# Week-10 Submission

Anandita Misra

2023-10-20

## Week-9

### Data Story Topic

The topic that I have chosen is **Sockonomics Unveiled**:This investigation explores the impact of missing socks in U.S. laundry between 2000-2019 on household finances, shedding light on how this seemingly trivial, unrelated and very bizarre issue can affect both current income and allocation of future expenses of households in the U.S. We evaluate the decrease in current income and change in allocation of future expenses.

### Data Sources curated so far

1- The data sources curated so far are from Kaggle on personal expenditures in the U.S. by State from 1997-2019- https://www.kaggle.com/code/davidbroberts/us-personal-expenditures-1997-2019 (https://www.kaggle.com/code/davidbroberts/us-personal-expenditures-1997-2019). There is also a data source from Data.gov which provides data on expenditures, income and demographic characteristics of consumers in the U.S. It provides it in different forms which I have accessed from here- https://catalog.data.gov/dataset/consumer-expenditure-survey-dbf32 (https://catalog.data.gov/dataset/consumer-expenditure-survey-dbf32) . These will tell me how much of income on average of a typical household in U.S. is allocated to clothing.

2- The third data source is Statista from where I have accessed Average annual expenditure on laundry and cleaning supplies per consumer unit in the United States from 2007 to 2022 https://www.statista.com/statistics/305499/us-expenditure-on-laundry-and-cleaning-supplies/ (https://www.statista.com/statistics/305499/us-expenditure-on-laundry-and-cleaning-supplies/).

3- To see the percentage of socks within the apparel expenditure, I have accessed- https://www.statista.com/topics/965/apparel-market-in-the-us/#topicOverview (https://www.statista.com/topics/965/apparel-market-in-the-us/#topicOverview) from Statista on the different categories of expenditure that make up the U.S. Apparel Market.

## Week-10

### Title-

Sockonomics Unveiled- how do lost socks from U.S. laundry shape American household finances?

### Why is it important-

In accordance to one report by Bureau of Labor Statistics, Americans spent $198 per annum on average, on socks and hosiery in 2019. This highlights a financial burden that lost socks may pose on families. For a thriving economy, the financial stability of households is a crucial component. The phenomenon of the missing socks may seem inconsequential, but when we extrapolate across all households(millions in US), it represents a substantial economic loss.

Secondly, by the emphasis on sustainable development goals (UNEP, 2020) missing socks encapsulates a broader consumer behavior and wastefulness. We have to understand the economic impact in order to stress on the seriousness of the matter on wastefulness. This knowledge is important for more efficient household

spending habits which are also sustainable. The World Economic Forum (WEF) in 2016 has shown that "seemingly small, everyday choices made by consumers can collectively have a substantial impact on resource use and waste generation."

Lastly, the allocation of future expenses depends to a large extent on purchasing decisions made today. Missing socks lead to increased sock purchases in the future as replacement. In the context of economic theory, seemingly trivial factors can have ripple effects throughout the economy. Behavioral economics research, cited by the World Bank, indicates that "present decisions can indeed have far-reaching effects on future expenses, often referred to as the"spillover effect" or "decision cascades""(World Bank, 2015). Hence, in the larger context of household financial planning, studying the economic impact of missing socks is significant. According to the behavioural economist Thaler in 1999, "presence of"friction" in decision-making processes, can influence consumer choices and have broader economic consequence."

## Specific Columns/Variables

Here is a survey that was conducted- This is the dataset- https://gwern.net/doc/psychology/2019-01-21-eric-socksurvey.csv (https://gwern.net/doc/psychology/2019-01-21-eric-socksurvey.csv) These were the survey questions. The data is based on a survey done with an international audience of sample size n=455. It was conducted on a personalized quiz website. For this dataset, I will be using the question (2), (3) relevant to my data story- (2) QN will be used to estimate how many socks will be washed per week. There is a correlation between number of socks owned and number of socks used. Hence, number of socks used will then be correlated to number of socks washed per week.

(3)QN will be used to find out future expenses on replacement socks. We will see the frequency of missing socks and then, take into account the frequency of the replacement through this survey and find out future expenditure on replacement socks.

Hence, column "Count" and "Frequency" will be used.

Do you have enough pairs of socks? Yes/No

How many pairs of socks do you have? (Numeric)

How often do you buy replacement socks? Monthly Semi-annually Annually Less or never

Who buys your socks? Me Spouse/significant-other Relative Other

Once we take this, we will use the sock loss formula to estimate sock losses in a week -

The Sock Loss Formula given by Samsung-

Sock loss index = (L+C)-(P x A)

Higher the value, the more likely you will lose socks. We will use an adapted version of the formula, and keep everything else constant (using standard value) except number of socks washed in a week: Prob= 0.38+(0.005 x L)+(0.0012 x C)-(0.0159 x P x A)3

Where:

L = Laundry size Calculated by multiplying the number of people in the household (p) with the frequency of washes in a week (f). (here f will be taken at average, p will be taken yearly from this dataset- https://usafacts.org/data/topics/people-society/population-and-demographics/population-data/average-family-size/ (https://usafacts.org/data/topics/people-society/population-and-demographics/population-data/average-family-size/) the column will be "Average number of people in a family")

C = Washing complexity Calculated by adding how many types of wash (t) households do in a week (darks + whites) and multiplying that by the number of socks washed in a week (s) (here t will be held constant and s will be estimated from the above dataset as mentioned)

P = The positivity towards doing laundry Measured on a scale of 1 to 5 with 1 being 'Strongly dislike doing clothes washing' to 5 which represents 'Strongly enjoy doing clothes washing' (we will take this at average)

A = Degree of Attention Which is the sum how many of these things you do at the start of each wash check pockets, unroll sleeves, turn clothes the right way and unrolling socks"

We will take this information, use the sock loss formula, to find out the number of missing socks by year (since we take the number of people in a household by year), then use the single average price of all types of socks in US from the CPI (generalise it to all years to control for price) and again use the average number of people in a household in US (yearly as above in dataset) to get an estimate of the lost current expenditure in that particular year on missing socks. To find future expenditure on replacement socks, we do the same thing, except also take into account the frequency of replacement socks from the first dataset (qn 3), to find the future expenditure arising from these replacement socks, with each particular year as the reference year for future expenditure from that year onwards.

https://www.kaggle.com/code/davidbroberts/us-personal-expenditures-1997-2019 (https://www.kaggle.com/code/davidbroberts/us-personal-expenditures-1997-2019) We then use the column of "Clothing and footwear" from the above dataset to find out the percentage of expenses in missing socks as part of clothing and footwear in years from 1997-2019 and also, as percentage of total expenditure.

# Challenges and Errors:

The main challenge in this investigation to find relevant accurate data from surveys conducted. Collecting data on missing socks and their financial or economic impact is challenging. It is difficult to distinguish the direct impact of missing socks from other factors affecting household finances and amongst laundry-related variables like the types of wash, or the positivity towards wash. We are taking these as constant to avoid confusion. We also have to make estimations or take an average number in order to provide a clear pathway of cause-effect of only lost socks on current and future expenditure and no other variable. We make an estimation of the number of socks washed per week from the number of socks owned and take averages like the average price of all types of socks from CPI- Consumer Price Index.

Furthermore, the dataset used does not specifically include a column for "missing socks." Therefore, it is necessary to use proxy variables or devise a methodology to estimate the financial effect which we did through the sock loss formula.

To ensure the accuracy of results, we need to account for potential errors, such as missing or partial data in the surveys, and we extrapolate the survey information of an international audience into the average price of socks, average number of people in a household in US, the average total expenditure, we have to make an assumption to get a constant average number. We also have to figure out a way to take into account, from missing socks, the frequency of replacement socks in order to find an estimate of future expenditure.