# Influences on Personal Income - An analysis for Germany

## 1. Introduction

The income affects how one lives, even if the person does not have to struggle at the poverty line. A higher income is associated with better health and increased quality of life. Therefore, it is interesting to see which personal characteristics that should not influence the value of the work are related to the level of income.

## 2. Hypotheses Development

The overarching research question is: "*Which solely personal characteristics can be associated with an increased monthly net income?*" To answer this, the influences of relationship status, gender, and region of living on net income will be considered while accounting for age and educational attainment level. The following hypotheses will be tested:

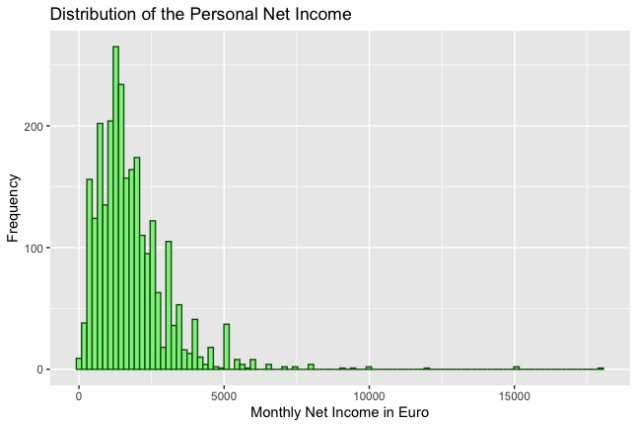
> *$H1_0$: Women have a lower net income than men.*
> *$H2_0$: People living in the West of Germany have a higher net income than people living in the East of Germany.*
> *$H3_0$: Married people have a higher net income than people who are not married.*

## 3. Data

The research is conducted for Germany and is based on the data set ALLBUS, the General Population Survey of Social Sciences ("*Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*"), from 2018 with 3477 observations. The survey is conducted every two years and focuses on attitudes, behaviors, and social structure of the population in Germany. Variables that will be utilized are monthly net personal income, gender, age, educational level, relationship status, and region.

| | Min. | Max. | Median | Mean | Variance | Cor(Y,) |
|---|---|---|---|---|---|---|
| $Y_{income}$ | 25 | 18000 | 1500 | 1789 | 1279.4 | 1 |
| $X_{sex}$ | 0 | 1 | 0 | 0.49 | 0.500 | -0.315 |
| $X_{education}$ | 0 | 2 | 1 | 1.15 | 0.793 | 0.276 |
| $X_{age}$ | 18 | 95 | 53 | 51.68 | 17.64 | 0.0542 |
| $X_{married}$ | 0 | 1 | 1 | 0.62 | 0.487 | 0.136 |
| $X_{eastwest}$ | 0 | 1 | 0 | 0.3135 | 0.434 | -0.147 |

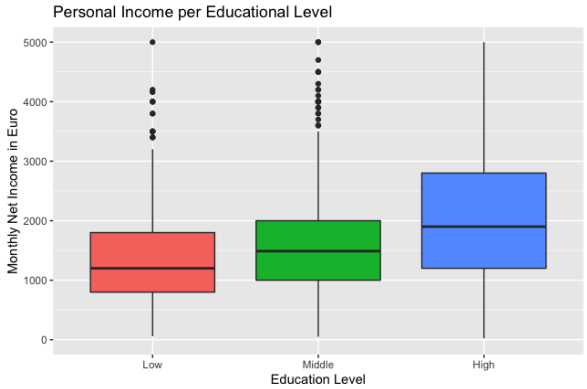
To see how those variables have been constructed in detail see Appendix 1 (Variable denotation see Appendix 1 - Table 1).

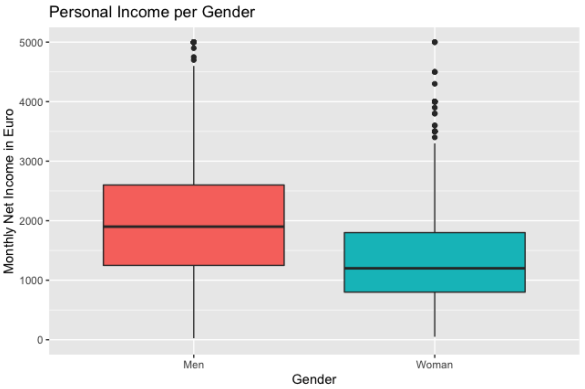



People's net monthly income has a wide span, the lowest being 25€, an income on which one is not able to sustain themselves, the highest being 18000€. The min value shows the shortcoming of the variable "personal net income", as the person could be living in a household in which their spouse, parent, or children are the breadwinners. The median net income of 1500€ is in line with government statistics, however, the mean of the participants is, as expected through self-reporting bias, lower than the country's average of about 3500€.

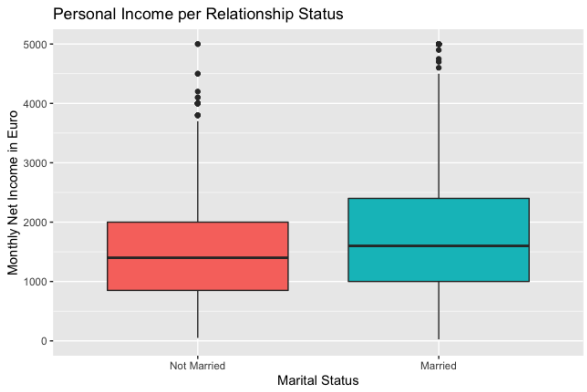
The data set's gender distribution is balanced (49% woman). Further, most people in the sample have attained a middle level of education (mean = 1.15). As only Germans from age 18-95 have participated, the median age of the sample is 53 which is slightly higher than the actual median age in Germany which is 47.3. Lastly, 31.3% of the participants are from Eastern Germany, which is an over-representation of the actual eastern german population (21%). The visualizations of the distributions for the explanatory variables can be found in Appendix 2 Graph 2-6.

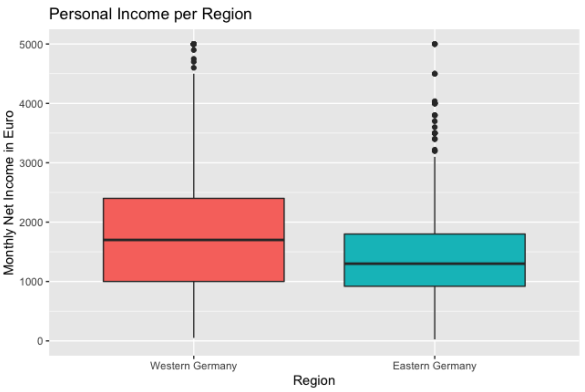
The Pearson correlation of the variables with income provides the first insights for future analysis. The differences in (mean and span of) personal net income for the categorical variables can be nicely visualized with the following Boxplots. The strongest correlation can be found between income and gender (-0.315) and income and educational level (0.276). While gender (male to female) is negatively and education (rising levels) positively correlated

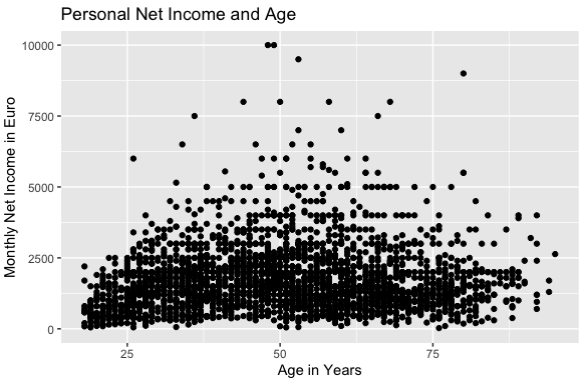



Marital Status (0.146) and the region (-0.147) show a medium correlation to income. Age shows the lowest correlation to net income (0.05). While marital status (Not Married to Married) and rising age are positively and Region (West to East) negatively correlated.




To summarize, whilst income looks fairly normally distributed with outliers to the left, all listed variables show some kind of correlation to personal net income. Whilst age is the least impactful. In the future, a linear regression can show whether the differences have a significant impact in explaining the variation in personal net income.

## 4. Appendix 1 - Data Wrangling

### 4.1 Personal Income

The net income of the respondent was retrieved by asking “How high is your own net income”. The income was reported in Euro.

### 4.2 Gender

For explanatory variables the variable gender (sex) with (1) “man” and (2) “woman”.
The expressions are re-coded to:

(0) “male” from (1)
(1) “female” from (2)
There are no missings.

### 4.3 Age

For age the metric age variable is used. People who refused to announce their age are re-coded as missing.

### 4.4 Educational Level

This variable represents the general school-leaving qualification of a person as this is a good indicator for educational achievement in Germany. The interviews could choose between: (1) "Finished school without graduation" (2) "Completed 9th-grade graduation" (3) "Completed 10th-grade graduation" (4) "Specialized upper secondary school" (5) "University entrance qualification" (6) "Other" and (7) "Still a student."
The expressions are re-coded to:

(0) "Low level" from (1) and (2)
(1) "Middle level" from (3)
(2) "High level" by (4) and (5)
All others were re-coded as missings.

### 4.5 Partnership

The marital status of the respondent was retrieved by asking whether a person is (1) "Married and living with a spouse," (2) “Married and living apart," (3) "Widowed," (4) “Divorced”, (5) “Single”, (6) "Registered civil partnership, living together," (7) "Registered civil partnership, living apart," (8) "Registered civil partner deceased," and (9) "Registered civil partnership dissolved."
The expressions are re-coded to:

(0) “Not Married” build from (3), (4), (5), (8) and (9)
(1) “Married” build from (1), (2), (6) and (7)
All others were re-coded as missings.

### 4.6 Region

The region should take west-east differences into account. The variable expresses people living in (1) “old federal states” and (2)” new federal states”.
The expressions are re-coded to:

(0) “west-germany” from (1)
(1) “east-germany” from (2)
There are no missings.

After this 2437 observations that show no NA in any of the variables can be used for a lineare regression.

## 5. Appendix 2 - Further Tables & Graphs

### Table 1 - Variable denotation

| Variable | Description | Type |
|---|---|---|
| $Y$ | Income | Metric on a scala from 25 to 18000 |
| $X_{sex}$ | Gender | Binominal with (0) Men and (1) Woman |
| $X_{education}$ | Educational attainment | Ordinal with (0) Low, (1) Middle and (2) High |
| $X_{age}$ | Age | Metric on a scala from 18 to 95 |
| $X_{married}$ | Marital Status | Binominal with (0) Not Married and (1) Married |
| $X_{eastwest}$ | Region | Binominal with (1) West Germany and (1) East Germany |

### Graph 2 - 6: Distribution of Variables