentage are in university?

57.7



Basic Statistics / Correlation and Regression / Pearson's R II

Exercise

### Pearson's R II

Which of the following Pearson's  $r$  values are likely to belong to this graph?

Instructions 50 XP

Possible Answers

- ☐ -0.26
- ☒ 0.418
- ☐ 1.0
- ☐ -0.77

Take Hint (-15 XP)

R Console

```
> |
```

Windows Taskbar: Type here to search, 12:14, 13.09.2020

Basic Statistics / Correlation and Regression / Pearson's R III

Exercise

### Pearson's R III

Which of the following Pearson's  $r$  values are likely to belong to this graph?

Instructions 50 XP

Possible Answers

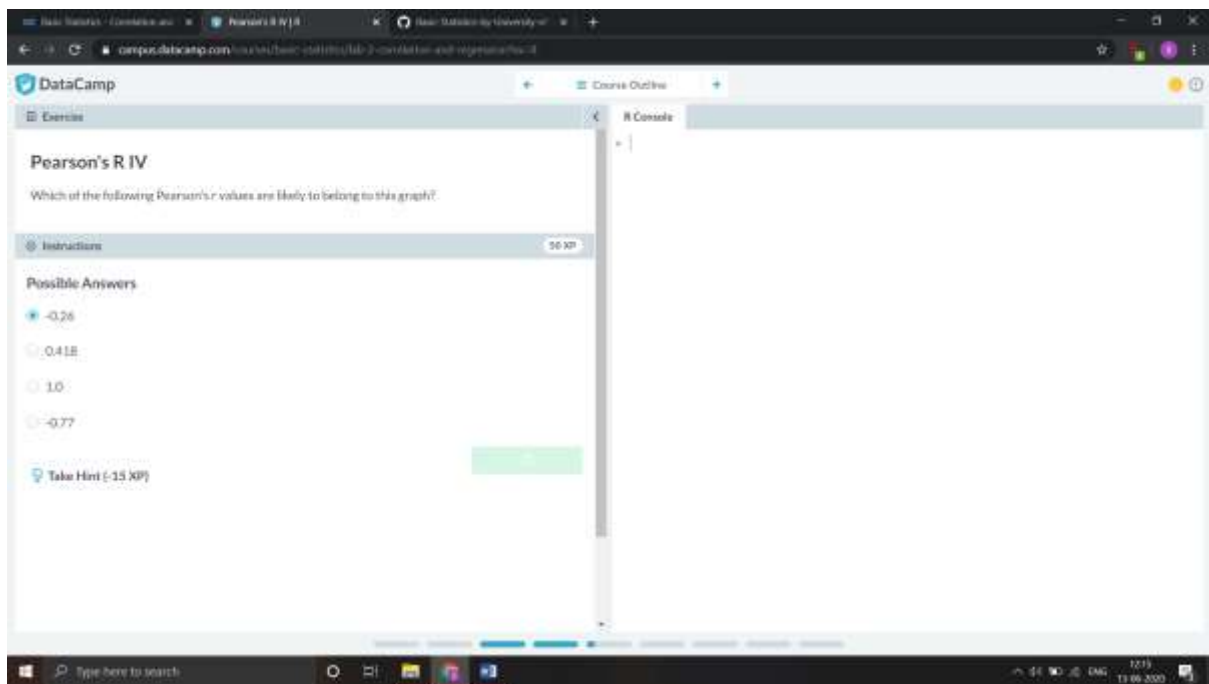
- ☐ -0.26
- ☐ 0.418
- ☐ 1.0
- ☒ -0.77

Take Hint (-15 XP)

R Console

```
> |
```

Windows Taskbar: Type here to search, 12:15, 13.09.2020



# Calculate the correlation between var1 and var2

```
cor(var1, var2)
```

# predicted values of y according to line 1

```
y1 <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
```

# predicted values of y according to line 2

```
y2 <- c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)
```

# actual values of y

```
y <- c(3, 2, 1, 4, 5, 10, 8, 7, 6, 9)
```

# calculate the squared error of line 1

```
sum((y1 - y) ^2)
```

# calculate the squared error of line 2

```
sum((y2 - y) ^2)
```

# How prosocial would we predict someone to be when they receive 6 units of money?

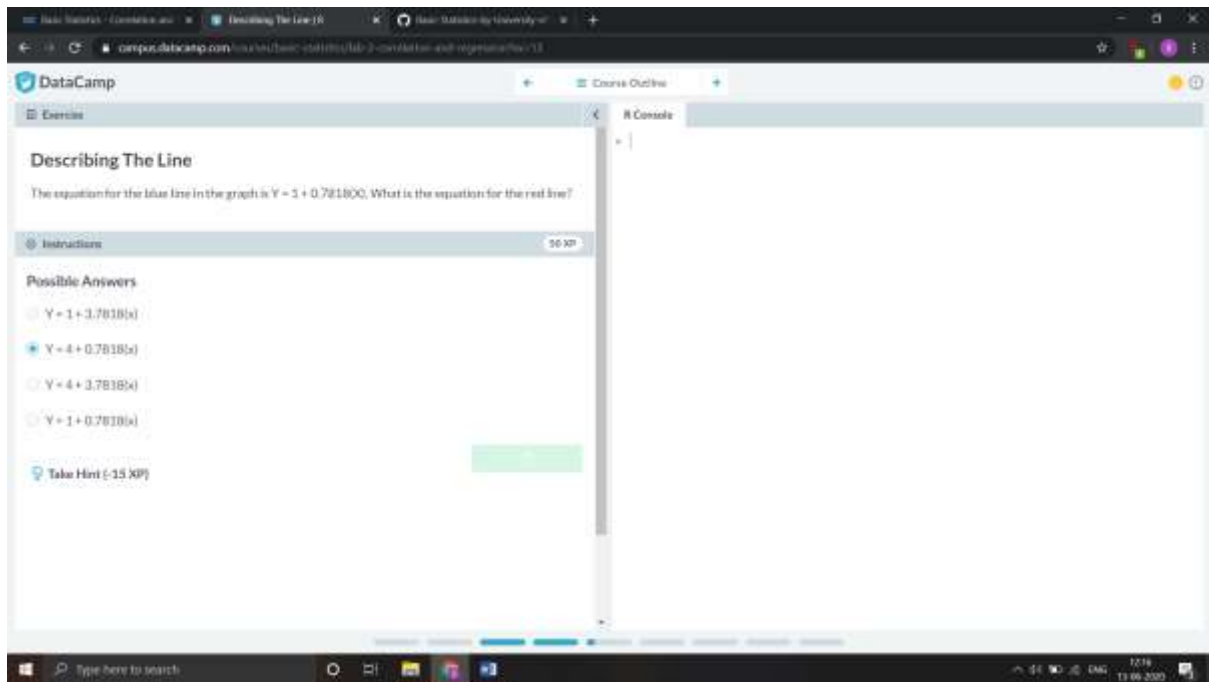
6

# How prosocial was the person who received 6 units of money in our study?

10

# What is the risk taking level of a person with an extraversion level of 7?

26.1



# Our data

```
money <- c(1,2,3,4,5,6,7,8,9,10)
```

```
prosocial <- c(3, 2, 1, 4, 5, 10, 8, 7, 6,9)
```

# Find the regression coefficients

```
lm(prosocial ~ money)
```

# Your plot

```
plot(money, prosocial, xlab = "Money", ylab = "Prosocial Behavior")
```

# Store your regression coefficients in a variable called "line"

```
line <- lm(prosocial ~ money)
```

# Use "line" to tell abline() to make a line on your graph

```
abline(line)
```

# Your plot

```
plot(money, prosocial, xlab = "Money", ylab = "Prosocial Behavior")
```

# Your regression line

```
line <- lm(prosocial ~ money)
```

```
abline(line)
```

# Add a line that shows the mean of the dependent variable

```
abline(mean(prosocial), 0)
```

# Calculate the R squared of prosocial and money

`cor(money, prosocial) ^2`

The screenshot shows a DataCamp exercise interface. The title is "R Squared II". The question asks: "In addition to being the reduction in residual error from using the regression line over the mean line, and, of course the Pearson correlation coefficient squared, how else can we describe the R squared?". There are four possible answers, with the third one selected: "The variation in the dependent variable explained by the independent variable". A "Submit Answer" button is visible.

The screenshot shows a DataCamp exercise interface. The title is "Correlation and Causation". The question asks: "You measured how much money people have and their education level in a town. The graph on the right shows the results. We cannot say that more education causes more money, we say it is related to more money. Which of the following is not a reason why education is only related to money?". There are four possible answers, with the first one selected: "There could be a third unmeasured variable that influences only money". A "Submit Answer" button is visible. To the right of the question is a scatter plot titled "Plots" showing a positive correlation between education and money, with a regression line. The plot has "Education" on the x-axis and "Money" on the y-axis. Below the plot are buttons for "Previous Plot" and "Next Plot".

# your data

```
money <- c(4, 3, 2, 2, 8, 1, 1, 2, 3, 4, 5, 6, 7, 9, 9, 8, 12)
```

```
education <- c(3, 4, 6, 9, 3, 3, 1, 2, 1, 4, 5, 7, 10, 8, 7, 6, 9)
```

# calculate the correlation between X and Y

```
cor(money, education)
```

# save regression coefficients as object "line"

```
line <- lm(money ~ education)
```

```
# print the regression coefficients  
line  
# plot Y and X  
plot(education, money, main="My Scatterplot")  
# add the regression line  
abline(line)  
# Percentage of people with high money that are university educated  
83.3  
# Percentage of people with low money that are high school educated  
72.7  
# What kind of education is linked to more money?  
"university"
```