PPHA 31002 | Statistics: Homework 2

Jay Ballesteros

STATISTICAL EXERCISES

Q1. What is the expected value of A?

$$E(A) = 0(0.3) + 2(0.4) + 4(0.3) = 0 + 0.8 + 1.2 = 2$$

Q2. What is the variance and standard deviation of A?

$$Var(A) = (0-2)^2(0.3) + (2-2)^2(0.4) + (4-2)^2(0.3) = 1.2 + 0 + 1.2 = 2.4$$

$$SD(A) = \sqrt{2.4} \approx 1.55$$

Q3. Is the skewness of A positive, negative, or zero? How do you know?

The skewness of A is zero. This is because the distribution is symmetric around the mean (2), with equal probabilities for values below and above the mean (0 and 4 are equally distant from 2, and both have the same probability of 0.3).

Q4. Is the expected value of B less than, greater than, or equal to the expected value of A? Briefly explain how you know.

$$E(B) = 0(0.4) + 2(0.2) + 4(0.4) = 0 + 0.4 + 1.6 = 2$$

The expected value of B is equal to the expected value of A.

Q5. Is the variance of B less than, greater than, or equal to the variance of A? Briefly explain how you know.

$$Var(B) = (0-2)^2(0.4) + (2-2)^2(0.4) + (4-2)^2(0.4) = 0 + 0.8 + 6.4 = 7.2$$

The variance of B is greater than the variance of A. This is because the probabilities in B are more concentrated at the extremes (0 and 4 both with 0.4) compared to A, leading to a higher spread of values around the mean.

Q6. Is the skewness of B the same as the skewness of A? If not, how is it different?

It's the same: 0. Because even if B differs to A in spread because of mode 0 or 4, the distribution is still symmetric around the mean (2).

Q7. What is the probability that a randomly drawn match was exciting?

$$P(X >= 6) = 1 - F(5))1 - 0.906)0.094$$

Q8. What is the mean number of goals scored in a game at this world cup?

Considering that the table is equivalent to a CDF, I need to translate it to a PMF in order to calculate the mean. Therefore:

$$E(X) = 0(0.109) + 1(0.157) + 2(0.281) + 3(0.203) + 4(0.063) + 5(0.093) + 6(0.047) + 7(0.031) + 8(0.016) = 2.672$$

Q9. What is the variance of goals in a game? What is the standard deviation?

$$Var(X) = \sum_x (x - 2.672)^2 p(x) = 3.5364$$

$$SD(X) = \sqrt{3.5364} \approx 1.88$$

Q10. Is the skewness of the goals scored in a game positive, negative, or zero? Briefly explain how you know.
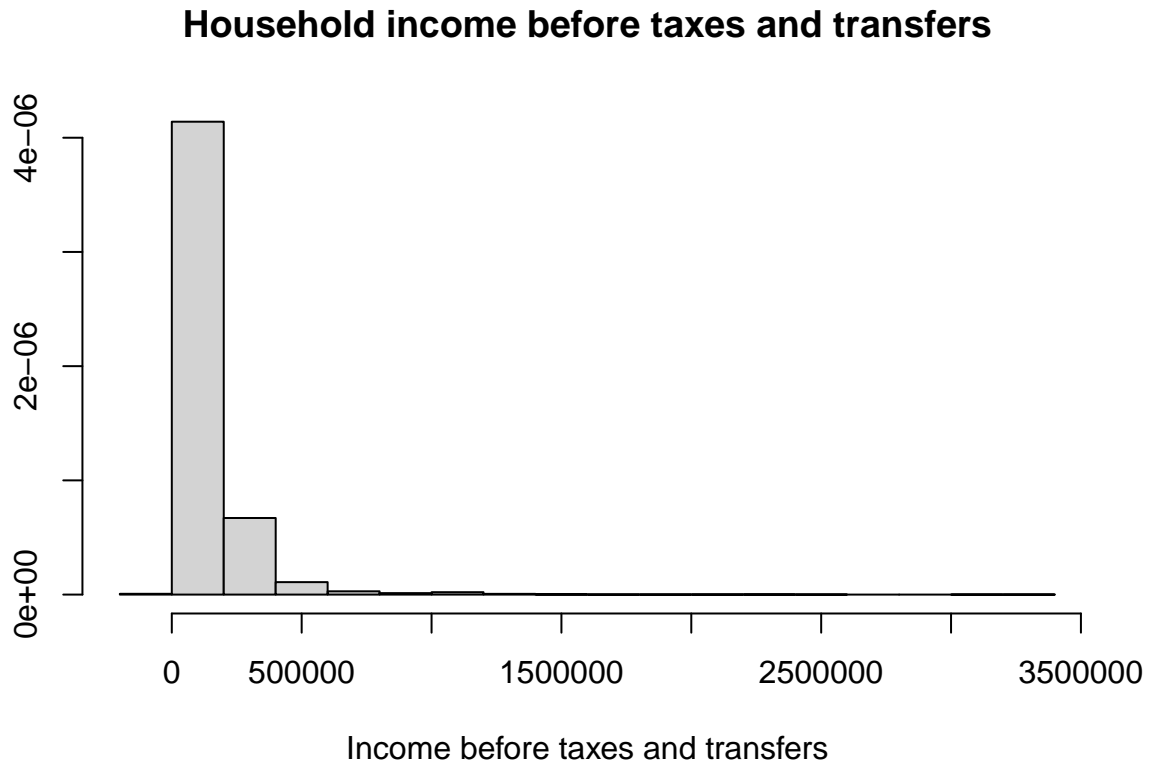
It is positive. First of all because the median doesn't equal the mean (median is 2 and mean 2.67). Second of all the distribution has a long right tail (values go up to 8).

CORE EXERCISE

Q20. How many observations are there in the sample? Are these households or individuals?

There are 127181 observations in the sample. These are households, since the data set contains rows per `hrecord` (unique identifier for each household), and the subsequent variables are at household level (e.g., household income, taxes, transfers).

Q21. Plot a histogram of household income before taxes and transfers ("htotval").

Household income before taxes and transfers

Income before taxes and transfers

Q22. What is the mean and median household income before taxes and transfers in this period?

The mean household income before taxes and transfers in this period is `135,477.2`. The median household income before taxes and transfers in this period is `101,000`.

Q23. Based on the above sample statistics and histogram, is skewness = 0, > 0 or < 0? Briefly explain.

The skewness is > 0. This is because the mean is greater than the median, and the histogram shows a longer tail on the right side of the distribution, indicating a positive skew.

Q24. The same household can pay taxes (federal or state taxes > 0) and receive transfers (transfers > 0). What does it mean that federal taxes < 0? What fraction of the sample has federal taxes < 0 at baseline?

A federal tax < 0 indicates that the household is receiving more in return than it is paying in taxes. That money in return, can be for example, tax credits or refunds.

The fraction of the sample that has federal taxes < 0 at baseline is 20.89%.

Q25. Create three variables that represent the difference between the BBB policy scenario and the baseline policy for the three sets of variables that start with inkind, states_response, and other_spending.

a) What is the mean for each of these new variables?

The mean for the difference in in-kind spending is -753.4031488. Whereas the mean for the difference in state response spending is 7.6006987. Finally, the mean for the difference in other spending is 104.0803874.

b) What does it imply if a mean value is positive?

A positive mean value implies that, on average, the spending in the BBB policy scenario is higher than in the baseline policy for that particular channel (variable or indicator).

Q26. Create a variable for the difference in the federal taxes and transfers in a way that is consistent with the CBO memo. What is the mean value of this new variable?

The mean value for the difference in the federal taxes and transfer is 2440.8460328.

Q27. Create a variable that is the sum of the four changes in questions 25-26. What is the mean value of this new variable?

The mean value of the new variable `sum_diff` is 1797.9356919.

Q28. Create this variable (cbo_income) at baseline in the data by adding transfers to htotval and subtracting off state and federal taxes. Calculate the mean and the median. Why would CBO use this definition of income? Answer in 5 lines or less. [2pt]

The mean of `cbo_income` is 120704.2, and the median is 95833.79. CBO uses this definition of income because it provides a more precise measure of a household's financial resources by accounting for both government transfers and tax obligations, which can significantly impact final disposable income.

Q29. Use the `quantile` function to create a new variable that divides the sample into 10 equal-sized groups according to income. How many households are in each decile group?
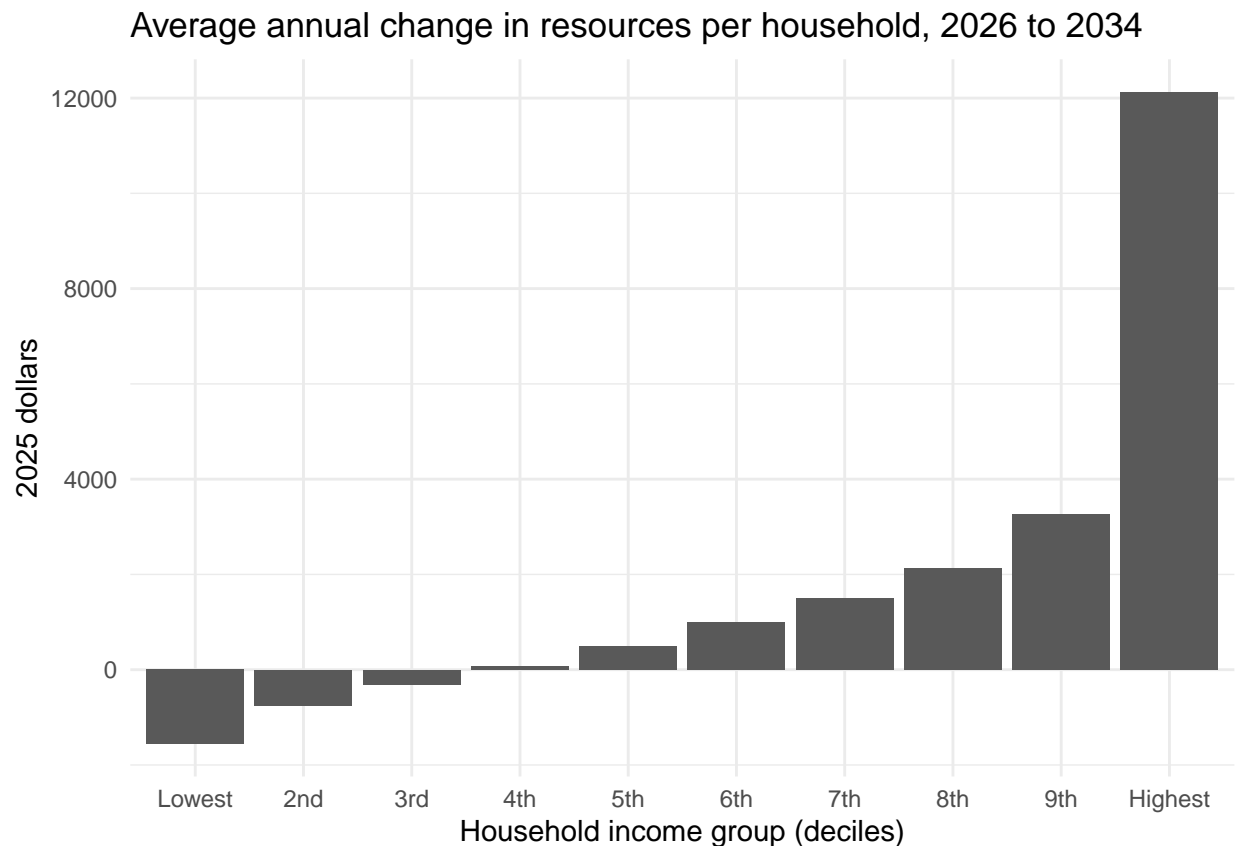
The following table illustrates how many households are in each decile group. The first row accounts for deciles and the second one for the number of households in each group.

```
##
##     1     2     3     4     5     6     7     8     9    10
## 12720 12717 12718 12721 12715 12718 12720 12720 12714 12718
```

Q30. The CBO excludes households with negative income from their analysis. How many households with negative CBO income are there? How many households with zero CBO income? Recorde the decible variable to be missing for these two types of households.
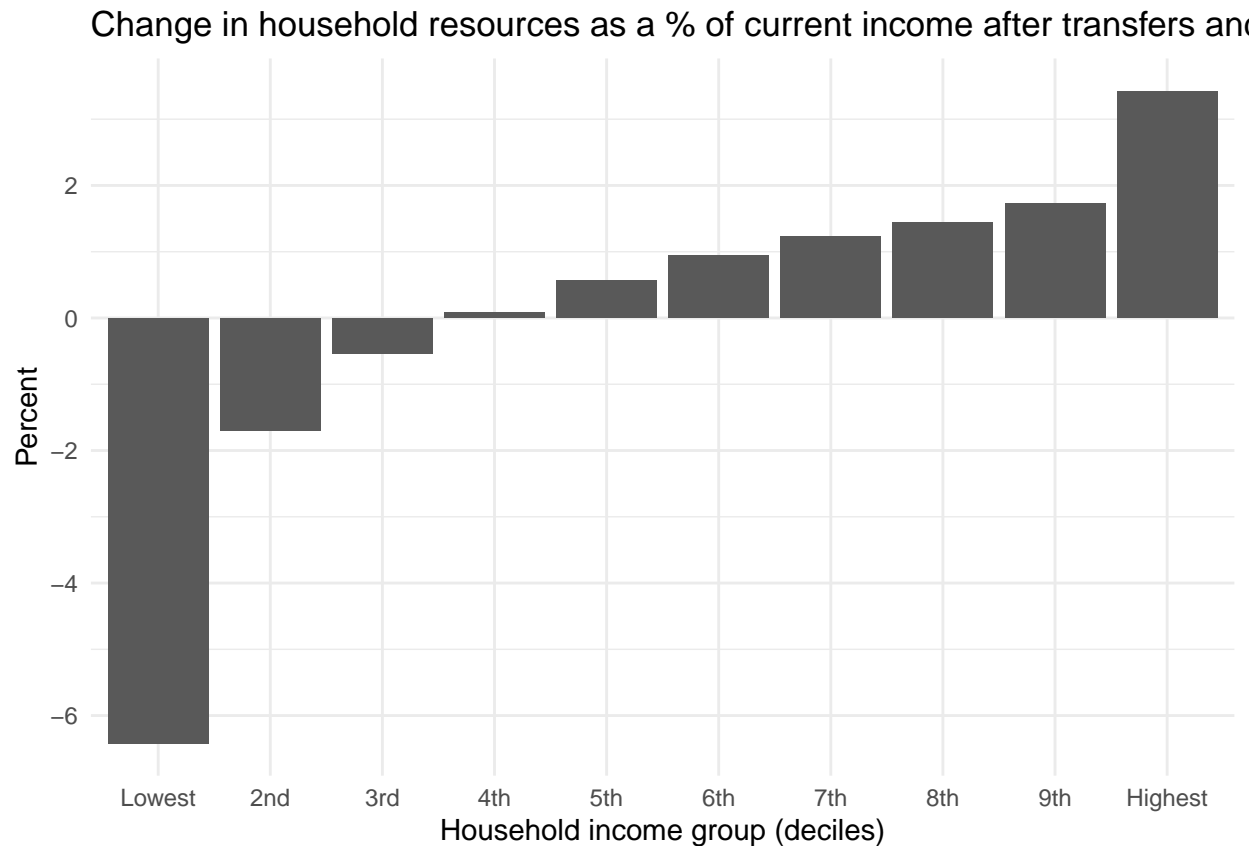
There are 40 households with negative CBO income and 44 households with zero CBO income.

Q31. Create a bar plot that displays the mean change in household resources for each income group from question 27.

The bar plot displays the same overall pattern as figure 1 in the CBO memo. The average change in household resources rises across income deciles, with the highest-income households seeing the largest gains. While the magnitudes differ slightly, but despite those differences, both figures show an upward trend in resources with household income.

Q32. Let's now reproduce part of Figure 2 from the CBO memo. In particular, we will create a bar plot in which the height of each bar corresponds to the height of the diamonds in Figure 2.Take the mean annual change in resources for a given decile that you calculated in question 31 and divide that value by the mean cbo_income for households in that decile.

Change in household resources as a % of current income after transfers and

The pattern in the bar plot matches the CBO figure: the change in household resources increases steadily across income deciles, with the lowest-income households experiencing the largest losses and higher-income households the largest gains as a share of income.