

# 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: age of household head

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: average household 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 values likely incurred business or other losses in the survey year, resulting in negative net income.

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

## Question 5: age of household head's spouse

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 size

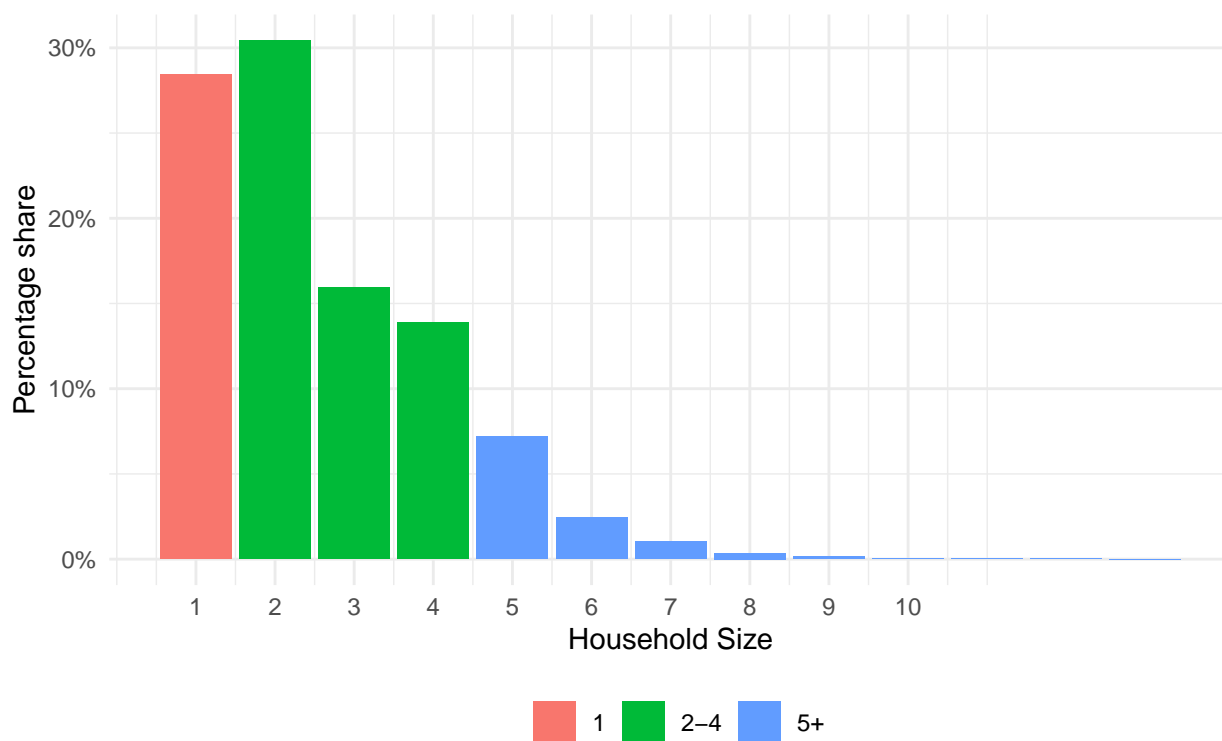
Households with only 1 member make up 28.45% of the data. Households with 5 or more members make up 11.29%. Compared to the histogram shown in class, the PSID has fewer 2-member households and more 3-5 member households. In both datasets, about 28% of households have one member. The PSID dataset has a smaller share of two-member households: only 30.43% compared to the nearly 35% shown on slide 2 [what dataset is this?]. PSID has a slightly larger share of households with 3 members: 15.95%, compared to the slide's 15%; a 13.88% share of 4-member households, compared to the slide's 12.5%; and a 7.2% share of 5-member households, compared to just past 5% in the slide.

Table 1: Percentage share of PSID households by size

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

### Most households have fewer than 5 members

Distribution of households by number of members, PSID



## Part 2: Income Distribution

### Question 1: distribution of household income

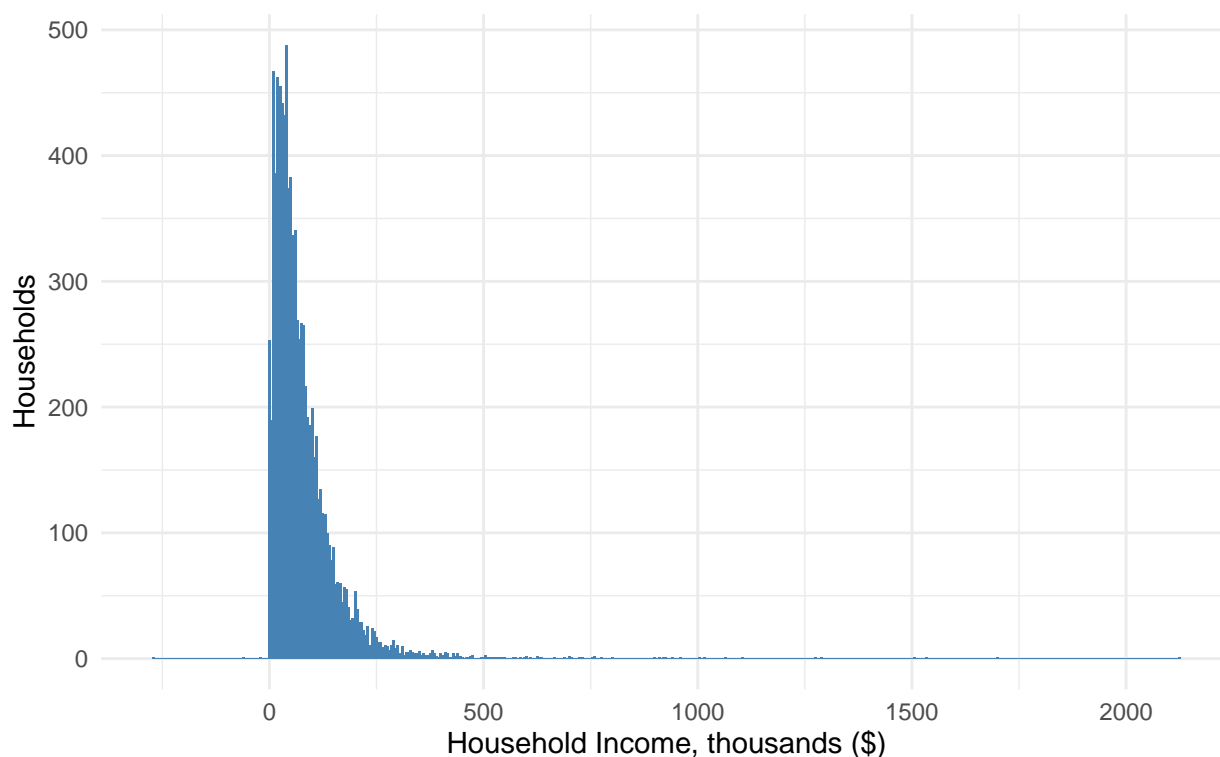
The following histogram shows the distribution of household incomes in the PSID, with values above the 99th percentile removed. The distribution is right-skewed, as the median of \$55,090 is lower than the mean of \$78,266. The majority of households earn less than \$100,682, which is the 75th percentile. A small minority of households surveyed have higher household incomes, up to \$2.13 million. 137 households have no income, and three households have negative income, with the lowest income being \$267,900.

Table 2: Distribution of Household Incomes

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-267900	28000	55090	78266	100681	2125100

### Most house

Distribution of households by household income, PSID



### Question 2: household income Lorenz curve

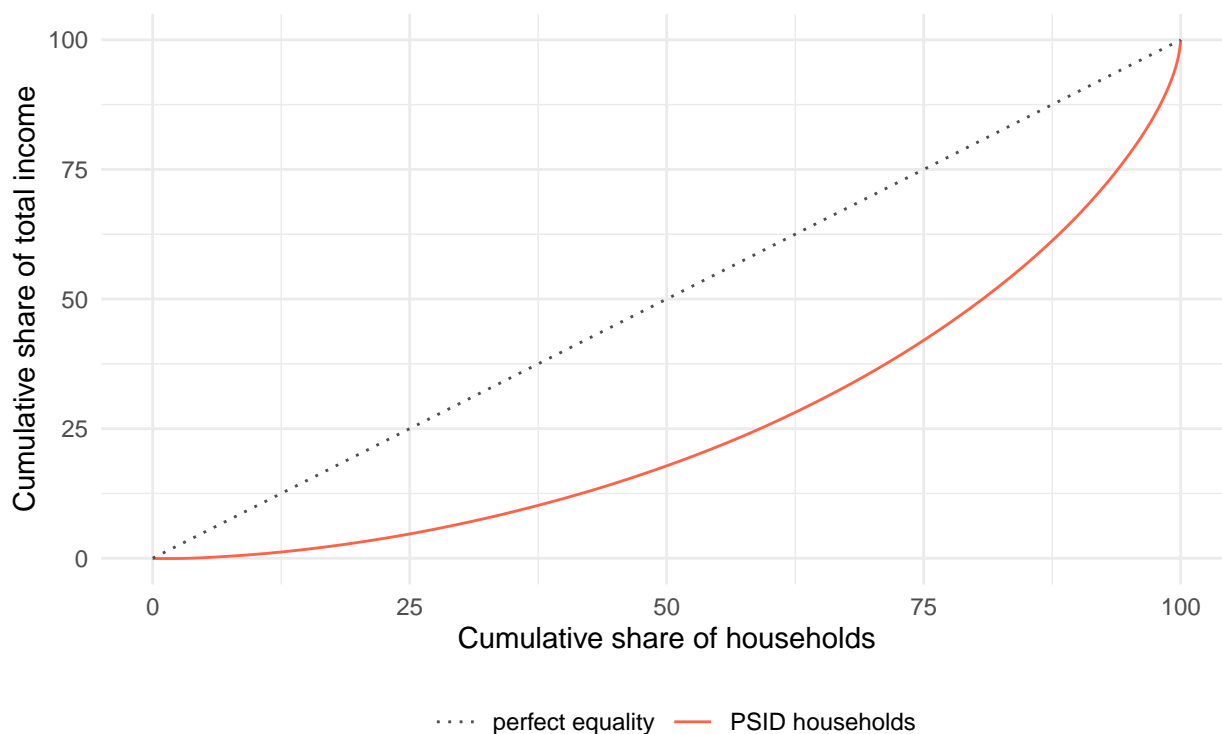
The following Lorenz curve visualizes the cumulative share of households (x) against the cumulative share of total household income (y). The dotted line is what the curve would look like at perfect inequality, where 50% of households possess 50% of total household income. The solid red line represents the relationship between cumulative household share and cumulative household income share.

We see on this curve that the lowest earning 50% of households in the PSID account for only 18% of total household income, and that the bottom 75% of households account for 42% of income. By contrast, the top 10% of highest earning households account for about a third of all household income, and the top 1% of households account for just above 8%. In this dataset, we can tell

that higher-earning households account for more than their proportional share of household income.

## Lorenz Curve of Household Income

Households and income, PSID



## ? SCF

### Question 3: total household income coefficient of variation

To further quantify the level of income inequality in the dataset, we can calculate the coefficient of variation by dividing the standard deviation in household income by the mean household income. This produces a coefficient of variation of 1.15, which is \_\_\_\_\_ compared to the \_\_\_\_\_ derived from the SCF.

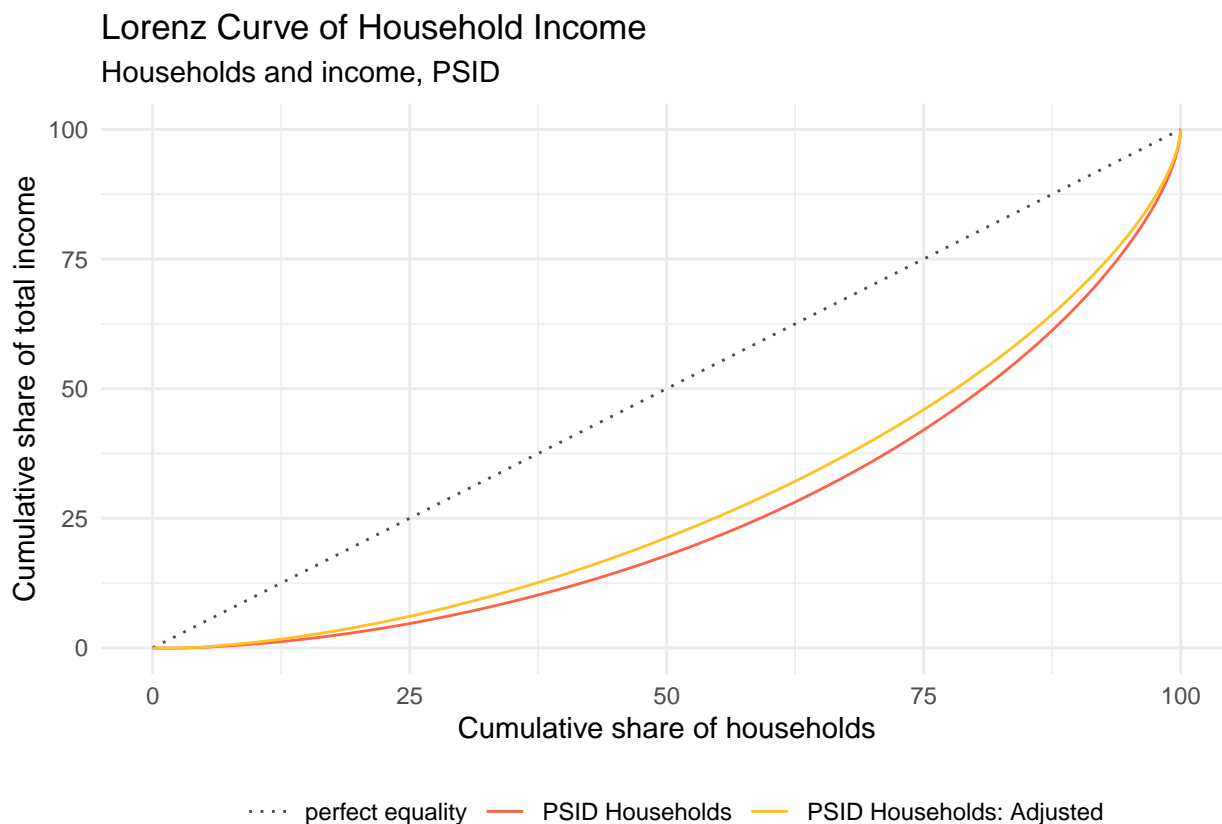
```
## [1] "Coefficient of variation of household income: 1.15"
```

```
##? adjustment method
```

```
## Question 4: adjusted total household income Lorenz curve
```

When we adjust this curve for the number of household heads present, the distance from the line of perfect equality decreases. On the yellow curve representing the relationship between adjusted income share and population share, we see that the bottom 50% of households account for 21% of income, compared to the 18% on the non-adjusted curve.

This indicates that some of the inequality among households may be attributed to differences in household size and the number of earners.



### Question 5: household income percentile ratios

The table below displays the 30th, 50th, 90th, and 99th percentiles of household income in the dataset. The wide gap between the 50th and 90th and 50th and 99th percentiles give us a sense of the income inequality among the households, but percentiles themselves are insufficient.

Table 3: PSID Household Income Percentiles

percentile	value
30th	\$33107.5
50th	\$55090
90th	\$161188
99th	\$396420

The percentile ratios in the table reveal a markedly unequal income distribution. Households at the top earn several times more than those in the middle or bottom. Specifically, households at the 90th percentile earn nearly three times the median income, while those at the 99th percentile earn 7.2 times the median, indicating extreme concentration of income at the very top. The 2.67 ratio between the 30th and 10th percentiles shows that inequality at the lower end exists but is considerably smaller than that among top earners.

Table 4: PSID Household Income Percentile Ratios

percentile_ratio	value
90-30	4.87
90-50	2.93
30-10	2.67

percentile_ratio	value
99-50	7.2

## ? SCF

- Comparison with SCF

### Question 6: mean household income and share by quintile

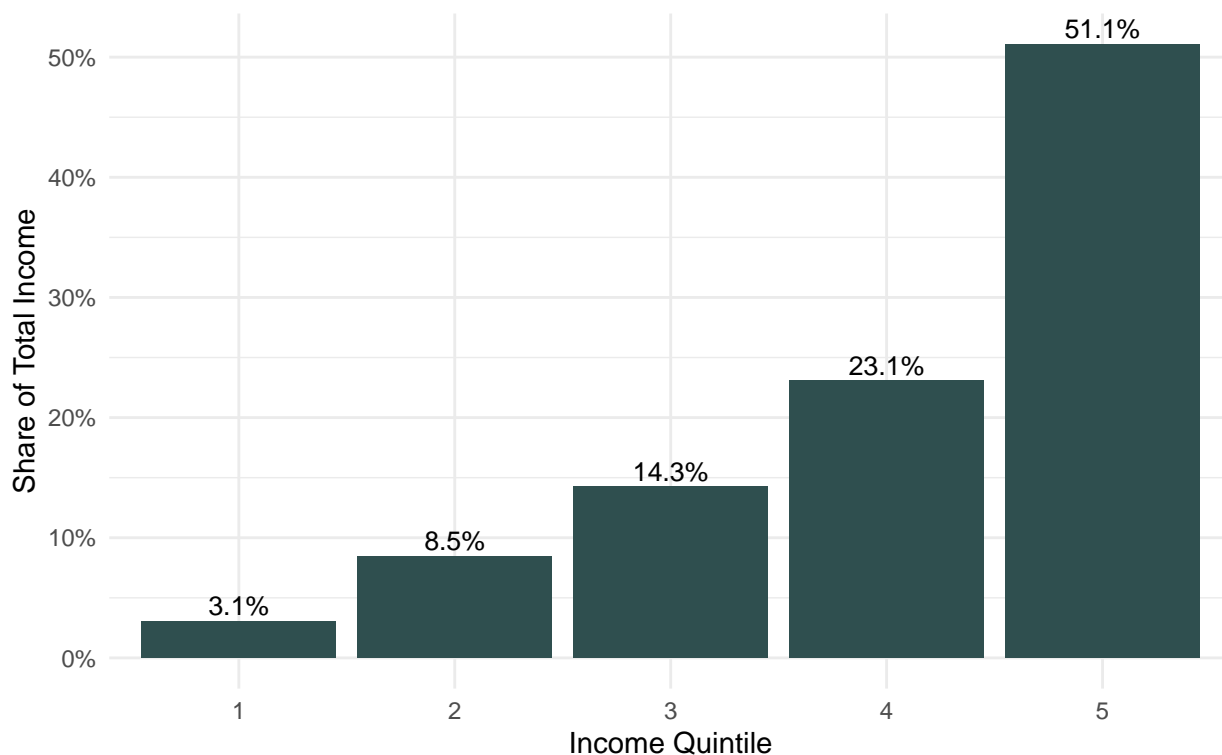
The following table further illustrates this pattern of income inequality among PSID households: the greatest disparities between consecutive quintiles are between quintiles 2 and 1 and quintiles 5 and 4. Quintile 2 has a mean income of \$33,188, 2.76 times the first quintile's mean income of \$12,003. The second quintile's income share of 8.5% is 2.74 times that of the first quintile (3.1%). The fifth quintile's income of \$199,894 is 2.21 times that of the fourth quintile (\$90,408). The fifth quintile's income share of 51.1% is 2.39 times that of the fourth quintile. It's also notable that the income share of the top quintile is greater than that of the remaining 4 quintiles. Taken together, this shows the high overall income inequality among PSID households, as well as the heightened inequality at the bottom and top of the distributions.

Table 5: Share of total household income and mean income by quintile

quintile	income_share	mean_income
1	3.1%	\$12,003
2	8.5%	\$33,188
3	14.3%	\$55,883
4	23.1%	\$90,408
5	51.1%	\$199,894

## Top 20% of PSID households earns more income than bottom 80%

Share of total household income by quintile



# ? SCF dollar(mean\_income)

### Question 7: mean total household income and share for the top 1%

As discussed in class, quintiles can obfuscate the inequality within the top of the distribution. Looking at the top 1% of high earning households, we see that their mean income (\$639,974) is 8.81 times that of the bottom 99% (\$72,570), and 3.2 times that of the top 20%. In terms of income share, the top 1% accounts for 8% of all household income, making up nearly one-sixth of the top quintile's share. Even within the top quintile, there is drastic inequality between the highest and lowest earners.

Table 6: Share of total household income and mean income held by the top 1%

group	income_share	mean_income
bottom 99%	92%	\$72,570
top 1%	8%	\$639,974

#? SCF

compare with scf distribution

## Part 3: Labor Income

### Question 1: household earnings share and mean by quintile

Looking at total labor earnings of households, the inequality within the bottom of the distribution is much higher. The table below lists the mean household earnings and share of household earnings by quintile. The

second quintile of households (\$14,976) earns 516 times the labor income of the first quintile (\$29). The earnings inequality at the top 40% of the distribution is also higher than total household income, but by a smaller margin.

Table 7: Share of total household labor income and mean labor income by quintile ## Question 2: total household income vs household earnings by quintile

Side by side, we can see that the distribution of earnings share is more unequal. The bottom quintile's share of earnings rounds to zero, while the bottom quintile's share of total household income is 3.1%. Quintiles 2 and 3 also hold smaller shares of earnings than they do total income. Quintile 4 has a slightly higher share of earnings than total income, and quintile 5's share of earnings is higher than total income by 6.3 percentage points. Compared to the distribution of total income, earnings is more unevenly distributed, with higher percentage shares concentrated in the top 2 quartiles.

quintile	labor_income_share	mean_labor_income
1	0.0%	\$29
2	5.0%	\$14,976
3	13.4%	\$39,309
4	24.1%	\$72,342
5	57.4%	\$170,395

Table 8: Total income vs earnings quintile shares

quintile	total_income_share	earnings_share
1	3.1%	0.0%
2	8.5%	5.0%
3	14.3%	13.4%
4	23.1%	24.1%
5	51.1%	57.4%

Calculating the coefficient of variation for both metrics confirms this: earnings have a coefficient of variation of 1.36, compared to 1.15 for household income. This indicates that there is more variability in earnings relative to the mean.

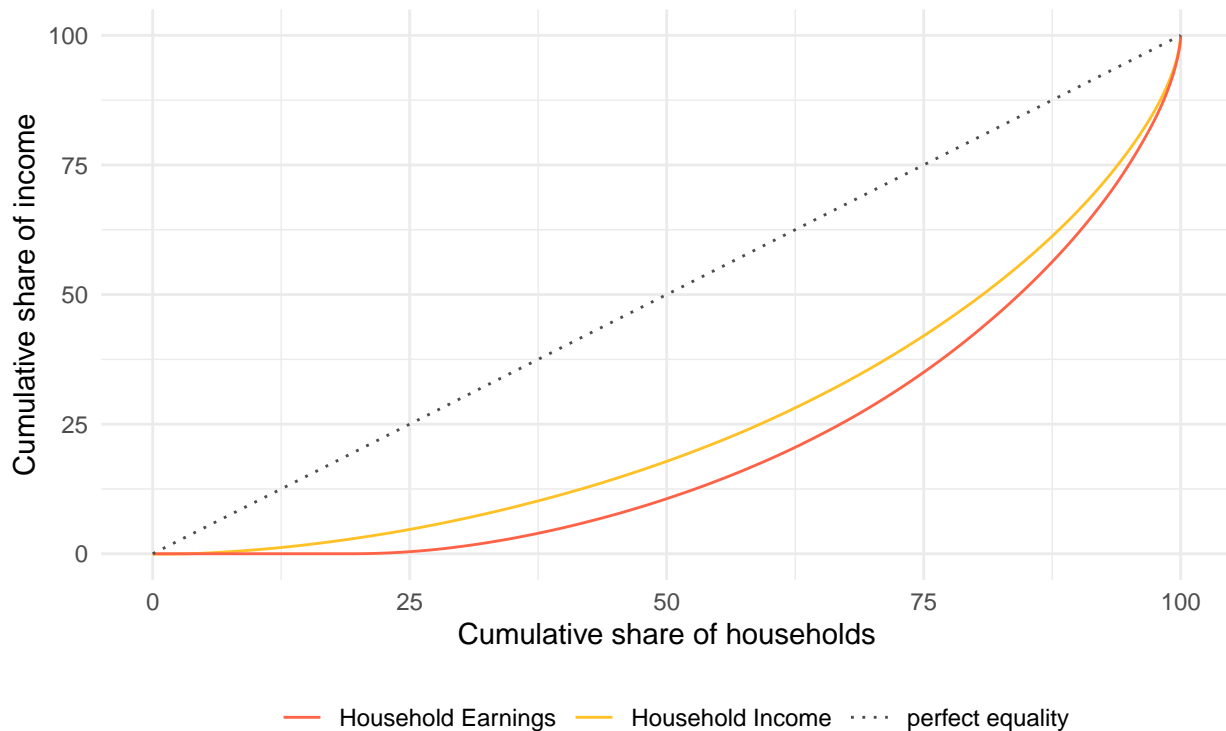


Table 9: Coefficient of Variation: Household Income and Earnings  
This trend is seen when plotted on a Lorenz curve, as well. The curve representing the relationship between cumulative shares of households and income for household earnings is more far away from the line of perfect inequality, indicating that its distribution is more unequal than that of total household income. As discussed in class, redistributive policies like social insurance and means-tested transfers may be contributing to the household income of households at the bottom of the income distribution who do not have labor income. Consequently, the bottom quintile can have a higher share of total household income than household earnings.

metric	cv
Household Income	1.15
Household Earnings	1.36

### Household earnings is more unequal than household income

Lorenz curve of total household income and household earnings, PSID



### Question 3: average share of total household income from earnings

The average share of earnings in total household income is 68.51%, indicating that households typically receive just over two-thirds of their income from labor income. This figure is excluding the 137 households with incomes of 0, as the share would be undefined.

```
## [1] "Mean share of earnings in total household income: 68.51"
```

```
## [1] "Houses with zero household income: 137"
```

## Question 4: mean and share of labor earnings by quintile

In the dataset, households that have more household income make a larger share of their income from labor earnings. The most stark difference is between quintile 1 and 2: the mean share of labor is 46.6% in the former, and increases by 18.9 percentage points for quintile 2. The difference in the earnings share between the remaining quintiles is modest: from quintile 2 to 5, the increase is less than that between quintile 1 and 2.

Table 10: Share of total household income from labor earnings by quintile, PSID # ? SCF

When we pull out the top 1% of households, however, we see that they only get 68.4% of their income from labor earnings, almost the same as the mean of all households. This is because, as discussed in class, the top earning households have a higher share of capital income in their total income, and rely less on labor. They also tend to be older, and some are retired. Compared to SCF, this figure

quintile	mean_labor_share
1	46.6%
2	65.5%
3	71.9%
4	76.8%
5	80.2%

Table 11: Share of total household income from labor earnings, top 1%, PSID

group	mean_labor_share
bottom 99%	77.1%
top 1%	68.4%

- Compare both to SCF
- Comment on differences

## Question 5: weekly wage, contribution of hours and wages to labor earnings inequality

In the dataset, there are 6,971 households where the household head has positive labor earnings and weeks of work. The mean weekly wage for these household heads is \$1,117.74.

```
## [1] 6971
```

```
## [1] 1117.736
```

- 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<sup>2</sup>, education, occupation
- Residuals
- Variance of residuals
- Share of inequality explained by age, education and occupation?