Task 1.1

A university determined the following joint frequency distribution of the variables "gender" (G), "subject" (F) and "accepted" (A) in applications for university places:

`('Fach'='subject', 'Geschlecht'='gender', 'Angenommen'='accepted')`

	Angenommen	0	1
Fach	Geschlecht		
Informatik	m	36	54
	w	18	27
IntMgmt	m	36	9
	w	108	27
WI	m	27	27
	w	18	18

a) Show that there is a relationship (=stochastic dependence) between G and A.

b) Use the data to check whether the correlation *can be explained* by the variable F. (This would mean that G has an influence on F, and F on A, but G does not have a direct influence on A.)

c) Interpretation:

   i. For machine learning: If you want to learn a classification model that predicts A, is there any point in using both, G and F as inputs, or would one of the variables suffice?

   ii. Is there evidence of gender discrimination, i.e. are men or women favored because of their gender?

d) Carry out the calculations for (a) and (b) ...

   i. ... by hand based on the table shown above. Make your calculations comprehensible.

   ii. ... in Python. The raw data and the code used to generate the table shown can be found in ILIAS under "Code examples/03_Discrete_Data_University_Applications ".

Hints and tips:

1. Hint for (b): If G influences F and F influences A, but G does not influence A directly, then G and A must be conditionally independent of each other.
2. A summary of the relevant statistical basics can be found in ILIAS under "Expected_prior_knowledge/00_Statistics_Foundations" - especially on slides 6 - 20