**Exploratory data analysis on the GSOEP dataset**

Sushma G Herakal Tejashwini Hosamani Dolaanand S

PES1UG19CS528 PES1UG19CS539 PES1UG19CS423

PES UNIVERSITY PES UNIVERSITY PES UNIVERSITY

Bangalore , Karnataka Bangalore , Karnataka Bangalore,Karnataka

[sushmagh92@gmail.com](mailto:sushmagh92@gmail.com) [onlyteju5@gmail.com](mailto:onlyteju5@gmail.com) [dolaanand737@gmail.com](mailto:dolaanand737@gmail.com)

# Abstract:

The dataset we considered was collected in order to investigate the determinants of secondary school choice of 14 year old students in Germany.

* We started our analysis with imputing the missing values using mean or median for numerical variables and mode for categorical variables.
* There was inconsistent capitalisation in the variable ‘state’ that was removed.
* We plotted various graphs to understand how the school choices varied among students.
* The dataset was normalised in order to change the values to a common scale, without affecting the differences in the range of values.
* We used correlation coefficients, scatter plots and heat maps to determine if there was any dependency between the variables.
* A hypothesis test was conducted to determine if the mean income of the population was 70000 or not.

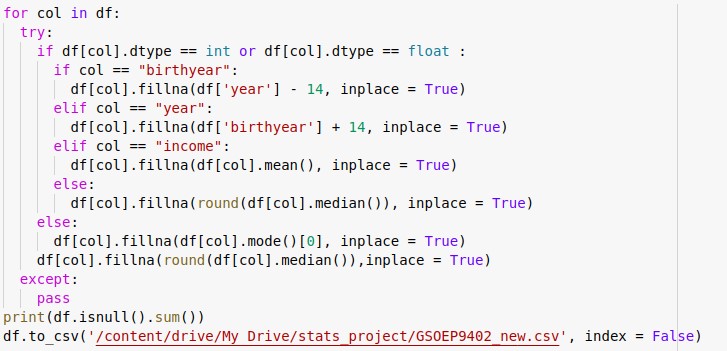
# Introduction

Our dataset is a cross section data for 675 14 year old children born between 1980 and 1988. The sample is taken from the German Socio - Economic Panel (GSOEP) for the years 1994 to 2002 to investigate the determinants of secondary school choice.

After cleaning up our dataset, our main analysis was based on the variations of the various continuous, discrete and categorical variables with the school choice of the children. We also tried to test the mean income of the population, the linear relationship between the variables and inferences were drawn from the results produced.

# Preprocessing or Data Cleaning

* Imputing missing values

There were around 20 missing values in each variable. Our dataset had 3.11% missing values. We replaced the missing values in the variable ‘income’ with the mean of the column. For the variables birth year and year, from the description of the dataset we knew that the children under consideration were 14 year olds, in order to calculate the missing birth year, 14 was deducted from the year. If the year was missing then 14 was added to the birth year. In case, both year and

birth year were missing, we replaced the values with the median. The other numerical missing values were replaced with the median of the variable and rounded off to the nearest integer.The missing values in categorical variables were replaced with mode of the variable.

* Fixing inconsistent capitalization

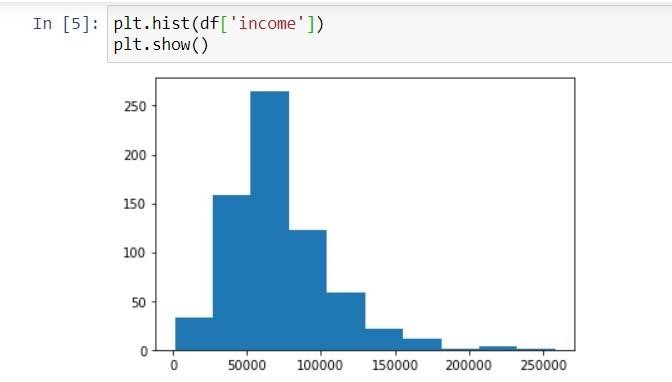


Only the variable ‘state’ had inconsistent capitalisation. All the words were first converted into lower and then the title method



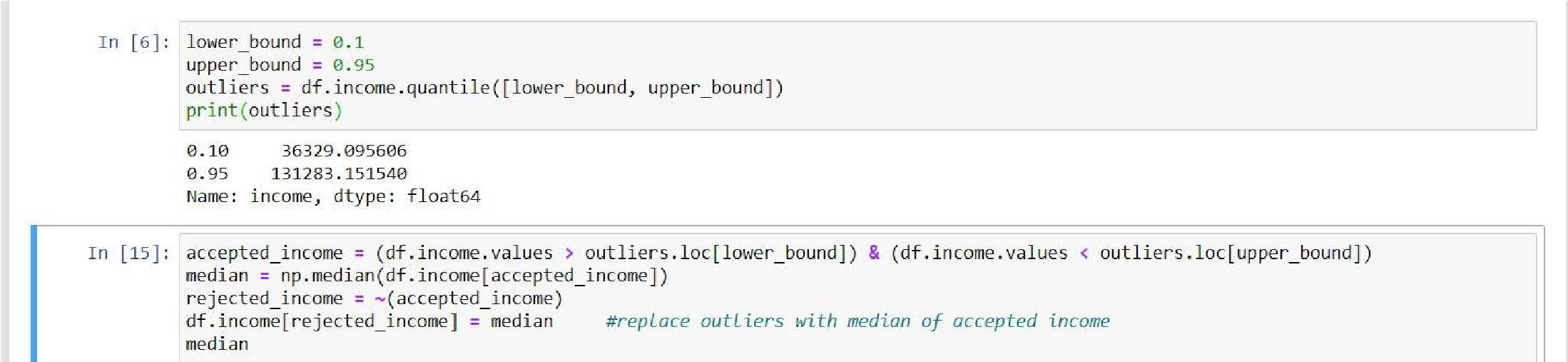
was used to fix the capitalization.

* Removal of outliers

When we plotted the histograms for the variables, it was observed that the variable ‘income’ had outliers in it. To remove these outliers, we used the Inter - Quartile Range (IQR) method.

In the IQR method, only those income values that lie between the 10th and the 95th percentile are accepted. The rejected income values were then replaced with the mean of the accepted range.

Code Snippet:

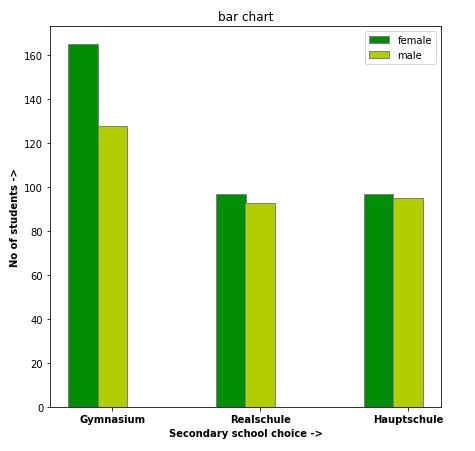
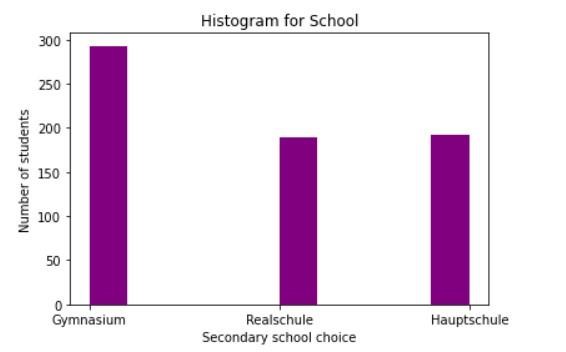


# Exploratory Data Analysis

The first part of data analysis involved graph visualisation. We plotted 3 graphs

1. Histogram
2. Barchart
3. Boxplot

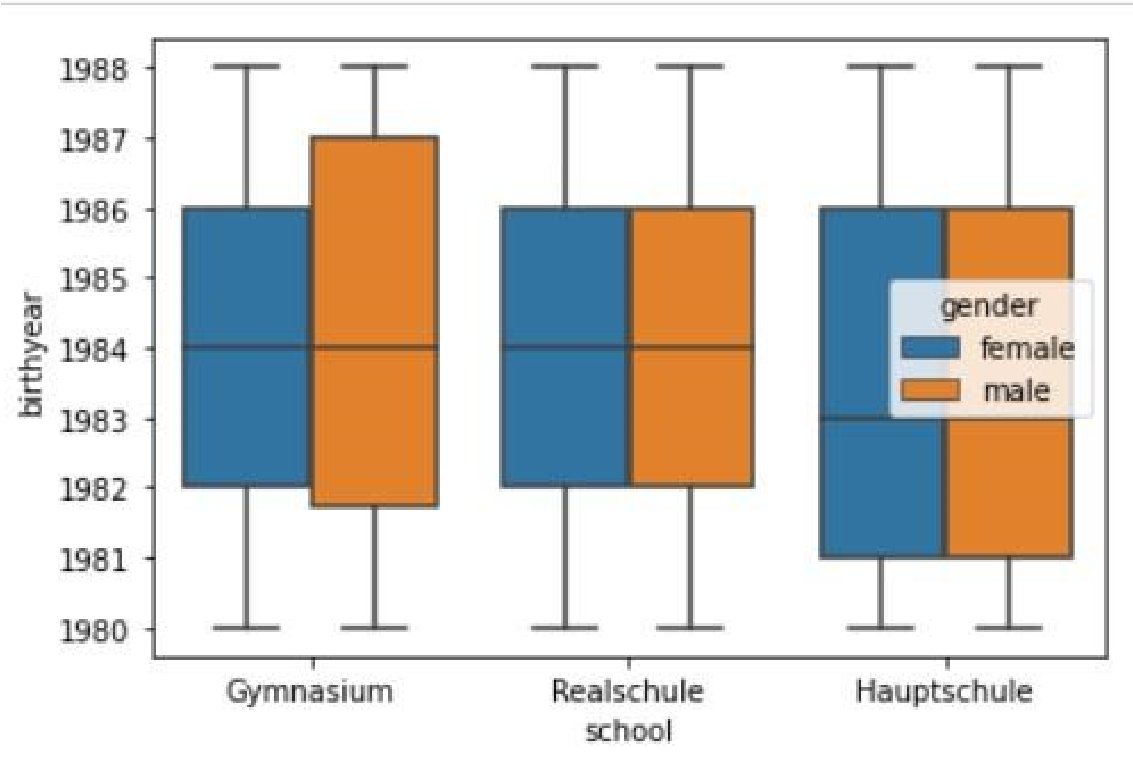
Histogram was plotted for the variable ‘school’. In order to determine the popularity of the schools among students, a histogram was plotted with school as the variable. The inferences drawn were:



Most of the students, (approximately 300 of them) opted for Gymnasium as their secondary school choice. Realschule and Hauptschule were equally popular among students.

A bar chart was plotted to compare the number of female and male students in each school.

The inferences drawn were: The number of female students in gymnasium(around 165) is more than the male students(around 125). While they are almost equally distributed in Realschule and Hauptschule.

In order to find out about the spread of values of birth year among the various schools with respect to gender, a box plot was constructed.

The inferences drawn were:

Realschule and Gymnasium have most of their students born in 1984, whereas Hauptschule students were born in 1983.

The spread of data for male students is much higher than female implying that the popularity of this school was more consistent among males. For both Realschule and Hauptschule, the 3rd Quartile is 1986 for both genders. The boxplot

for these schools are similar for both genders indicating similar spread/popularity among the genders.

- Normalization

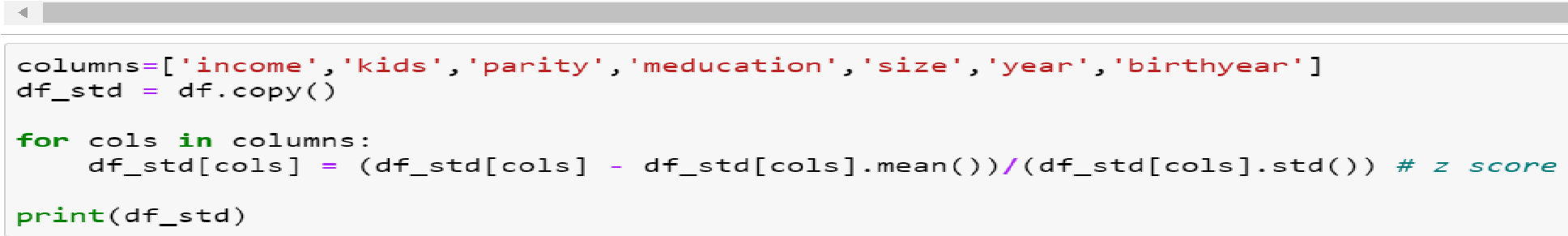
Normalization of ratings means adjusting values measured on different scales to a notionally common scale, often prior to averaging. The aim of Normalization is to change the values in the dataset to a common scale, without changing the differences in the range of values.

Normalization only changes the numeric values of mean, median and mode without affecting the overall behaviour and distribution of the dataset.

There are many Normalization techniques.

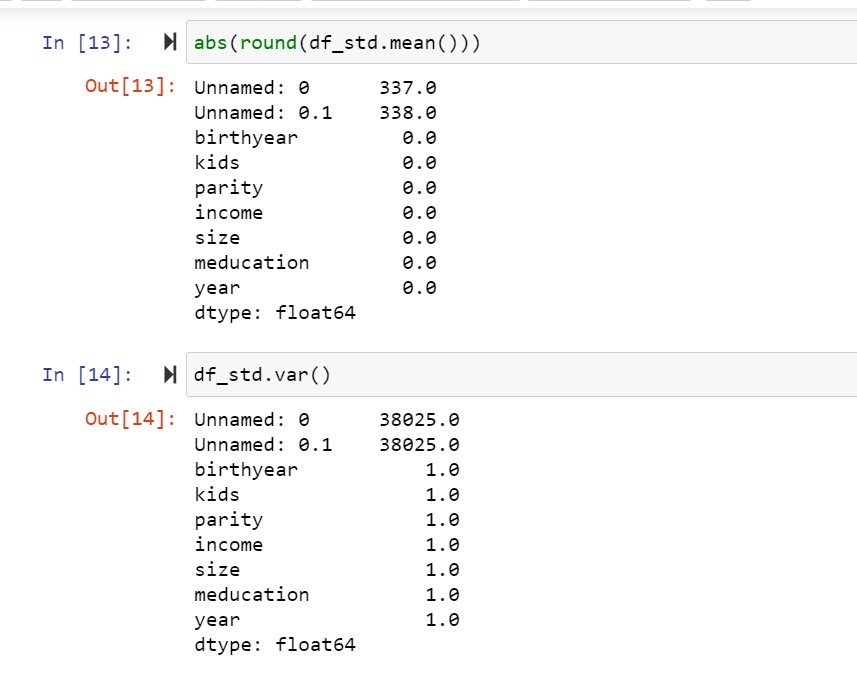
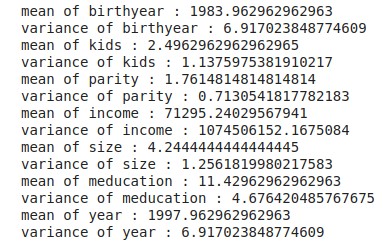
* The maximum absolute scaling
* The min-max feature scaling
* Z-score method

We have used the Z-score method in which each numerical value is replaced by Z score. Z = (x - μ) / σ



df.mean() and df.std() are the formulae used for calculating mean and standard deviation respectively.

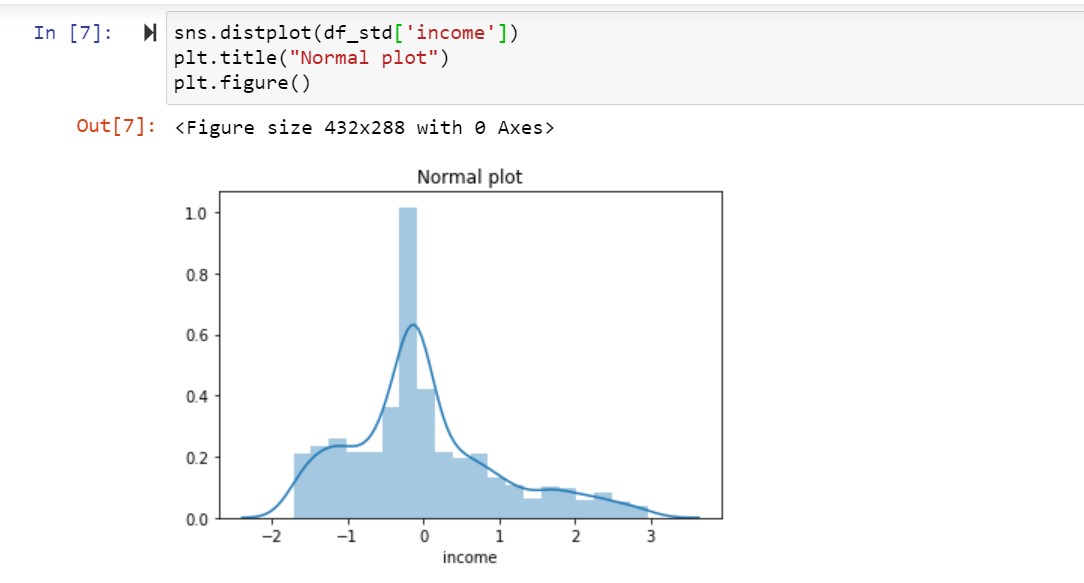
Mean and variance before Normalization. Mean and Variance after Normalization



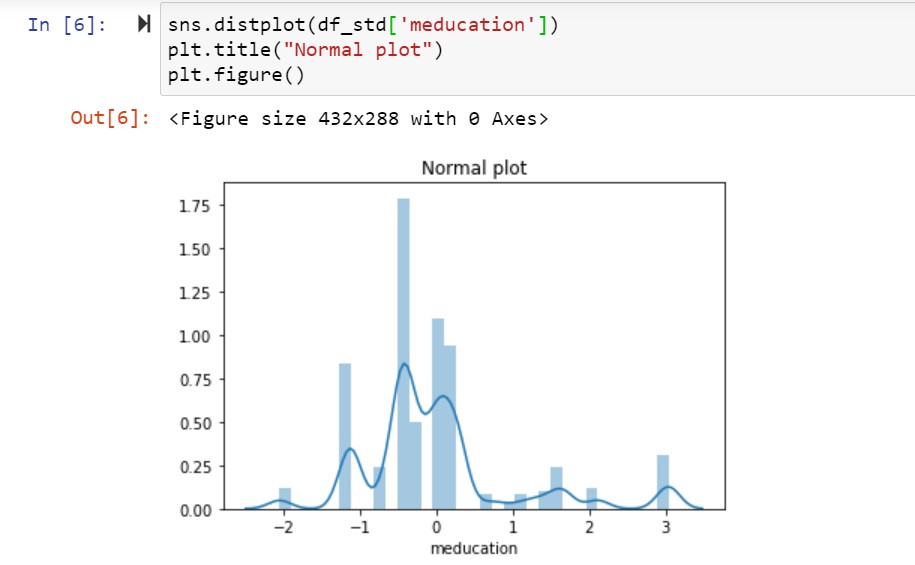
Mean and Variance after Normalization are 0 and 1 respectively.

Graphs after Normalization are as follows

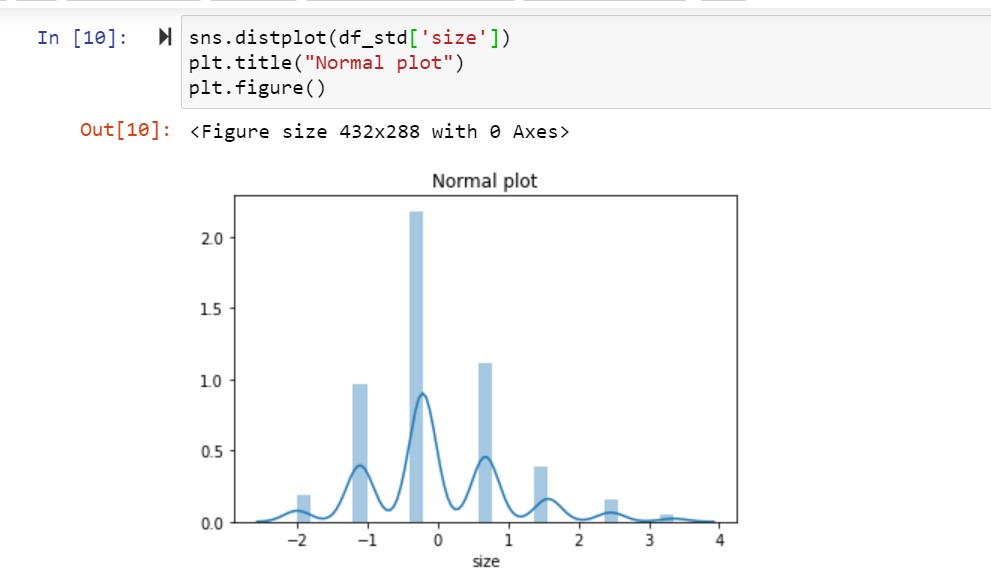
* Income



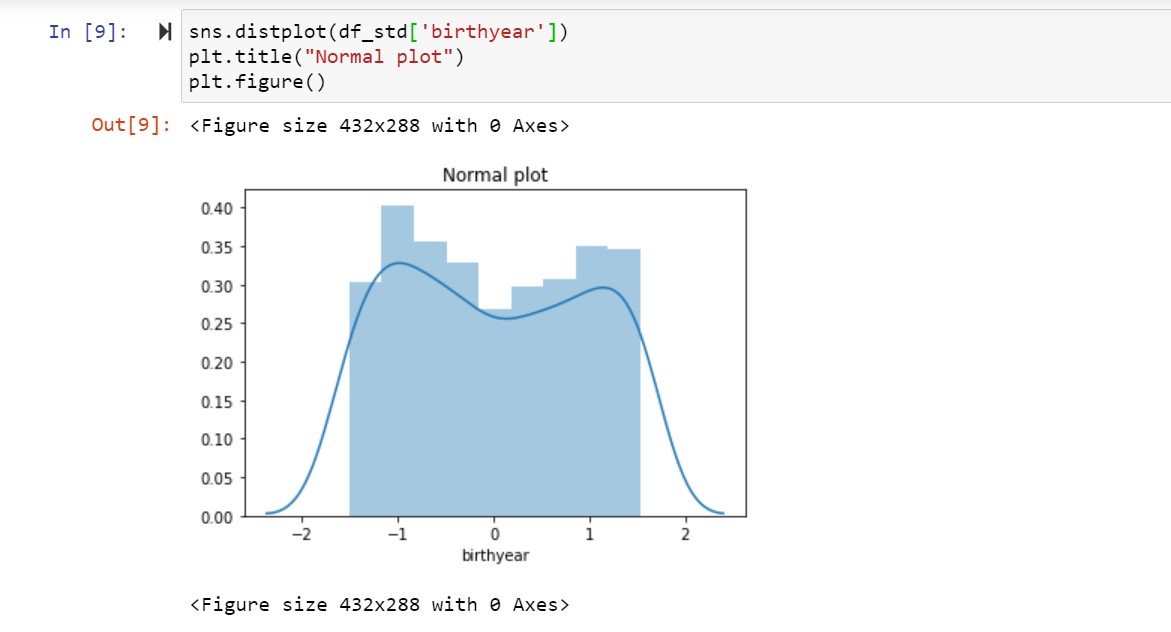
* Meducation



* Size



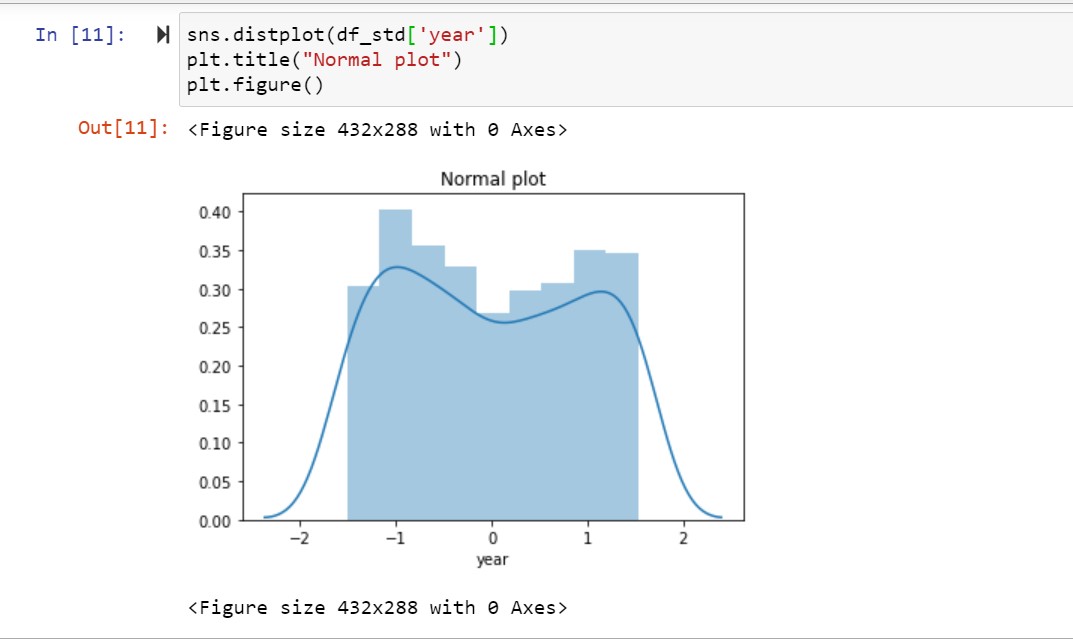
* Birthyear



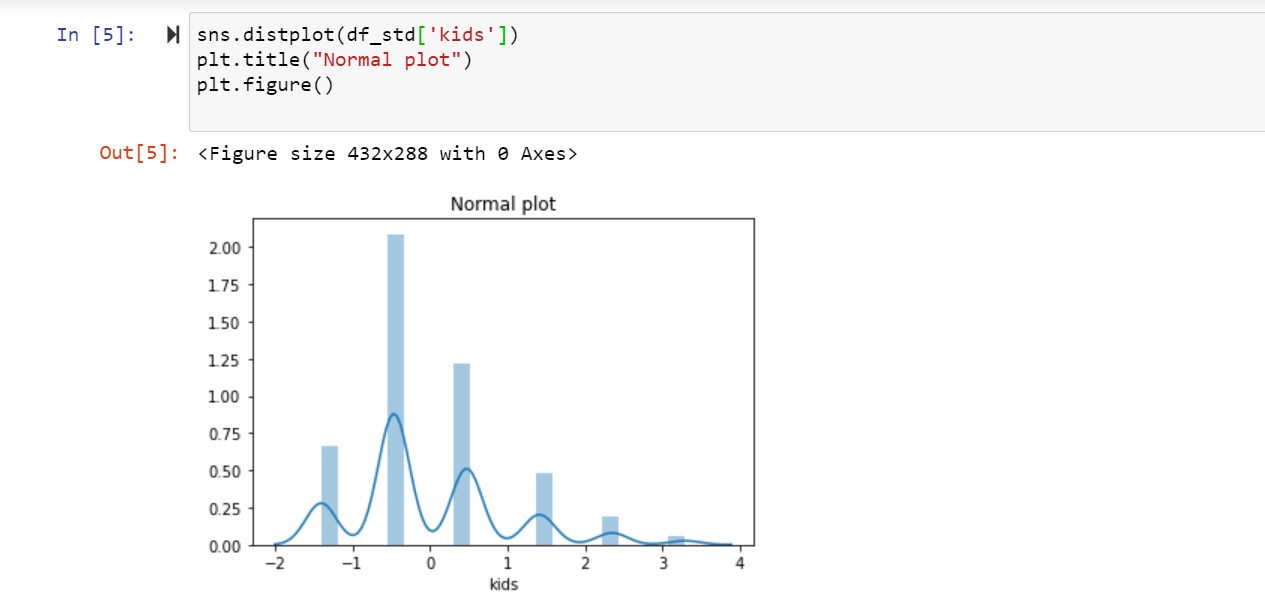
* Parity



* Year



* Kids

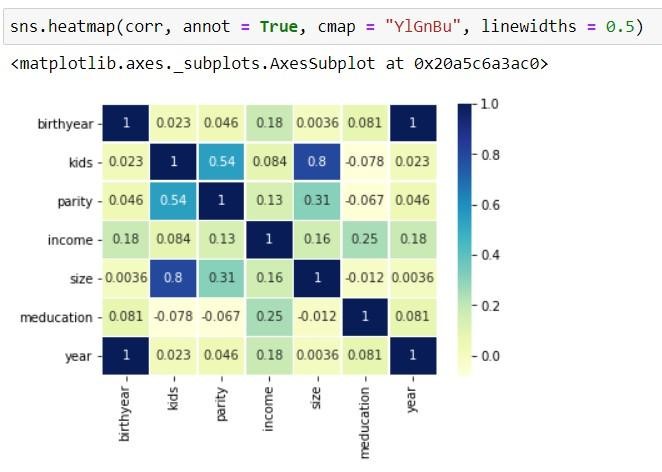


From the graphs , we can see that the mean is approximately equal to 0 and variance 1.

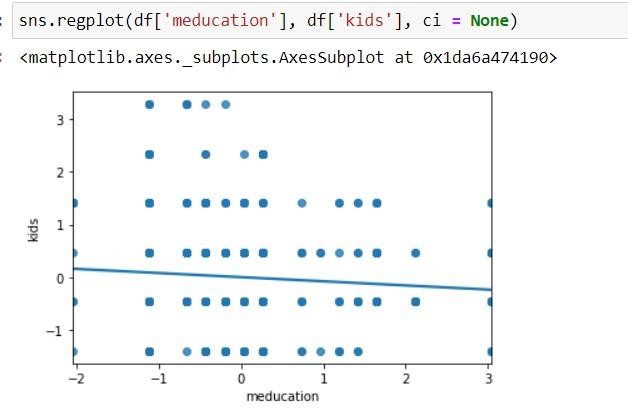
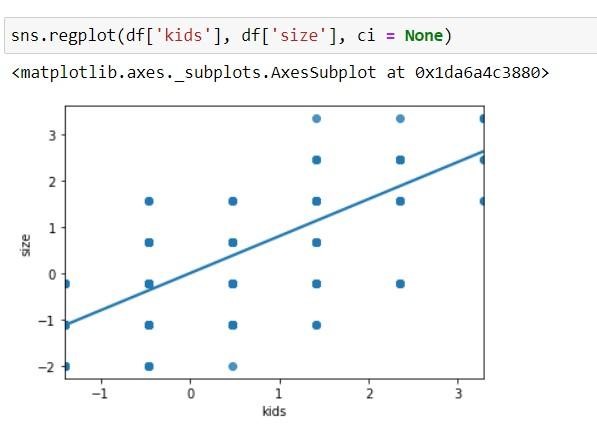
- Correlation among Variables

Correlation coefficient is used to determine the interdependence of variable quantities.

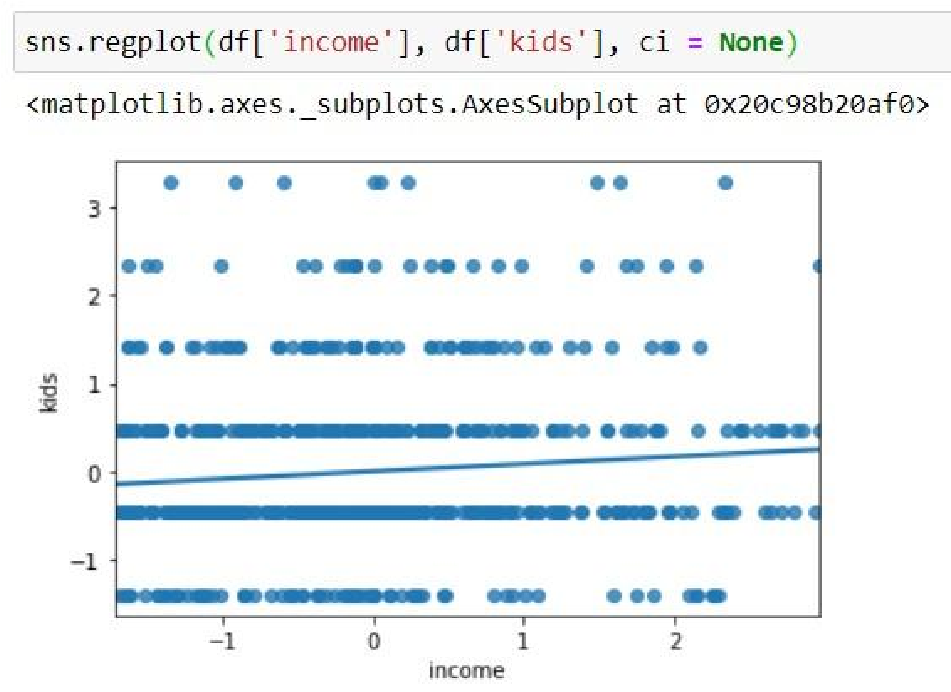
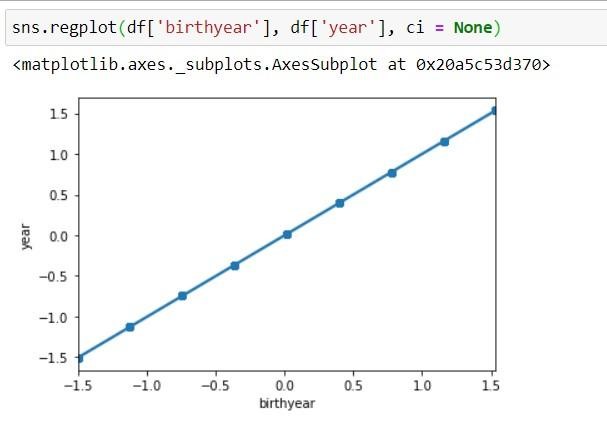
We first constructed the correlation matrix, that gives the values of the correlation coefficient for the variables.

A heatmap was then plotted for the same for better visualisation.

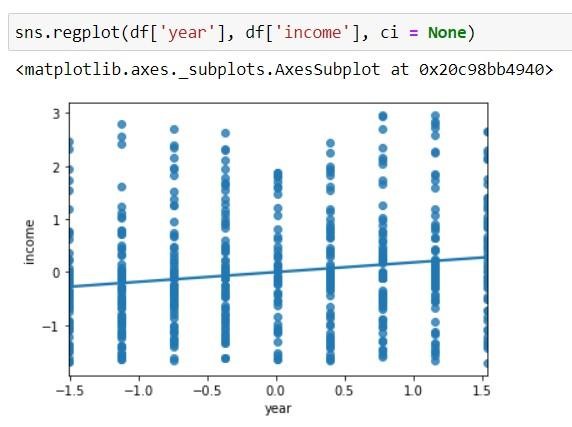
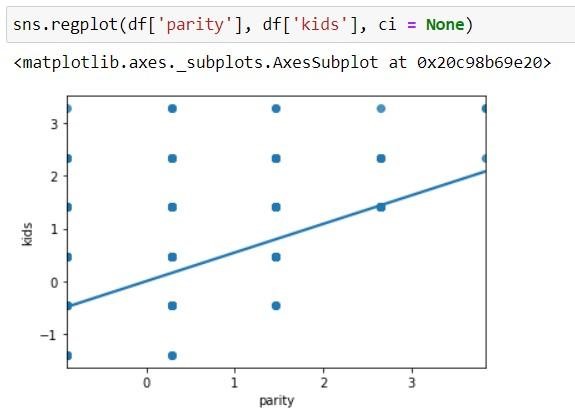
In order to confirm the findings of the heatmap and correlation matrix, scatter plots were constructed for all variables.



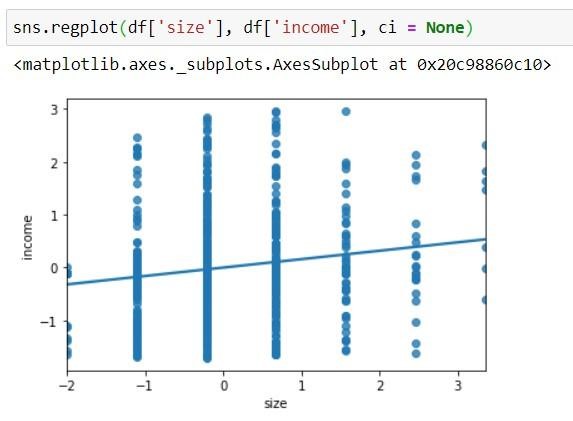
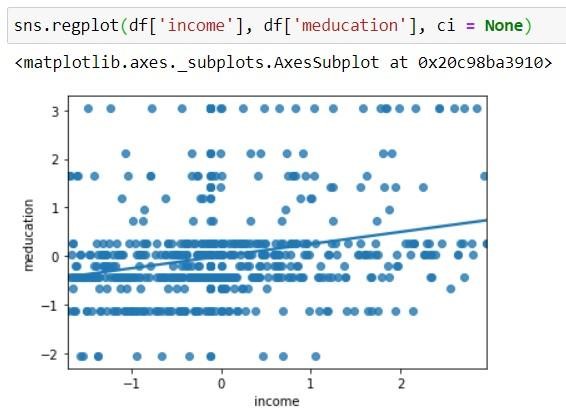
kids vs size meducation vs kids



birth year vs year income vs kids



parity vs kids year vs income



income vs meducation size vs income Inferences drawn from the scatter plot:

1. The variables year and birth year have a very strong positive correlation of 1. Since all the students under consideration were 14 year olds, we could use this data to calculate the missing values in these variables either by adding/subtracting 14 from the variables.
2. There is positive correlation between the size of the family and number of kids.
3. There is a positive correlation between parity and number of kids.
4. For the rest of the variables, the correlation coefficient is very close to zero, or in the scatter plots, a large number of data values are away from the regression line indicating that the 2 variables may be independent i.e, there is no linear relationship between them.

# Hypothesis Testing

Research hypothesis is population mean of income is less than 70,000.

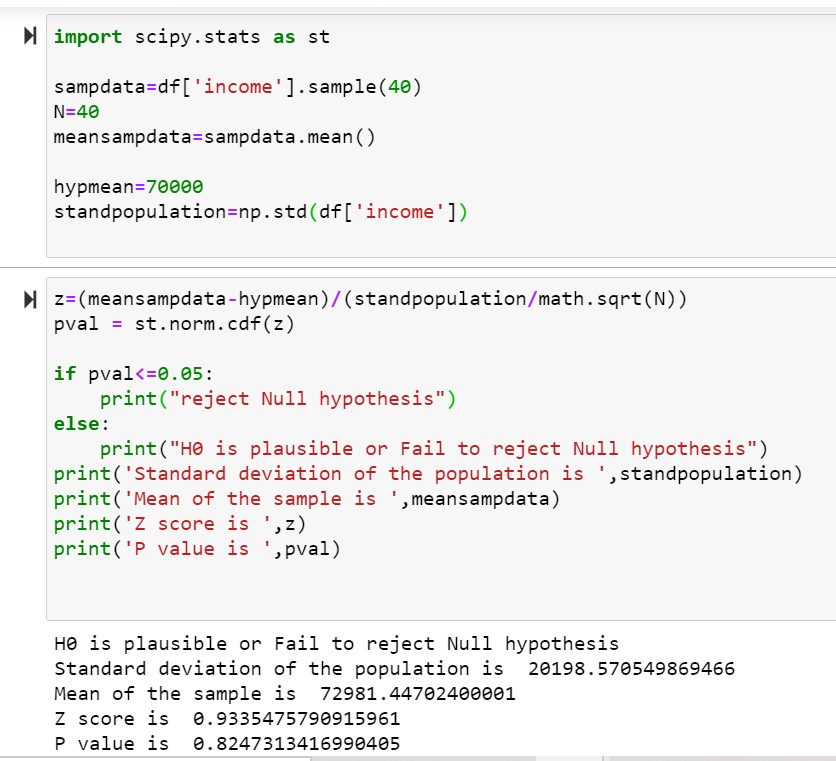
Alternate Hypothesis is

H1: μ ＜ 70,000

Null Hypothesis is

H0: μ ≥ 70,000

It is a Left-tailed test.



# LITERATURE REVIEW

##### **1.Data description**

The German Socio-Economic Panel (SOEP) is a longitudinal survey of approximately 11,000 private households in the Federal Republic of Germany from 1984 to 2019 and the eastern German länder from 1990 to 2019 (release 2021). The database is produced by the [Deutsches Institut für Wirtschaftsforschung](http://www.diw.de/en/diw_02.c.221180.en/research_data_center_soep.html) (DIW), Berlin. Variables include household composition, employment, occupation, earnings, health and satisfaction indicators.

Wave 36 of the German Socio-Economic Panel was released in April 2021. Information about the update is on this [DIW support page](https://www.diw.de/sixcms/detail.php?id=diw_01.c.814095.en).

##### **Time period**

* German SOEP Wave 36 was released in April 2021
* The Innovation Sample (I.S.) was updated in August 2021
* Coverage is from 1984 to 2019.

GSOEP

## **2 Basic features**

### 2.1 Sampling and weighting

The target population to be represented by the SOEP is Germany’s resident population. The two initial samples in 1984 include private households with a German national as household head (Sample A, n=4,528) and, oversampled, households with a Greek, Italian, Spanish, Turkish, or Yugoslavian household head (Sample B, n=1,393).

To maintain the cross-sectional representativeness in the presence of influx to the underlying target population, the enlargement samples are integrated (see Online Appendix Table A1 for an overview of all samples, and [Kroh et al. 2015](https://www.degruyter.com/document/doi/10.1515/jbnst-2018-0022/html" \l "j_jbnst-2018-0022_ref_008_w2aab3b7d293b1b6b1ab2b1b8Aa) for further details). One enlargement sample is the East German sample integrated shortly before German unification in 1990 (Sample C, n=2,179), making SOEP unique among other household panel surveys worldwide. Other enlargement samples are migrant samples, typically integrated after periods of increased gross influx. This applies to the immigration of large numbers of ethnic Germans following the collapse of the Soviet Union (Sample D, n=531, integrated in 1994/5), the immigration of EU citizens from Central and Eastern Europe after freedom of movement was implemented in the EU (Sample M1, n=2,732, integrated in 2013, and Sample M2, n=1,096, integrated in 2015), and the recent immigration of refugees from the Middle East, in particular Syria (Sample M3/4, n=3,554, integrated in 2016 and Sample M5, n=1,555, integrated in 2017). **[1]** Enlargement samples of immigrants of SOEP aim at consecutively covering all immigration years, i. e. the target population of the initial Sample B is immigration until 1983, Sample D covers subsequent immigration years 1984 to 1994, Sample M1 1995 to 2010, Sample M2 2011 to 2013, Sample M3/4 2013 to 2015, and finally Sample M5 immigration in 2016.

Panel attrition is another issue challenging representativeness and a reasonable sample size. Refreshment samples of the residential population of Germany are a means to address both issues. These refreshment samples either have the form of a general population sample or as a boost sample, the latter focusing on specific population subgroups that are the focus of the research community or policy makers. General population refreshments were integrated in 1998 (Sample E, n=1,056), in 2000 (Sample F, n=6,043), in 2006 (Sample H, n=1,506), in 2011 (Sample J, n=3,136), in 2012 (Sample K, n=1,526), and in 2017 (Sample N, n=2,314).

Boost samples focused on high-income households in 2002 (Sample G, n=1,224), families with newborn and young children in 2010 (Sample L1, n=2,074), and low income/large families/single parents in 2010/11 (Sample L2, n=2,500, Sample L3, n=924).

Two SOEP subsamples that originate from other studies have also been successfully integrated. The L-samples originate from the Families in Germany (FiD) project. FiD started in 2010 in order to enrich the available stock of data on single parents, low income families, large families with many and, in particular, young children (see [Schröder et al. 2013](https://www.degruyter.com/document/doi/10.1515/jbnst-2018-0022/html#j_jbnst-2018-0022_ref_016_w2aab3b7d293b1b6b1ab2b1c19Aa)). FiD data are the foundation underlying the evaluation of family policy measures in Germany. In 2014, FiD was fully integrated into SOEP-Core version 31 (<https://doi.org/10.5684/soep.v31>). Sample N originates from data for the Programme for the International Assessment of Adult Competencies Longitudinal (PIAAC-L). **[2]** Examining the basic skills necessary for adults to participate successfully in society and working life, PIAAC-L is the world’s first internationally comparable longitudinal study on competencies and their significance across the life course (see [Rammstedt et al. 2017](https://www.degruyter.com/document/doi/10.1515/jbnst-2018-0022/html" \l "j_jbnst-2018-0022_ref_009_w2aab3b7d293b1b6b1ab2b1c13Aa)).

## **3 Augmenting the basic SOEP data**

### 3.1 SOEP related studies

Several studies in Germany have incorporated SOEP questions to validate their results with representative sample data (“SOEP as Reference Data”, see [Siedler et al. 2009](https://www.degruyter.com/document/doi/10.1515/jbnst-2018-0022/html" \l "j_jbnst-2018-0022_ref_018_w2aab3b7d293b1b6b1ab2b1c21Aa)). Although these SOEP-Related Studies (SOEP-RS) are externally funded, they are designed and implemented in close cooperation with the SOEP team and follow the SOEP structure. This makes it possible to link the SOEP-RS datasets with either the original SOEP questionnaire (SOEP-Core) or with the SOEP-IS questionnaires, thus making it possible to analyze the data jointly. Examples of SOEP-RS studies include:

1. BASE-II – Berlin Aging Study II: BASE-II, an extension and expansion of the Berlin Aging Study (BASE), complements the analysis of cognitive development across the lifespan by including socio-economic and biological factors, such as living conditions, health, and genetic preconditions (see [Bertram et al. 2014](https://www.degruyter.com/document/doi/10.1515/jbnst-2018-0022/html#j_jbnst-2018-0022_ref_002_w2aab3b7d293b1b6b1ab2b1b1Aa)). It entails about 2,200 respondents.
2. Bonn Intervention Study (BIP) and Bremen Initiative to Foster Early Childhood Development (BRISE; <http://brise-bremen.de/wissenschaft/>): BIP and BRISE provide an intervention for children and new born babies that tries to emulate a Random Control Treatment.
3. Early childhood education and care quality in the Socio-Economic Panel (K2ID-SOEP). K2ID-SOEP has collected new data on the quality of ECEC institutions that are attended by children below school age who are sample members of SOEP (see [Spieß et al. 2018](https://www.degruyter.com/document/doi/10.1515/jbnst-2018-0022/html" \l "j_jbnst-2018-0022_ref_013_w2aab3b7d293b1b6b1ab2b1c22Aa)).

## **4 Data access and user support**

The SOEP scientific use file with anonymized microdata is made available free of charge to universities and research institutes for research and teaching purposes around the world in various data formats. Interested users must sign a user contract (<http://www.diw.de/soep-contractmanagement>) and, after approval, the data can be downloaded from the website via a secure data transfer system.

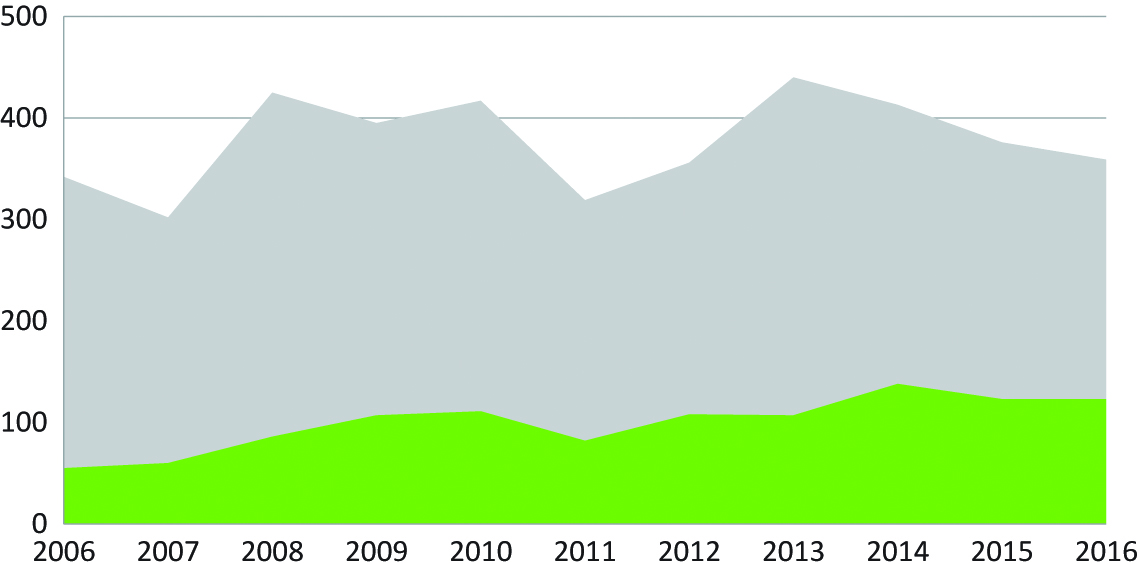
SOEP offers different forms of user support:

1. SOEP hotline (soepmail@diw.de) with instant services related to user-contracts, data distribution, as well as support for general and specialized questions on data structure and data analysis.
2. Paneldata.org (<https://paneldata.org/>) is SOEP’s documentation system that provides basic information on each variable. Item-correspondence tables indicate changes in variable names and/or value labels across time. A script-generator produces syntax files for standard software programs to combine and generate datasets.
3. SOEP-in-Residence (<https://www.diw.de/en/diw_02.c.222617.en/soep_in_residence.html>) provides SOEP users the opportunity for research stays at SOEP at DIW Berlin to discuss data matters and research projects with the SOEP team. Since 2017, European researchers can apply for visitation grants via the EU’s InGRID-2 project (<http://www.inclusivegrowth.eu/visiting-grants>).
4. SOEPcampus (<http://www.diw.de/soepcampus>) is a modular training program that familiarizes SOEP users with the data. Campus events take place at the SOEP offices in Berlin, at German universities, or as pre-conference workshops (in cooperation with other household-panels).

## **5 Science impact**

It is a challenge to describe all of the scientific contributions made possible with the SOEP in a comprehensive manner, as SOEP provides the empirical basis for researchers in disciplines as diverse as economics and sociology; political sciences and psychology; demography and gerontology; transportation, architecture and city planning; nutrition and dietetics; as well as genetics and neuro science.

As of the end of 2017, SOEP had about 3,500 users worldwide, with about 50 percent resident in Germany. As seen in [Figure 1](https://www.degruyter.com/document/doi/10.1515/jbnst-2018-0022/html#j_jbnst-2018-0022_fig_001_w2aab3b7d293b1b6b1ab1b4b4Ab2), the yearly number of SOEP-based publications amounts to between 300 and 400 annually, thereof about 25 percent in (S)SCI journals. SOEP-data is used in various internationally recognized studies, including OECD reports on the development of income inequality in OECD countries (see OECD 2008, 2011, 2015), or Education at a Glance. SOEP is also an integral database for official government reports in Germany, including the German Federal Government’s 5th Report on Poverty and Wealth (Federal Ministry of Labour and Social Affairs 2017) and the report of the German Federal Government on Wellbeing in Germany (Federal Press Office 2016).



### Figure 1:

# Results and Discussions

From the histogram we concluded that most of the students opted for Gymnasium as their secondary school choice. By comparing the number of female and male students in each school, it was observed that the number of female students is more than the male students in gymnasium, while they were almost equally distributed in the other schools. This analysis helped us understand how the popularity of schools varied through the years in Germany.

The correlation coefficients are scatter plots were used to conclude that, most of the variables were independent.The variables year and birth year have a very strong positive correlation of 1. There is positive correlation between the size of the family and number of kids. There is a positive correlation between parity and number of kids. The hypothesis test concludes, it is plausible that the mean income of the population may be greater than or equal to 70000.

### References And Links :

Bertram, L., A. Böckenhoff, I. Demuth, S. Düzel, R. Eckardt, S.-C. Li, U. Lindenberger, G. Pawelec, T. Siedler, G.G. Wagner, E. Steinhagen-Thiessen (2014), Cohort Profile: The Berlin Aging Study II (BASE-II). International Journal of Epidemiology 43: 703–712. [Search in Google Scholar](https://scholar.google.com/scholar?q=Bertram,%20L.,%20A.%20B%C3%B6ckenhoff,%20I.%20Demuth,%20S.%20D%C3%BCzel,%20R.%20Eckardt,%20S.-C.%20Li,%20U.%20Lindenberger,%20G.%20Pawelec,%20T.%20Siedler,%20G.G.%20Wagner,%20E.%20Steinhagen-Thiessen%20(2014),%20Cohort%20Profile:%20The%20Berlin%20Aging%20Study%20II%20(BASE-II).%20International%20Journal%20of%20Epidemiology%2043:%20703%E2%80%93712.%20Bertram%20L.%20B%C3%B6ckenhoff%20A.%20Demuth%20I.%20D%C3%BCzel%20S.%20Eckardt%20R.%20Li%20S.-C.%20Lindenberger%20U.%20Pawelec%20G.%20Siedler%20T.%20Wagner%20G.G.%20Steinhagen-Thiessen%20E.%202014%20Cohort%20Profile:%20The%20Berlin%20Aging%20Study%20II%20(BASE-II)%20International%20Journal%20of%20Epidemiology%2043%20703%20712)

Deaton, A. (2015), Nobel Prize Lecture by Angus Deaton. Available under:<http://www.nobelprize.org/mediaplayer/index.php?id=2585> [Search in Google Scholar](https://scholar.google.com/scholar?q=Deaton,%20A.%20(2015),%20Nobel%20Prize%20Lecture%20by%20Angus%20Deaton.%20Available%20under:%20Deaton%20A.%202015%20Nobel%20Prize%20Lecture%20by%20Angus%20Deaton%20Available%20under:)

Eisnecker, Ph., K. Erhardt, M. Kroh, P. Trübswetter (2017), The Request for Record Linkage in the IAB -SOEP Migration Sample. SOEP Survey Papers 291, C. [Search in Google Scholar](https://scholar.google.com/scholar?q=Eisnecker,%20Ph.,%20K.%20Erhardt,%20M.%20Kroh,%20P.%20Tr%C3%BCbswetter%20(2017),%20The%20Request%20for%20Record%20Linkage%20in%20the%20IAB%20-SOEP%20Migration%20Sample.%20SOEP%20Survey%20Papers%20291,%20C.%20Eisnecker,%20Ph.%20K.%20Erhardt,%20M.%20Kroh,%20P.%20Tr%C3%BCbswetter%202017%20The%20Request%20for%20Record%20Linkage%20in%20the%20IAB%20-SOEP%20Migration%20Sample.%20SOEP%20Survey%20Papers%20291,%20C)

Federal Ministry of Labour and Social Affairs (2017), The German Federal Government’s 5th Report on Poverty and Wealth. Bonn. [Search in Google Scholar](https://scholar.google.com/scholar?q=Federal%20Ministry%20of%20Labour%20and%20Social%20Affairs%20(2017),%20The%20German%20Federal%20Government%E2%80%99s%205th%20Report%20on%20Poverty%20and%20Wealth.%20Bonn.%20Federal%20Ministry%20of%20Labour%20and%20Social%20Affairs%202017%20The%20German%20Federal%20Government%E2%80%99s%205th%20Report%20on%20Poverty%20and%20Wealth.%20Bonn)

Federal Press Office (2016), Government Report on wellbeing in Germany. Berlin. [Search in Google Scholar](https://scholar.google.com/scholar?q=Federal%20Press%20Office%20(2016),%20Government%20Report%20on%20wellbeing%20in%20Germany.%20Berlin.%20Federal%20Press%20Office%202016%20Government%20Report%20on%20wellbeing%20in%20Germany.%20Berlin.)

Goebel, J., B. Pauer (2014), Datenschutzkonzept zur Nutzung von SOEPgeo im Forschungsdatenzentrum SOEP am DIW Berlin. Zeitschrift für amtliche Statistik Berlin-Brandenburg 3: 42–47. [Search in Google Scholar](https://scholar.google.com/scholar?q=Goebel,%20J.,%20B.%20Pauer%20(2014),%20Datenschutzkonzept%20zur%20Nutzung%20von%20SOEPgeo%20im%20Forschungsdatenzentrum%20SOEP%20am%20DIW%20Berlin.%20Zeitschrift%20f%C3%BCr%20amtliche%20Statistik%20Berlin-Brandenburg%203:%2042%E2%80%9347.%20Goebel%20J.%20Pauer%20B.%202014%20Datenschutzkonzept%20zur%20Nutzung%20von%20SOEPgeo%20im%20Forschungsdatenzentrum%20SOEP%20am%20DIW%20Berlin%20Zeitschrift%20f%C3%BCr%20amtliche%20Statistik%20Berlin-Brandenburg%203%2042%2047)

Kroh, M., R. Pischner, M. Spiess, G. Wagner (2008), On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP): Tracing, Weighting, and Frequencies. Methoden, Daten, Analysen. Zeitschrift Für Empirische Sozialforschung 2: 179–198. [Search in Google Scholar](https://scholar.google.com/scholar?q=Kroh,%20M.,%20R.%20Pischner,%20M.%20Spiess,%20G.%20Wagner%20(2008),%20On%20the%20Treatment%20of%20Non-Original%20Sample%20Members%20in%20the%20German%20Household%20Panel%20Study%20(SOEP):%20Tracing,%20Weighting,%20and%20Frequencies.%20Methoden,%20Daten,%20Analysen.%20Zeitschrift%20F%C3%BCr%20Empirische%20Sozialforschung%202:%20179%E2%80%93198.%20Kroh%20M.%20Pischner%20R.%20Spiess%20M.%20Wagner%20G.%202008%20On%20the%20Treatment%20of%20Non-Original%20Sample%20Members%20in%20the%20German%20Household%20Panel%20Study%20(SOEP):%20Tracing,%20Weighting,%20and%20Frequencies.%20Methoden,%20Daten,%20Analysen%20Zeitschrift%20F%C3%BCr%20Empirische%20Sozialforschung%202%20179%20198)

Kroh, M., R. Siegers, S. Kühne (2015), Gewichtung und Integration von Auffrischungsstichproben am Beispiel des Sozio-oekonomischen Panels (SOEP) in: J. Schupp, C. Wolf (Hrsg.), Nonresponse Bias: Qualitätssicherung sozialwissenschaftlicher Umfragen. Wiesbaden, Springer VS Verlag. [Search in Google Scholar](https://scholar.google.com/scholar?q=Kroh,%20M.,%20R.%20Siegers,%20S.%20K%C3%BChne%20(2015),%20Gewichtung%20und%20Integration%20von%20Auffrischungsstichproben%20am%20Beispiel%20des%20Sozio-oekonomischen%20Panels%20(SOEP)%20in:%20J.%20Schupp,%20C.%20Wolf%20(Hrsg.),%20Nonresponse%20Bias:%20Qualit%C3%A4tssicherung%20sozialwissenschaftlicher%20Umfragen.%20Wiesbaden,%20Springer%20VS%20Verlag.%20Kroh%20M.%20Siegers%20R.%20K%C3%BChne%20S.%202015%20Gewichtung%20und%20Integration%20von%20Auffrischungsstichproben%20am%20Beispiel%20des%20Sozio-oekonomischen%20Panels%20(SOEP)%20Schupp%20J.%20Wolf%20(Hrsg.)%20C.%20Nonresponse%20Bias:%20Qualit%C3%A4tssicherung%20sozialwissenschaftlicher%20Umfragen%20Wiesbaden%20Springer%20VS%20Verlag)

Kroh, M., S. Kühne, R. Siegers, V. Belcheva (2018), SOEP-Core – Documentation of Sample Sizes and Panel Attrition (1984 until 2016). SOEP Survey Paper 480, Series C. [Search in Google Scholar](https://scholar.google.com/scholar?q=Kroh,%20M.,%20S.%20K%C3%BChne,%20R.%20Siegers,%20V.%20Belcheva%20(2018),%20SOEP-Core%20%E2%80%93%20Documentation%20of%20Sample%20Sizes%20and%20Panel%20Attrition%20(1984%20until%202016).%20SOEP%20Survey%20Paper%20480,%20Series%20C.%20Kroh%20M.%20K%C3%BChne%20R%20Siegers%20S.%20Belcheva%20V.%202018%20SOEP-Core%20%E2%80%93%20Documentation%20of%20Sample%20Sizes%20and%20Panel%20Attrition%20(1984%20until%202016).%20SOEP%20Survey%20Paper%20480,%20Series%20C)

OECD (2008), Growing unequal? Income distribution and poverty in OECD countries. Paris. [Search in Google Scholar](https://scholar.google.com/scholar?q=OECD%20(2008),%20Growing%20unequal?%20Income%20distribution%20and%20poverty%20in%20OECD%20countries.%20Paris.%20OECD%202008%20Growing%20unequal?%20Income%20distribution%20and%20poverty%20in%20OECD%20countries.%20Paris)

OECD (2011), Divided We Stand: Why Inequality Keeps Rising. Paris. [Search in Google Scholar](https://scholar.google.com/scholar?q=OECD%20(2011),%20Divided%20We%20Stand:%20Why%20Inequality%20Keeps%20Rising.%20Paris.%20OECD%202011%20Divided%20We%20Stand:%20Why%20Inequality%20Keeps%20Rising.%20Paris)

OECD (2015), In It Together: Why Less Inequality Benefits All. Paris. [Search in Google Scholar](https://scholar.google.com/scholar?q=OECD%20(2015),%20In%20It%20Together:%20Why%20Less%20Inequality%20Benefits%20All.%20Paris.%20OECD%202015%20In%20It%20Together:%20Why%20Less%20Inequality%20Benefits%20All.%20Paris)

Rammstedt, B., S. Martin, A. Zabal, C. Carstensen, J. Schupp (2017), The PIAAC Longitudinal Study in Germany: rationale and design. Large-Scale Assessments in Education 5: 4. [Search in Google Scholar](https://scholar.google.com/scholar?q=Rammstedt,%20B.,%20S.%20Martin,%20A.%20Zabal,%20C.%20Carstensen,%20J.%20Schupp%20(2017),%20The%20PIAAC%20Longitudinal%20Study%20in%20Germany:%20rationale%20and%20design.%20Large-Scale%20Assessments%20in%20Education%205:%204.%20Rammstedt%20B.%20Martin%20S.%20Zabal%20A.%20Carstensen%20C.%20Schupp%20J.%202017%20The%20PIAAC%20Longitudinal%20Study%20in%20Germany:%20rationale%20and%20design%20Large-Scale%20Assessments%20in%20Education%205%204)

Rendtel, U. (1995), Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität. Campus Verlag, Frankfurt/New York. [Search in Google Scholar](https://scholar.google.com/scholar?q=Rendtel,%20U.%20(1995),%20Lebenslagen%20im%20Wandel:%20Panelausf%C3%A4lle%20und%20Panelrepr%C3%A4sentativit%C3%A4t.%20Campus%20Verlag,%20Frankfurt/New%20York.%20Rendtel%20U.%201995%20Lebenslagen%20im%20Wandel:%20Panelausf%C3%A4lle%20und%20Panelrepr%C3%A4sentativit%C3%A4t%20Campus%20Verlag,%20Frankfurt/New%20York)

Richter, D., J. Rohrer, M. Metzing, W. Nestler, M. Weinhardt, J. Schupp (2017), SOEP Scales Manual. SOEP Survey Papers 423, Series C. [Search in Google Scholar](https://scholar.google.com/scholar?q=Richter,%20D.,%20J.%20Rohrer,%20M.%20Metzing,%20W.%20Nestler,%20M.%20Weinhardt,%20J.%20Schupp%20(2017),%20SOEP%20Scales%20Manual.%20SOEP%20Survey%20Papers%20423,%20Series%20C.%20Richter%20D.%20Rohrer%20J.%20Metzing%20M.%20Nestler%20W.%20Weinhardt%20M.%20Schupp%20J.%202017%20SOEP%20Scales%20Manual%20SOEP%20Survey%20Papers%20423,%20Series%20C)

Richter, D., J. Schupp (2015), The SOEP Innovation Sample (SOEP IS). Schmollers Jahrbuch 135(3): 389–399. [Search in Google Scholar](https://scholar.google.com/scholar?q=Richter,%20D.,%20J.%20Schupp%20(2015),%20The%20SOEP%20Innovation%20Sample%20(SOEP%20IS).%20Schmollers%20Jahrbuch%20135(3):%20389%E2%80%93399.%20Richter%20D.%20Schupp%20J.%202015%20The%20SOEP%20Innovation%20Sample%20(SOEP%20IS)%20Schmollers%20Jahrbuch%20135%203%20389%20399)

Schonlau, M., M. Kroh, N. Watson (2013), The Implementation of Cross-Sectional Weights in Household Panel Surveys. Statistics Surveys 7: 37–57. [Search in Google Scholar](https://scholar.google.com/scholar?q=Schonlau,%20M.,%20M.%20Kroh,%20N.%20Watson%20(2013),%20The%20Implementation%20of%20Cross-Sectional%20Weights%20in%20Household%20Panel%20Surveys.%20Statistics%20Surveys%207:%2037%E2%80%9357.%20Schonlau%20M.%20Kroh%20M.%20Watson%20N.%202013%20The%20Implementation%20of%20Cross-Sectional%20Weights%20in%20Household%20Panel%20Surveys%20Statistics%20Surveys%207%2037%2057)

Schonlau, M., N. Watson, M. Kroh (2011), Household Survey Panels: How Much Do following Rules Affect Sample Size? Survey Research Methods 5: 53–61. [Search in Google Scholar](https://scholar.google.com/scholar?q=Schonlau,%20M.,%20N.%20Watson,%20M.%20Kroh%20(2011),%20Household%20Survey%20Panels:%20How%20Much%20Do%20following%20Rules%20Affect%20Sample%20Size?%20Survey%20Research%20Methods%205:%2053%E2%80%9361.%20Schonlau%20M.%20Watson%20N.%20Kroh%20M.%202011%20Household%20Survey%20Panels:%20How%20Much%20Do%20following%20Rules%20Affect%20Sample%20Size?%20Survey%20Research%20Methods%205%2053%2061)

Schröder, M., R. Siegers, C.K. Spieß (2013), “Familien in Deutschland” – FiD. Schmollers Jahrbuch 133(4): 595–606. [Search in Google Scholar](https://scholar.google.com/scholar?q=Schr%C3%B6der,%20M.,%20R.%20Siegers,%20C.K.%20Spie%C3%9F%20(2013),%20%E2%80%9CFamilien%20in%20Deutschland%E2%80%9D%20%E2%80%93%20FiD.%20Schmollers%20Jahrbuch%20133(4):%20595%E2%80%93606.%20Schr%C3%B6der%20M.%20Siegers%20R.%20Spie%C3%9F%20C.K.%202013%20Familien%20in%20Deutschland%22%20%E2%80%93%20FiD.%20Schmollers%20Jahrbuch%20133(4):%20595%E2%80%93606)

Schupp, J. (2015), Forty Years of Social Reporting and Quality of Life Reserach in Germany. 107–126 in: G. Trommsdorff, W.R. Assmann (Eds.), A Look Back and Prospects for the Future, Forschung fördern. Am Beispiel von Lebensqualität im Kulturkontext. Konstanz und München, UVK. [Search in Google Scholar](https://scholar.google.com/scholar?q=Schupp,%20J.%20(2015),%20Forty%20Years%20of%20Social%20Reporting%20and%20Quality%20of%20Life%20Reserach%20in%20Germany.%20107%E2%80%93126%20in:%20G.%20Trommsdorff,%20W.R.%20Assmann%20(Eds.),%20A%20Look%20Back%20and%20Prospects%20for%20the%20Future,%20Forschung%20f%C3%B6rdern.%20Am%20Beispiel%20von%20Lebensqualit%C3%A4t%20im%20Kulturkontext.%20Konstanz%20und%20M%C3%BCnchen,%20UVK.%20Schupp%20J.%202015%20Forty%20Years%20of%20Social%20Reporting%20and%20Quality%20of%20Life%20Reserach%20in%20Germany%20107%20126%20Trommsdorff%20G.%20Assmann%20W.R.%20A%20Look%20Back%20and%20Prospects%20for%20the%20Future%20Forschung%20f%C3%B6rdern.%20Am%20Beispiel%20von%20Lebensqualit%C3%A4t%20im%20Kulturkontext%20Konstanz%20und%20M%C3%BCnchen%20UVK)

Siedler, T., J. Schupp, C.K. Spieß, G.G. Wagner (2009), The German Socio-Economic Panel (SOEP) as Reference Data Set. Schmollers Jahrbuch 129(2): 367–374. [Search in Google Scholar](https://scholar.google.com/scholar?q=Siedler,%20T.,%20J.%20Schupp,%20C.K.%20Spie%C3%9F,%20G.G.%20Wagner%20(2009),%20The%20German%20Socio-Economic%20Panel%20(SOEP)%20as%20Reference%20Data%20Set.%20Schmollers%20Jahrbuch%20129(2):%20367%E2%80%93374.%20Siedler%20T.%20Schupp%20J.%20Spie%C3%9F%20C.K.%20Wagner%20G.G.%202009%20The%20German%20Socio-Economic%20Panel%20(SOEP)%20as%20Reference%20Data%20Set%20Schmollers%20Jahrbuch%20129%202%20367%20374)

Spieß, C.K., Schober, P.S., Stahl, J.F. (2018), Early Childhood Education and Care Quality in the Socio-Economic Panel (SOEP) – the K2ID-SOEP Study. Journal of Economics and Statistics (online first), doi: <https://doi.org/10.1515/jbnst-2018-0001>. [Search in Google Scholar](https://scholar.google.com/scholar?q=Spie%C3%9F,%20C.K.,%20Schober,%20P.S.,%20Stahl,%20J.F.%20(2018),%20Early%20Childhood%20Education%20and%20Care%20Quality%20in%20the%20Socio-Economic%20Panel%20(SOEP)%20%E2%80%93%20the%20K2ID-SOEP%20Study.%20Journal%20of%20Economics%20and%20Statistics%20(online%20first),%20doi:%20.%20Spie%C3%9F%20C.K.%20Schober%20P.S.%20Stahl%20J.F.%20Zoch%20G.%20Camehl%20G.F.%202018%20Early%20Childhood%20Education%20and%20Care%20Quality%20in%20the%20Socio-Economic%20Panel%20(SOEP)%20%E2%80%93%20the%20K2ID-SOEP%20Study.%20Journal%20of%20Economics%20and%20Statistics%20(online%20first),%20doi:)

Wagner, G.G., J.R. Frick, J. Schupp (2007), The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements, Schmollers Jahrbuch. Journal of Applied Social Science Studies 127(1): 139–169. [Search in Google Schola](https://scholar.google.com/scholar?q=Wagner,%20G.G.,%20J.R.%20Frick,%20J.%20Schupp%20(2007),%20The%20German%20Socio-Economic%20Panel%20Study%20(SOEP)%20%E2%80%93%20Scope,%20Evolution%20and%20Enhancements,%20Schmollers%20Jahrbuch.%20Journal%20of%20Applied%20Social%20Science%20Studies%20127(1):%20139%E2%80%93169.%20Wagner%20G.G.%20Frick%20J.R.%20Schupp%20J.%202007%20The%20German%20Socio-Economic%20Panel%20Study%20(SOEP)%20%E2%80%93%20Scope,%20Evolution%20and%20Enhancements,%20Schmollers%20Jahrbuch%20Journal%20of%20Applied%20Social%20Science%20Studies%20127%201%20139%20169)