An Evaluation of Weekly Working Hours and Overtime in the 1994 U.S. Labor Force

Name: Yuhan Zhuang (Raine Zhuang)

Student number: 53183687

Group leader: Yuhan Zhuang (Raine Zhuang)

Contributions: This is an independent project.

**Introduction**

This study investigates the mean weekly working hours and the proportion of individuals working over 40 hours per week across four occupation types in the United States workforce in 1994. The four occupation categories are: High-Responsibility/High-Skill Roles, Manual Labor/Trade Occupations, Service/Clerical Occupations, and Other/Flexible Occupations. The focus on a 40-hour threshold stems from its status as a global standard for full-time work (e.g., Working Time standard), and offers insights into work-life balance, labor market dynamics, and employee well-being.

While much existing research has examined annual working hours (e.g., OECD reports on usual working hours), fewer studies specifically explore weekly working hours, particularly the proportion of workers exceeding 40 hours. Moreover, previous studies have typically used broad occupation categories, often overlooking the nuanced distinctions that exist within the workforce. This gap motivates our focus on both the mean weekly hours and the proportion of workers exceeding the 40-hour mark across a more refined set of occupation types. We aim to provide a deeper understanding of working hours by segmenting the workforce into categories that better reflect the diversity of job roles and working conditions.

The study's primary objectives are: 1) to estimate the mean weekly working hours; 2) to assess the proportion of individuals working more than 40 hours per week; 3) to explore potential implications for work-life balance, labor policies, and employee health. Understanding these patterns is particularly relevant in light of ongoing concerns about employee burnout and mental health. Although the data comes from 1994, these issues remain highly relevant today, as reports from organizations such as the OECD continue to highlight overtime and long working hours in various sectors, including healthcare, finance, and law. Despite technological advancements and flexible work arrangements, many workers still face extended workweeks, making the historical trends identified in this study important for understanding current labor market dynamics.

The population for this study consists of U.S. individuals aged 17-90 from the 1994 U.S. Census dataset. The dataset, compiled by Ronny Kohavi and Barry Becker, includes demographic and employment-related information, such as occupation, education, marital status, and weekly working hours. It is representative of a wide range of industries, education levels, and demographic groups within the labor force. Our analysis will focus on the variables related to occupation and hours worked per week, excluding other factors like income, education, and race for simplicity. Missing values in categorical variables such as work class, occupation, and native country are assumed to be random and will not introduce bias into the analysis.

By examining working hours across these varied occupations, this study aims to offer valuable insights into persistent challenges around long work hours, their impacts on workers, and potential avenues for policy improvement.

**Data Collection and Data Summaries**

Two sampling methods are used in this study: Simple Random Sampling (SRS) and Stratified Sampling. Both aim to ensure a representative sample of individuals from the U.S. workforce in 1994, allowing for accurate estimates of mean weekly working hours and the proportion of individuals working over 40 hours per week.

**1. Simple Random Sampling (SRS)**

The sample size is determined using the formula (M.O.E stands for margin of error) and . Here we will use a 95% confidence level, which means . We can also assume M.O.E is 0.05, a commonly used default value in statistical analysis. Since we consider the proportion of individuals working over 40 hours per week as one of the parameters of interest, a conservative guess of the sample variance would be . Therefore, , which is around 385. Since the population size N is known, we can take into account FPC by using , which is around 381. For convenience, we denote 381 as n in the following analysis procedure.

To implement SRS, we use a random number generator with a seed of 1 in R, ensuring each individual has an equal chance of selection from the total population. The sample size for the SRS will be 381, as determined based on the above formulas. Note that we do not take factors like the occupation category into account when selecting.

**2.Stratified Sampling**

We have 14 general occupations: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces. The population is divided into four strata based on occupation category, as there might be a strong relationship between occupation and working hours. The occupations are stratified as follows:

* High-Responsibility / High-Skill Occupations (H): Includes roles like Exec-managerial, Prof-specialty, and Tech-support. These jobs are often associated with long working hours due to high pressure and specialized skills.
* Manual Labor / Trade Occupations (M): Includes Craft-repair, Machine-op-inspct, Transport-moving, Farming-fishing, and Handlers-cleaners. These roles may require irregular hours due to seasonality or task-based deadlines.
* Service / Clerical Occupations (S): Includes Adm-clerical, Sales, and Protective-serv. These roles typically have more predictable working hours but may include occasional overtime during busy periods.
* Other / Flexible Occupations (O): Includes Other-service, Priv-house-serv, and Armed-Forces. These occupations often involve flexible or unpredictable hours due to the nature of the work (e.g., caregiving or military service).

This stratification is important since it simplifies the process of understanding how different occupations might relate to weekly working hours and the likelihood of working more than 40 hours a week, making it easier for further analysis and insights. Moreover, by stratifying the population based on occupation, we can ensure that individuals from each of the four occupation strata are adequately represented in the sample, which could lead to more precise and relevant results.

Once the strata are defined, we have to determine the sample size for each stratum. In this study, we assume that the cost of sampling is the same across strata. Hence, the burden of sample size choice falls onto the population size of the stratum and its guessed standard deviation.

When choosing the strata sample sizes to estimate the proportion of individuals working over 40 hours per week, we can guess strata standard deviations based on the SRS we took. The proportion of individuals working over 40 hours per week is for each stratum. Then the individual stratum standard deviations are calculated using the sample standard deviation formula for binary variables, which is . The strata standard deviations are summarized in the table below:

A screenshot of a computer

Description automatically generated

Table . Note that "H", "M", "O", and "S" stand for High-Responsibility / High-Skill Occupations, Manual Labor / Trade Occupations, Other / Flexible Occupations, and Service / Clerical Occupations. The four capital letters stand for the four strata throughout the analysis.

Since strata standard deviations are different, strata sample sizes will depend both on strata population sizes and estimated strata standard deviations (i.e., optimal allocation). Let and denote the sample size and population size for the stratum High-Responsibility / High-Skill Occupations; denote the sample size and population size for the stratum Manual Labor / Trade Occupations; and denote the sample size and population size for the stratum Other / Flexible Occupations; and denote the sample size and population size for the stratum Service / Clerical Occupations. The strata population size notations, the actual strata population sizes, and their descriptions are summarized in the following table:

A screenshot of a data

Description automatically generated

Table

For optimal allocation, each stratum sample size is proportional to the stratum population size and the guessed stratum standard deviation. The formula used is . We know that , and . The guessed strata standard deviations are in Table 1. The total sample size should be 381, which is the same as in SRS. Thus, the calculated strata sample sizes are as follows:

A screenshot of a computer

Description automatically generated

Table

After determining the sample size for each stratum, we perform random sampling within each stratum using the sample sizes in Table 3. This means that a random sample is selected separately from each of the four strata by using a random number generator, such as by setting the seed to 1 in R.

When choosing strata sample sizes to estimate the mean weekly working hours, we can guess individual stratum standard deviation based on the SRS we took. The stratum standard deviation of working hours for each stratum is calculated using the sample standard deviation formula for continuous variables, which is . The four strata standard deviations are summarized in the table below:

A screenshot of a computer

Description automatically generated

Table

Since strata standard deviations are different, each stratum sample size will depend both on the stratum population size and the guessed stratum standard deviation (i.e., optimal allocation). Let and denote the sample size and population size for the stratum High-Responsibility / High-Skill Occupations; and denote the sample size and population size for the stratum Manual Labor / Trade Occupations; and denote the sample size and population size for the stratum Other / Flexible Occupations; and denote the sample size and population size for the stratum Service / Clerical Occupations.

For optimal allocation, the stratum sample size is proportional to the stratum population size and the guessed stratum standard deviation. The formula used is . We know that and . The guessed strata standard deviations are in Table 4. The total sample size should be 381, which is the same as in SRS. Thus, the calculated strata sample sizes are as follows:

A screenshot of a computer

Description automatically generated

Table

After determining the sample size for each stratum, we perform random sampling within each stratum using the sample sizes in Table 5. This means that a random sample is selected separately from each of the four strata by using a random number generator, such as by setting the seed to 1 in R.

**Data analysis**

**1.SRS**

The population consists of males and females, aged 17 to 90, living in the United States in 1994. It includes individuals from a wide range of occupations, education levels, marital statuses, races, and native countries, with a focus on those who are part of the labor force. The continuous parameter is the mean weekly working hours reported by individuals in the United States workforce in 1994, and the binary parameter is the proportion of individuals working over 40 hours weekly in 1994.

When using SRS, a first advantage is that it is easy, cost-effective, and time-efficient to implement. This is because we can randomly generate a representative sample of the population by randomly selecting a subset without worrying about complex stratification or weighting. Besides, every individual in the population has an equal chance of being selected. In our case, a random sample ensures that every individual, regardless of occupation, age, or income, has the same likelihood of being included. This minimizes the chance of sampling bias.

SRS, however, has some limitations as well. First, it does not account for variability among different categories of occupations. If certain categories have significantly more inherent variability in working hours, a purely random sample might lead to large standard errors, decreasing the efficiency of our estimate. Second, bias could still exist due to the fact that individuals in smaller categories are less likely to be in the sample. This might impact the accuracy of our estimate.

Assume that all individuals in our population have an equal chance of being selected and that they are all selected independently of one another. Also, assume that our sample is also large enough for us to apply the Central Limit Theorem (CLT), which states that as we take more and more samples, the distribution of these averages (called the sampling distribution) starts to look more and more like a normal distribution, no matter the shape of the original population. The formula for estimating mean weekly working hours is . The formula for calculating the standard error is , where . We can also compute a 95% confidence interval for our estimate, and the formula involved is . The summarized results are in Table 6. Table 7 display the simulation-based bootstrap estimate:

A screenshot of a computer code

Description automatically generated

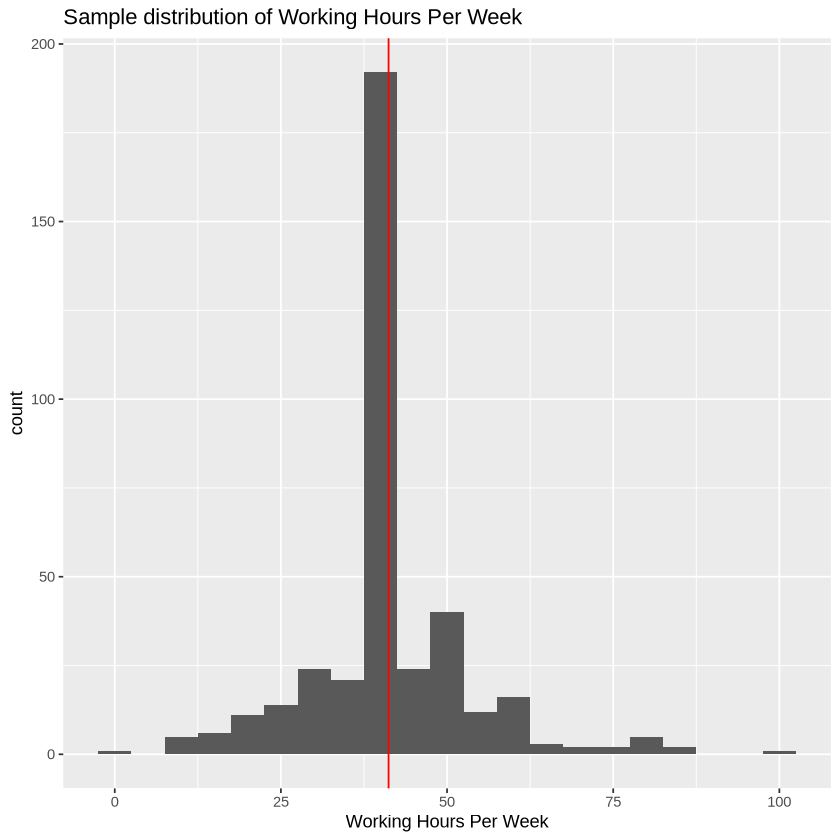
Table . CLT-based estimate

A screenshot of a computer code

Description automatically generated

Table . Simulation-based estimate after taking 3000 resamples from the original SRS sample

The following graphs are the sample distribution and bootstrap distribution with the estimate marked with a vertical line and confidence interval highlighted in red:

  
A graph of a normal distribution

Description automatically generated




Figure 1

Figure . A bootstrap distribution of mean working hours per week after taking 3000 resamples from the original SRS sample.

According to our CLT-based results, the estimate for mean weekly working hours is around 41.19 hours, with a standard error of approximately 0.61. We are 95% confident that the true mean weekly working hours lies somewhere between 40.00 and 42.40. The sample distribution is slightly skewed to the right, with most of the weekly working hours reported by sampled individuals centered around 41.19 hours, the estimate for the population mean. Note that there are outliers on either end of the distribution (0 at the lower end and 100 at the upper end). Given the non-normal sample distribution, the underlying population might not be normal. To check whether the CLT conditions are met given the sample size we have, we will use a bootstrap distribution randomly generated by taking 3,000 resamples. The reason for using a bootstrap distribution is that its results do not rely on the CLT yet can be used to verify the CLT-based results.

The bootstrap distribution in Figure 2, which estimates the sampling distribution of samples means, is approximately normal, suggesting that our sample size is large enough for the CLT to hold. Also, as we can see from Table 6 and 7, the simulation-based bootstrap results and the CLT-based results are pretty similar to each other, with the bootstrap one having a slightly higher standard error and wider confidence interval. Since both methods produce similar results, we are confident that our findings might be generalizable in a broader population context.

Next, we estimate the proportion of individuals working over 40 hours per week. Assume that our sample is large enough for us to apply the CLT. The formula for estimating the proportion is where . The formula for calculating the standard error is , where . We can then compute a 95% confidence interval for our estimate, using the formula . The summarized results are in Table 8. Table 9 display the simulation-based bootstrap estimate:

A screenshot of a computer code

Description automatically generated

Table . CLT-based results

A screenshot of a computer code

Description automatically generated

Table . Simulation-based results after taking 3000 resamples from the original SRS sample

The following graphs are the proportion divided into two categories: hours or > 40 hours per week and the bootstrap distribution of sample proportions:

A graph of a number of hours

Description automatically generated with medium confidence

Figure . Proportion divided into two categories: hours or > 40 hours per week

A graph of a diagram

Description automatically generated with medium confidence

Figure . The bootstrap distribution with the estimate marked with a vertical line and confidence interval highlighted in red.

According to our CLT-based results, the estimate for the proportion of individuals working over 40 hours per week is around 0.29, with a standard error of approximately 0.0229. We are 95% confident that the true mean weekly working hours lies somewhere between 0.2460 and 0.3369. From Figure 3, we can see that the estimated proportion of individuals working over 40 hours per week is nearly half of the estimated proportion of individuals working no more than 40 hours per week, indicating that nearly a third of our sampled individuals work over 40 hours per week. Again, to check whether the CLT conditions are met given the sample size we have, we will use a bootstrap distribution randomly generated by taking 3,000 resamples.

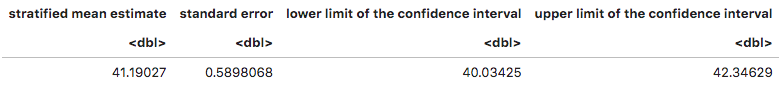
The bootstrap distribution in Figure 4, which estimates the sampling distribution of samples proportions, is approximately normal, suggesting that our sample size is large enough for the CLT to hold. Also, as we can see from Table 8 and 9, the simulation-based bootstrap results and the CLT-based results are pretty similar to each other. Since both methods produce similar results, we are confident that our findings might be generalizable in a broader population context.

**2.Stratified Sampling**

When it comes to stratified sampling, the two parameters of interest remain the same. However, we now have four different populations with different sizes, which are documented in Table 2.

Since the stratified sampling method means drawing samples from each strata population and combining strata results to estimate parameters, we do not need to worry about the large standard errors resulted from potentially large variability between strata. Hence, the precision of our estimates might increase. Nevertheless, stratified sampling is typically more costly and time-consuming than SRS is because we have to sample from each stratum. On top of that, if the strata are poorly chosen, the estimation might not perform better than that of SRS.

We will estimate the mean weekly working hours first. Assuming that all individuals in each stratum have an equal chance of being selected and that they are all selected independently of one another. Also, assume that our samples are all large enough for us to apply the CLT. The formula for estimating mean weekly working hours is . The formula for calculating the standard error is , where . We can compute a 95% confidence interval for our estimate. The formula involved is . The summarized results are in Table 10:



Table

This is the bootstrap distribution of sample means from above with the vertical line at the stratified mean estimate:

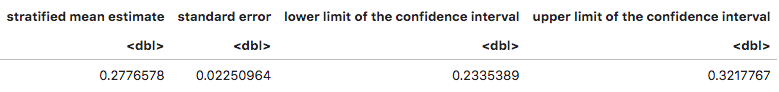
A graph of a normal distribution

Description automatically generated

Figure

According to our CLT-based results, the estimate for mean weekly working hours is around 41.19 hours, with a standard error of approximately 0.59. We are 95% confident that the true mean weekly working hours lies somewhere between 40.03 and 42.35. The mean estimate is extremely similar to that resulted from SRS. We can also see that from Figure 5, where the stratified estimate for the mean is still at the center of the previous bootstrap distribution of sample means. The standard error is only slightly smaller than that of SRS and the confidence interval only slightly narrower. This is a bit surprising as we predict the standard error will be largely reduced due to the elimination of variability between strata. The slight decrease in standard error is probably due to our initial assumption that all other categorical variables are not related to weekly working hours, and this will be further discussed in the conclusion section.

Then we will estimate the proportion of individuals working over 40 hours per week using stratified sampling. The assumptions are the same as those in the mean estimation. The formula for estimating the proportion is , where . The formula for calculating the standard error is , where . We can compute a 95% confidence interval for our estimate. The formula involved is . The summarized results are in Table 11:



Table

The graph below is the proportion divided into two categories: hours or hours per week:

A graph of a bar chart

Description automatically generated with medium confidence

Figure

This is the bootstrap distribution of sample proportions with the vertical line at the stratified proportion estimate:

A graph of a graph with a blue line

Description automatically generated

Figure

According to our CLT-based results, the estimate for the proportion of individuals working over 40 hours per week is around 0.28, still nearly a third of the sampled individuals, with a standard error of approximately 0.0225. We are 95% confident that the true mean weekly working hours lies somewhere between 0.2335 and 0.3218. The stratified estimate and its confidence interval being slightly to the left of the SRS results is likely due to how stratified sampling adjusts for strata differences. The method weights strata differently, which can shift the overall estimate if certain strata with a lower proportion but higher variability are weighted more. This leads to a more precise estimate, but with a slight downward adjustment compared to the SRS estimate, reflecting the population's underlying heterogeneity. We can also see this adjustment from Figure 7, where the stratified estimate is to the left of the center of the previous bootstrap distribution of sample proportions. The standard error is only slightly smaller than that of SRS, while the confidence interval is slightly narrower. This suggests that stratified sampling may improve the precision of the proportion estimate. This could be useful in applications where knowing the exact proportion of workers exceeding 40 hours is critical for policy or business decisions (e.g., labor force management, wage policy). However, the improvement is modest. This could be due to our choice of strata and our assumption that other variables would not affect working hours, which will be further discussed in the conclusion section.

**Conclusions and discussion**

Our final estimate of the mean weekly working hours in the United States for 1994 is 41 hours, with approximately 28-29% of individuals working over 40 hours per week. These findings highlight that a significant portion of the workforce was working long hours, with nearly a third exceeding the standard 40-hour workweek. Given the increasing work pressure and longer working hours in today's labor market, this result emphasizes the persistence of extended working hours. The widespread nature of long working hours suggests potential strain on worker well-being, including higher stress levels and burnout. The high proportion of workers exceeding 40 hours could also indicate a lack of sufficient labor protections or overtime compensation, which might affect productivity and job satisfaction. These findings are critical for shaping labor policies aimed at improving work-life balance, regulating overtime, and fostering healthier, more sustainable workforce practices.

However, several limitations must be acknowledged. First, we assume that other variables are constant across strata. This assumption limits the accuracy of our stratified estimates, as it ignores potential interactions between these variables and weekly working hours. If factors such as age, education, or marital status influence working hours differently across strata, this assumption could lead to biased or underestimated effects. As a result, the stratified estimates may not fully capture the complexity of the data, affecting both the reliability of our conclusions about the true distribution of working hours and the generalizability of our results to the broader population. This failure to capture data complexity likely explains why the precision of our stratified estimates (for both the mean and the proportion) did not significantly improve over the SRS estimates.

Second, the strata were chosen based on general assumptions rather than data-driven criteria. This subjectivity in the stratification process could introduce bias and overlook more nuanced or statistically significant differences within occupations that influence working hours. As a result, the estimates may be less accurate because the occupational categories may not fully reflect the diversity of working conditions or work-hour patterns within the population. Future studies should adopt a more data-driven approach to stratification to better capture these subtleties.

In conclusion, this study provides valuable insights into the dynamics of working hours in 1994, which are essential for shaping labor policies that promote a healthier and more sustainable workforce. While the sample used in this analysis is representative of the U.S. labor force in 1994, caution should be exercised when generalizing these results to other populations or time periods. The sample was drawn using simple random and stratified sampling methods, which likely capture a broad cross-section of the workforce. However, differences in work patterns across industries, geographic regions, or demographic groups might affect the generalizability of the findings. Future research could improve the generalizability by considering the interaction of multiple factors with working hours, including more diverse populations, or examining trends over multiple years, providing a clearer understanding of how working hour patterns have evolved and how they might differ when diverse factors are taken into account.

**References**

[1]“LMF2.1: Usual Working Hours per Week by Gender.” OECD Family Database, Oe.Cd/Fdb, Aug. 2022, [www.oecd.org/els/family/LMF\_2\_1\_Usual\_working\_hours\_gender.pdf](http://www.oecd.org/els/family/LMF_2_1_Usual_working_hours_gender.pdf). Accessed 12 Nov. 2024.

[2]“Hours Worked: Average Usual Weekly Hours Worked - Averages (Edition 2023).” OECD iLibrary, [www.oecd-ilibrary.org/employment/data/oecd-employment-and-labour-market-statistics/hours-worked-average-usual-weekly-hours-worked-averages-edition-2023\_82bc9d9c-en](http://www.oecd-ilibrary.org/employment/data/oecd-employment-and-labour-market-statistics/hours-worked-average-usual-weekly-hours-worked-averages-edition-2023_82bc9d9c-en). Accessed 12 Nov. 2024.

[3]“Working Time.” Wikipedia, Wikimedia Foundation, <en.wikipedia.org/wiki/Working_time#:~:text=Standard%20working%20hours%20of%20countries,above%20the%20normal%20hourly%20payments>. Accessed 12 Nov. 2024.

[4] Becker, Barry and Ronny Kohavi. "Adult." UCI Machine Learning Repository, 1996, <https://doi.org/10.24432/C5XW20>. Accessed 12 Nov. 2024.

**Appendix**

**R code:**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

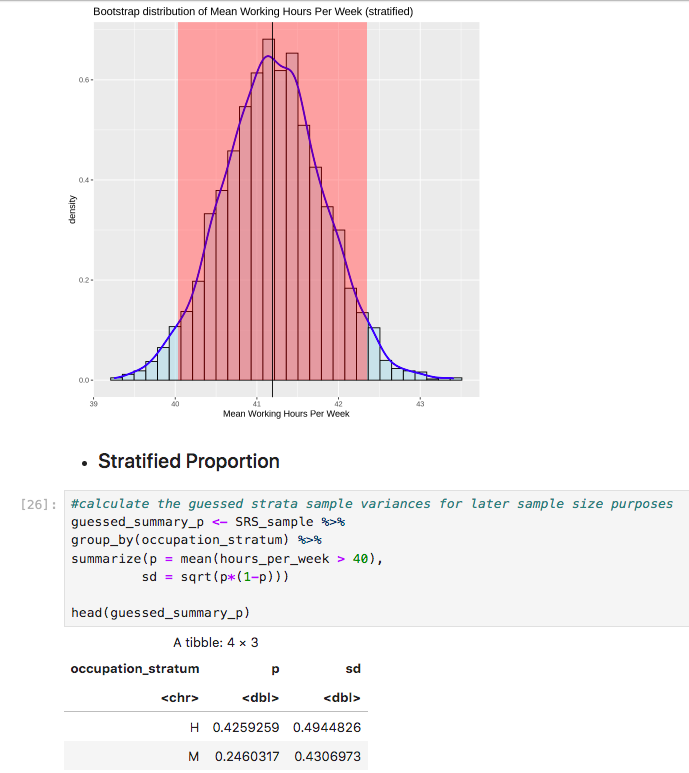
Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

A graph of time and hours

Description automatically generated with medium confidence