

ECON4225 Homework 1

Anna Duan

2025-10-06

Part 1: Overview of the Dataset

Question 1: household and variable count

This dataset has 9066 households and 38 variables, including the household ID.

```
## [1] "rows: 9066"

## [1] "ID" "AGE_HEAD" "AGE_SPOUSE" "SEX_HEAD"
## [5] "SEX_SPOUSE" "EDU_HEAD" "EDU_SPOUSE" "OCC_HEAD"
## [9] "OCC_SPOUSE" "STATE" "MARITAL" "HOUSEHOLD_SIZE"
## [13] "CHILDREN" "SPOUSE_PRESENT" "WEIGHT" "INCOME"
## [17] "HEAD_LABOR" "SPOUSE_LABOR" "WEEKS_HEAD" "WEEKS_SPOUSE"
## [21] "WEEKS_OUT_HEAD" "WEEKS_OUT_SPOUSE" "EXP" "GAS"
## [25] "FOOD_HOME" "FOOD_DELIV" "FOOD_OUT" "RENT"
## [29] "EDUC_EXP" "CHILDCARE" "TRIPS" "WEALTH"
## [33] "HOUSE_VALUE" "MORTGAGE" "MORT_PAY" "HOME_EQUITY"
## [37] "BUSINESS_VAL" "PROPERTY_TAX"
```

Question 2

The average age of the household head is 46.25. The highest age is 102 and the lowest age is 18.

```
## [1] "Mean age: 46.25"
## [1] "Highest age: 102"
## [1] "Lowest age: 18"
```

Question 3: Income

The average household income is \$78,265.69. This is substantially lower than the \$142k reported by the SCF in 2022 (Section 1.2, slide 9). One possible reason for this difference is that the SCF surveys a small sample, focusing on high-income individuals, resulting in a less representative mean. The PSID has a larger sample, with less emphasis on the top of the income distribution.

```
## [1] "Mean household income: 78265.69"
```

Question 4: income range

The household incomes in the dataset have a wide range: the lowest is -\$267,900 and the highest is \$2,125,100. The households with negative income are likely in debt.

```
## [1] "Lowest household income: -267900"
## [1] "Highest household income: 2125100"
```

Question 5: spouse age

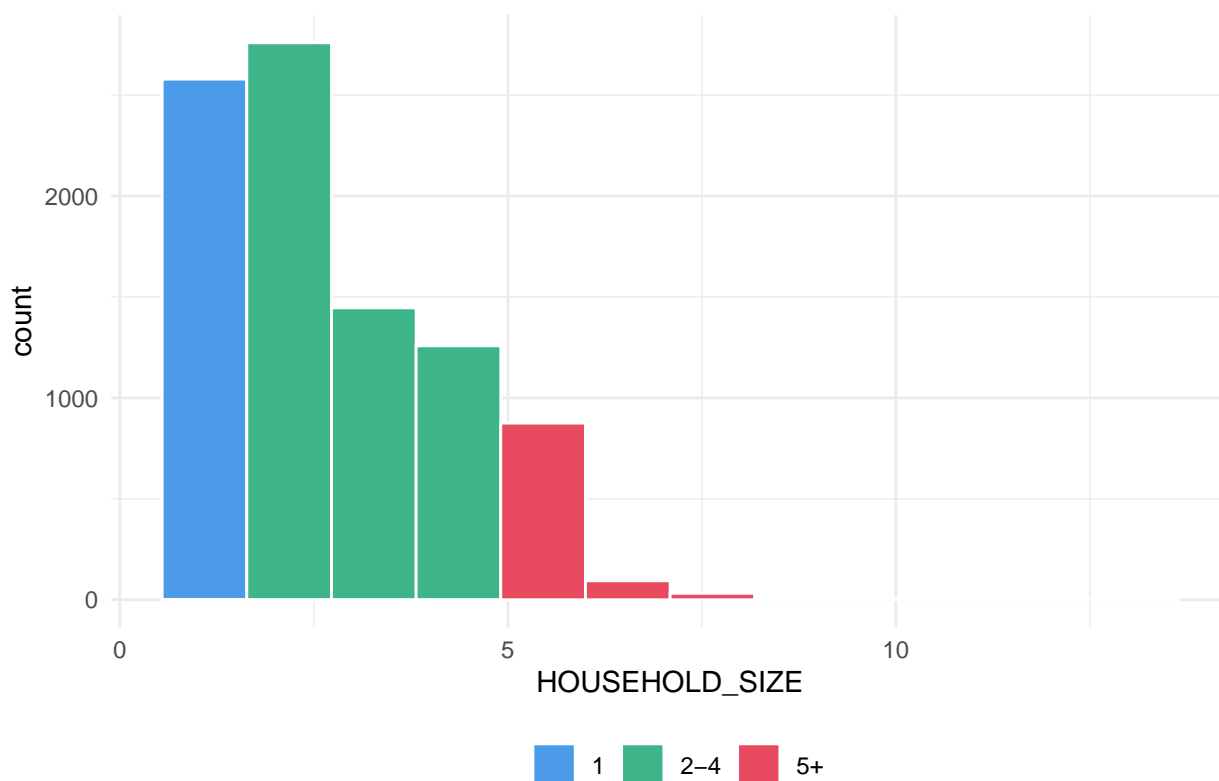
Of the 4,523 households with a spouse of the household head present, the average age of the spouse is 45.64 years.

```
## [1] "Mean spouse age: 45.64"
```

Question 6: household members

Households with only 1 member make up 28.45% of the data. Households with 5 or more members make up

Distribution of Household Size in the PSID



11.29%.

```
psid_hh_size %>%  
  group_by(size_group) %>%  
  summarise(count = n(), .groups = "drop") %>%  
  mutate(pct = 100*round(count/sum(count),4)) %>%  
  pander()
```

size_group	count	pct
1	2579	28.45
2-4	5463	60.26
5+	1024	11.29

Part 2: Income Distribution

Question 1

- Histogram of household income with bins of \$5k
- Description of histogram findings

Question 2

- Lorenz Curve of household income
- Description of lorenz curve findings

Question 3

- Coefficient of variation
- Compare to SCF

Question 4

- Income adjusted for number of household heads (filter only for households with spouses, divide income by 2)
- Lorenz curve of adjusted income alongside regular income
- Conclusions from comparison

Question 5

- 30th, 50th, 90th, 99th percentile of income
- 90-30, 90-50, 30-10, 99-50 ratios
- Interpretation of ratios
- Comparison with SCF

Question 6

- Share of income by quintile
- Average income by quintile
- Interpretation of results
- Comparison with SCF (lecture)

Question 7

- Average income and income share for top 1%
- Comparison with SCF

Part 3: Labor Income

Question 1

- Create variable of total labor income of household
- Earnings and share of earnings for each quintile

Question 2

- Compare distribution of household earnings vs income
- Compare share of quintiles
- Coefficient of variation
- Lorenz curve of labor earnings and total household income
- Compare relative inequality in labor earnings vs total household income

Question 3

- Create variable of share of labor earnings in total household income
- Average share of labor earnings

Question 4

- Average share of labor earnings by quintile of income distribution
- Share of labor earnings for top 1%
- Compare both to SCF
- Comment on differences

Question 5

- Filter by households where household head has positive work weeks and positive earnings
- Weekly wage of household head
- Variance of log of earnings, wages, weeks worked
- Relative contribution of hours and wages to inequality in labor earnings?
- Caveats to this calculation?

Question 6

- Regression of log wages of household head on age, age^2 , education, occupation
- Residuals
- Variance of residuals
- Share of inequality explained by age, education and occupation?