

# Factors influencing Total Income in the United States\*

## Research and Statistical Analysis on Factors influencing Total Income in the United States of America

Wen Han Zhao

December 1, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

### Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>3</b>
2.1	Overview . . . . .	3
2.2	Measurement . . . . .	3
2.3	Outcome variables . . . . .	3
2.4	Predictor variables . . . . .	3
<b>3</b>	<b>Model</b>	<b>3</b>
3.1	Model set-up . . . . .	5
3.1.1	Model justification . . . . .	6
<b>4</b>	<b>Results</b>	<b>6</b>
<b>5</b>	<b>Discussion</b>	<b>6</b>
5.1	First discussion point . . . . .	6
5.2	Second discussion point . . . . .	6
5.3	Third discussion point . . . . .	6
5.4	Weaknesses and next steps . . . . .	6
	<b>Appendix</b>	<b>7</b>

---

\*Code and data are available at: <https://github.com/younazhao/Income-Influencing-Factors.git>.

<b>A Additional data details</b>	<b>7</b>
<b>B Model details</b>	<b>7</b>
B.1 Posterior predictive check . . . . .	7
B.2 Diagnostics . . . . .	7
<b>References</b>	<b>8</b>

# 1 Introduction

Income inequality is a central concern in economic research, reflecting disparities in access to opportunities, resources, and outcomes. In the United States, total income is influenced by a variety of factors, including personal, professional, and demographic characteristics. This paper explores the relationship between total income and predictors such as marital status, sex, educational level, labor force participation, and occupation. By employing statistical modeling techniques, we aim to uncover the extent to which these variables shape income levels and contribute to broader income disparities.

The estimand of this study is the expected total income for individuals based on their social, personal and occupational characteristics. Specifically, the analysis seeks to quantify how factors such as marital status, sex, educational attainment, participation in the labor force, and occupational category predict variations in total income across the U.S. population. This includes identifying both independent and among these variables.

Preliminary results indicate that educational attainment is one of the strongest predictors of total income, with higher levels of education yielding significantly greater earnings. Marital status also plays a notable role, with married individuals generally earning more than their single counterparts. Gender disparities in income persist, with men earning more on average than women across most occupational categories. Labor force participation and occupation further highlight critical differences, as full-time workers and those in higher-skilled professions consistently report higher incomes.

Understanding the drivers of income disparities is essential for designing policies that promote economic equity and mobility. By identifying how marital status, gender, education, labor force participation, and occupation contribute to income levels, this research offers actionable insights for policymakers, employers, and educators. Addressing these disparities not only fosters greater fairness but also enhances the overall economic productivity and social cohesion of the United States.

To do this, total income was analyzed in Section 2 Section 2. The prediction from the model has show accuracy in lower total income group. The model has shown difficulties in predicting higher total income. Section 3 Section 3 has further explanation on the Bayesian Generalized Linear Model. Section 4 ?@sec-results show the coefficient beta from the model. Section 5 ?@sec-discussion focus on the weakness of this study and further possible improvements.

Lastly, Section 6 [?@sec-appendix](#) contains methodology to perform a survey to collect data for Total Income.

## 2 Data

### 2.1 Overview

The Survey of Consumer Finances (SCF), conducted in 2022 covering the period 1989 to 2022 by the Board of Governors of the Federal Reserve System of the United States ([scfdata?](#)), provides a comprehensive snapshot of household financial conditions across the nation. This dataset includes detailed information related to total income, capturing key predictors such as sex, education level, occupation category, salary income, and income from other sources. By offering rich, granular data, the SCF serves as an invaluable resource for analyzing the factors influencing income distribution and economic disparities within the United States.

### 2.2 Measurement

The survey contained a panel element over 2 periods. Respondents to the 1983 survey were re-interviewed in 1986 and 1989. Respondents to the 2007 survey were re-interviewed in 2009. In order to ensure the representativeness of the study, respondents are selected randomly in order to attempt families from all economic strata.

### 2.3 Outcome variables

Histogram of the predicted variables, Total income.

### 2.4 Predictor variables

Predictors variables selected are Age, Sex, Marital Status, Education Level, Labour Force, Work Status, and Occupation category.

## 3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in [Appendix B](#).

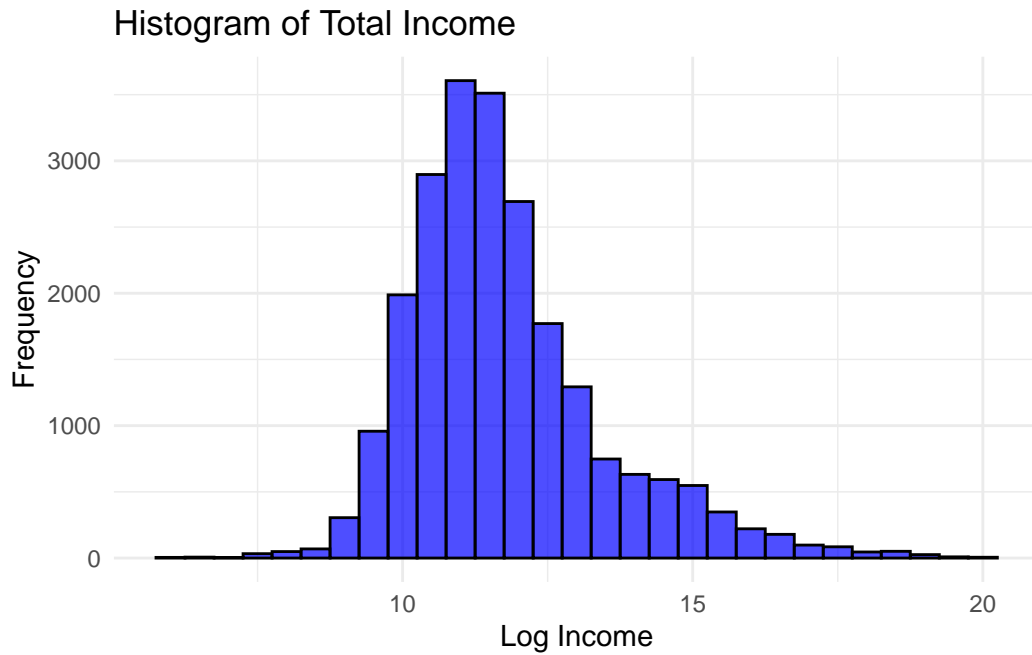


Figure 1: Histogram of Total Income

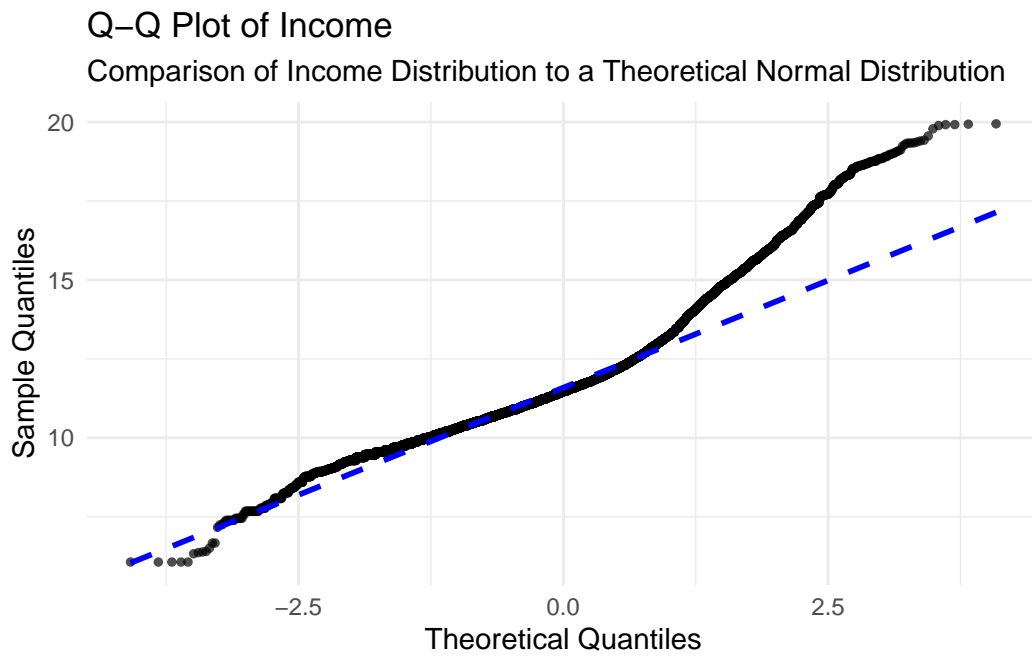


Figure 2: Comparison of Income Distribution using QQ Plot

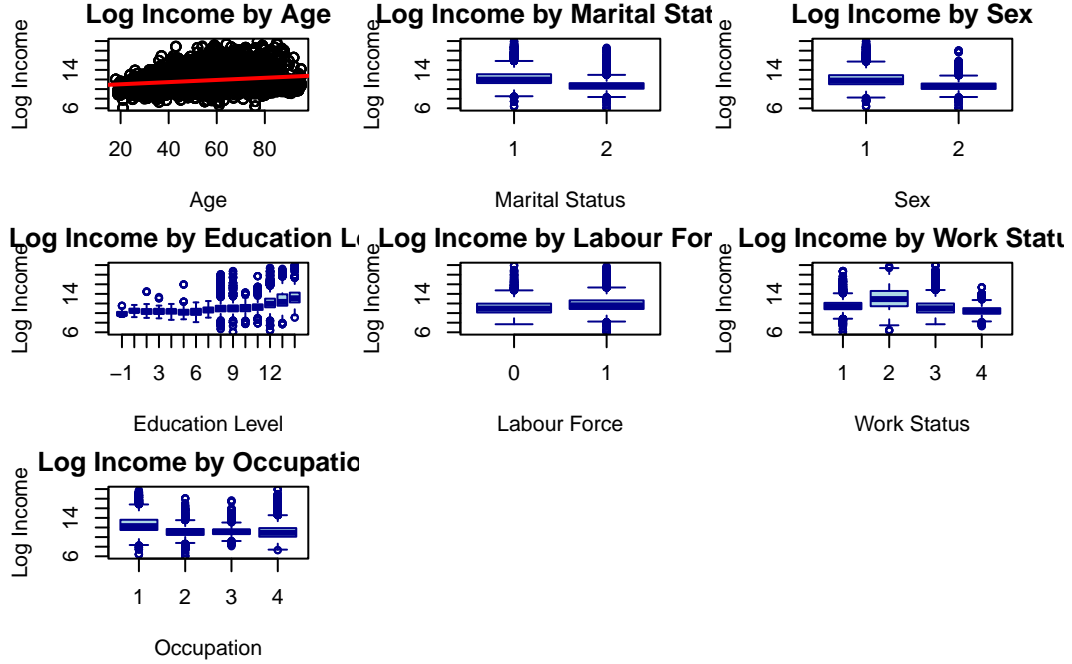


Figure 3: Relationship of Predictors variables with Total Income

### 3.1 Model set-up

Define  $y_i$  as the number of seconds that the plane remained aloft. Then  $\beta_i$  is the wing width and  $\gamma_i$  is the wing length, both measured in millimeters.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (1)$$

$$\mu_i = \alpha + \beta_i + \gamma_i \quad (2)$$

$$\alpha \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\gamma \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\sigma \sim \text{Exponential}(1) \quad (6)$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

### **3.1.1 Model justification**

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance  $\theta$ .

## **4 Results**

Our results are summarized in Table ??.

## **5 Discussion**

### **5.1 First discussion point**

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

### **5.2 Second discussion point**

Please don't use these as sub-heading labels - change them to be what your point actually is.

### **5.3 Third discussion point**

### **5.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.

## Appendix

### A Additional data details

### B Model details

#### B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected  
by, the data

#### B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algo-  
rithm

## References

- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.