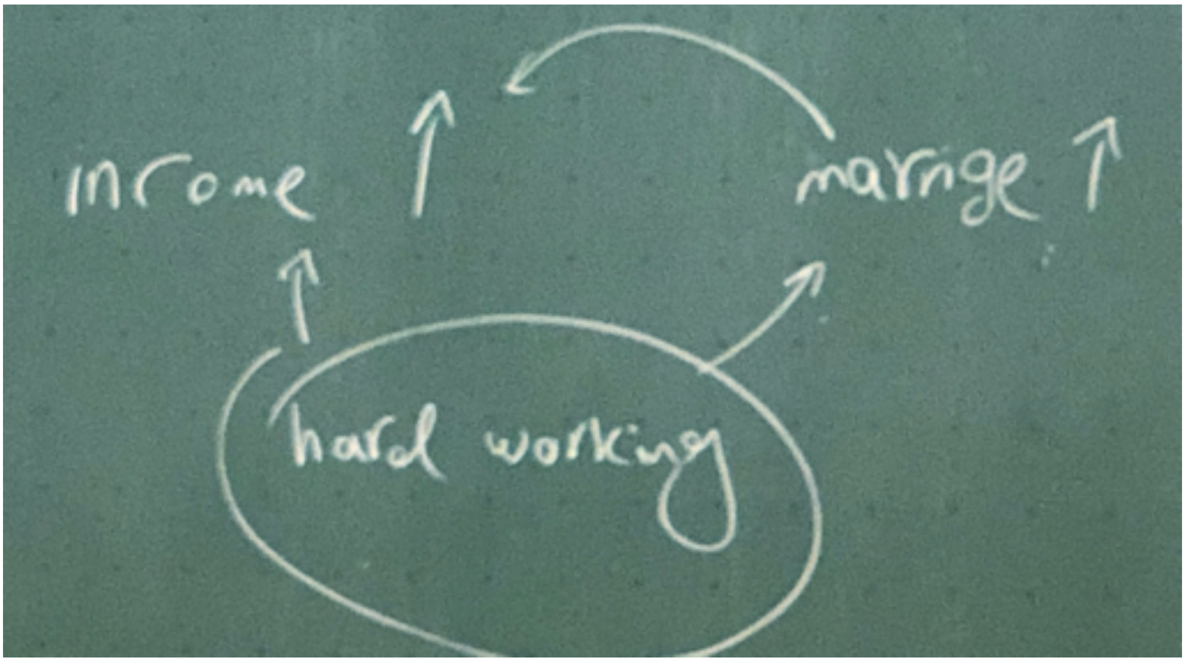
P1 Study questions

Where do we use aggregated data instead of segregated data and why it's useful?
A: The blood pressure example from the lecture. We do not use it for initial decision making, rather looking at the causation "after the fact".

1.2.1
What is wrong with the following?

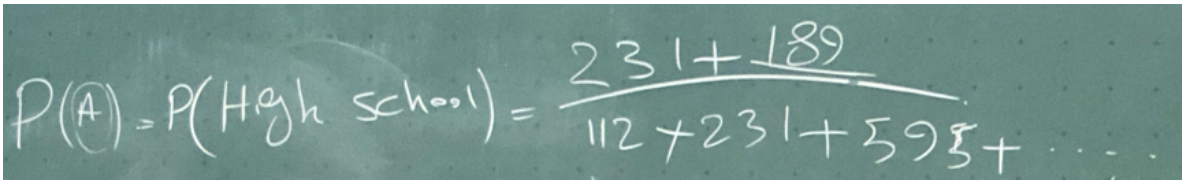
a) Correlation is confused with causality. There is no direct correlation while there might be additional variables like "hard working"



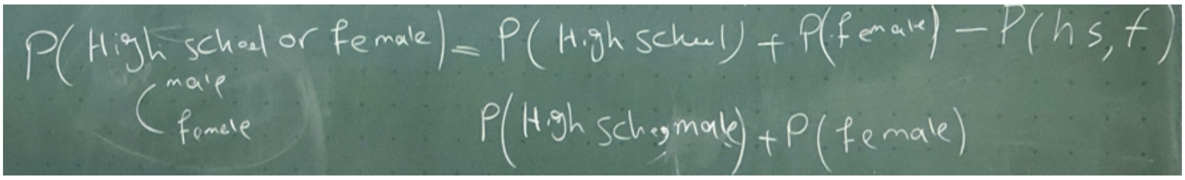
b) Direct correlation between fires and firefighters but not the other way around. We increase the fires, firefighters increases but not the other way (decreasing firefighters does not mean we have less fires).
c) You left late so you hurry but not the other way around. There can be a third variable "sleeping in".

1.2.3
a) Stone size-specific data
b) Difficulty-specific data

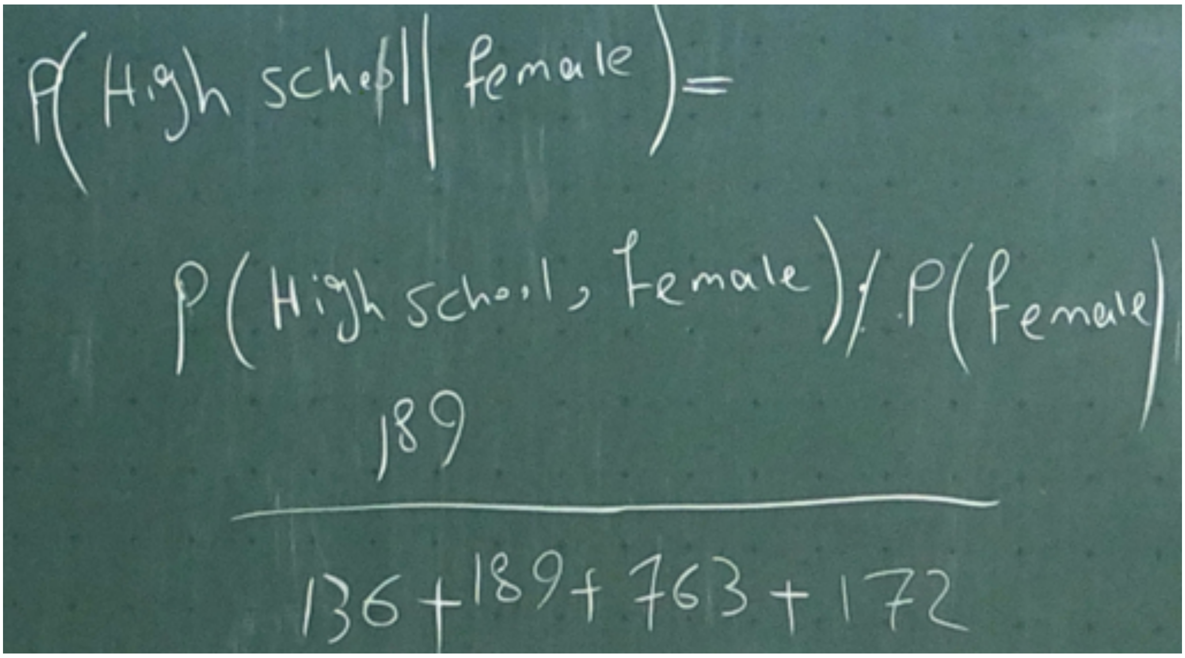
1.3.2
a)



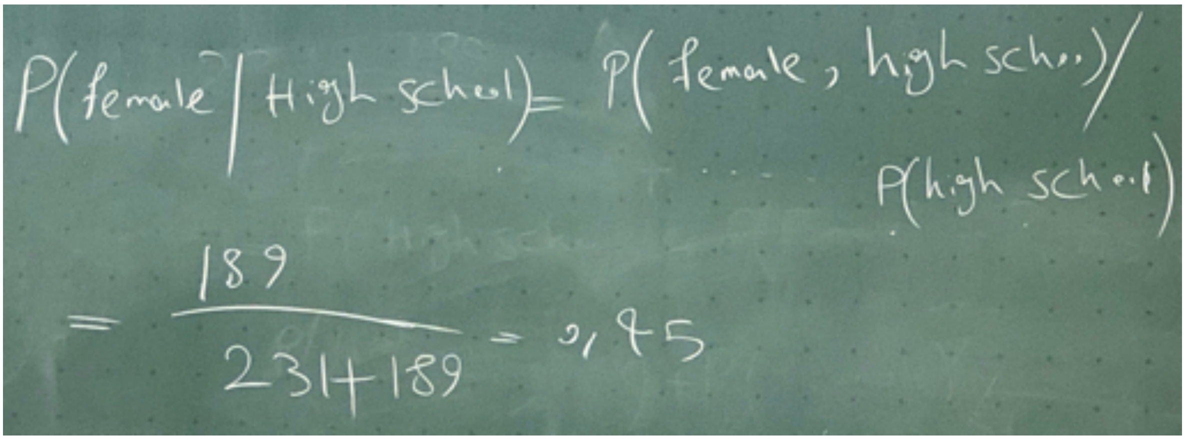
b)



c)



d)

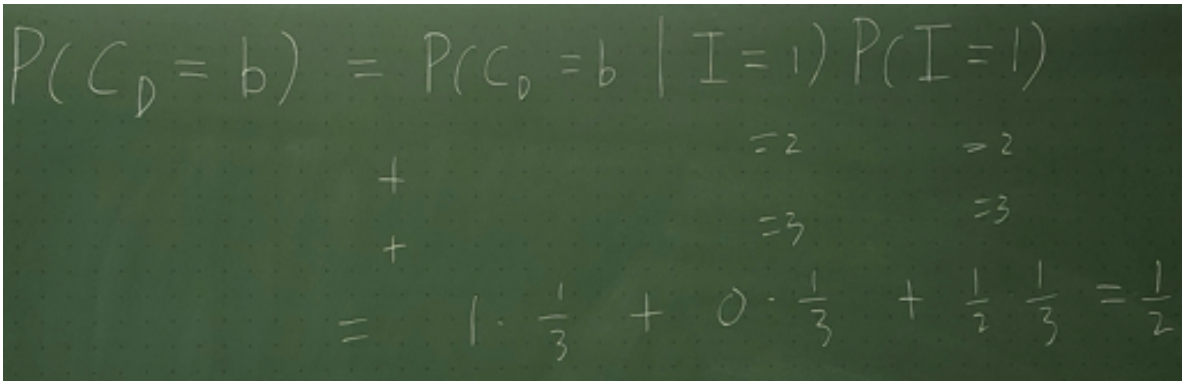


1.3.4

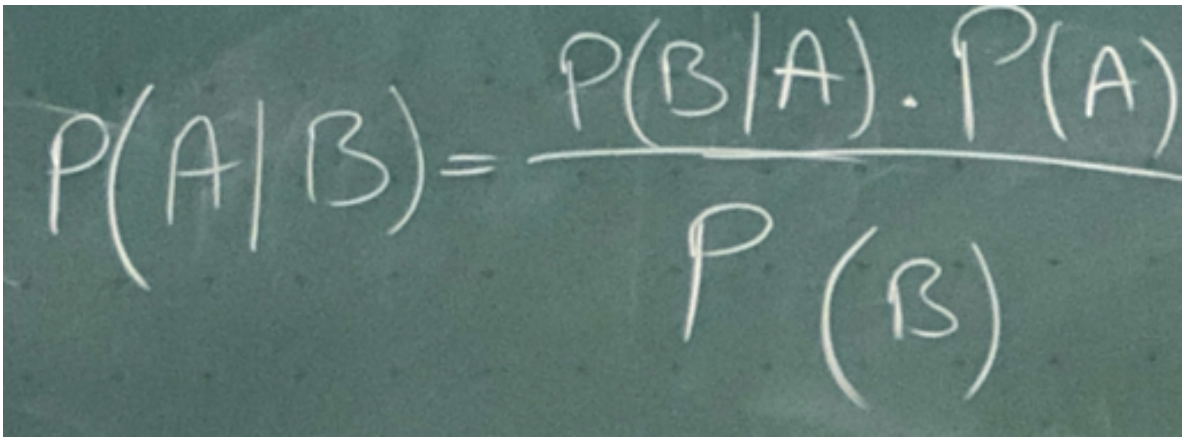
a) It is more than 1/2

b) If we choose a black-black card, then you know that the P of being black is 1. Also, you have 1/2 P that, the next card is white-black.

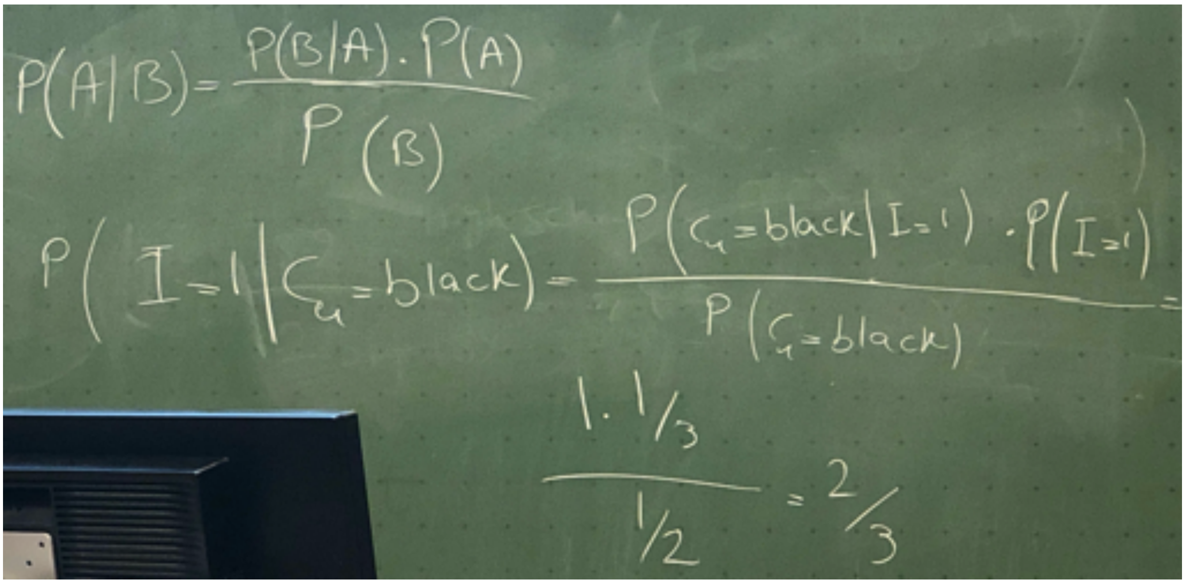
Find the probability that the face-down....



c)
Bayes theorem:

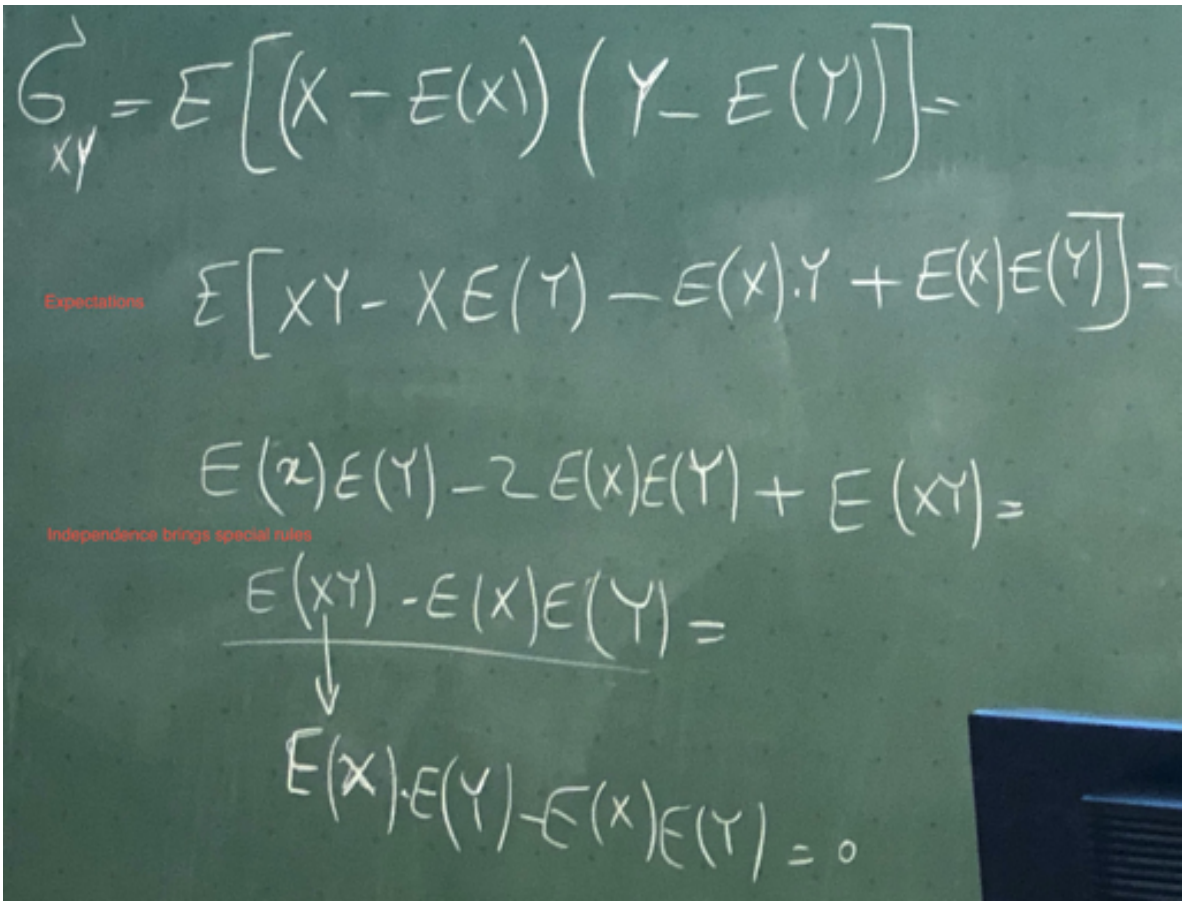


Solution to the question:



1.3.6

a) (If the two variables are zero, show that the covariance is also zero. Use the covariance equation.)



For homework

- 1.3.6 b)
- Read up on statistics and the covariance equation
- 1.3.1
- 1.3.3
- 1.3.7