Regression Analysis
Ava Tuccio and Meg Simon
Final Project
December 3, 2024

---

**Data Description**

The data selected from IPUMS (Integrated Public Use Microdata Series) for the year 2023 includes a set of demographic, geographic and economic variables including sex, age, hours worked, region, marital status, educational attainment beyond college, number of children, and race. The analysis is focused specifically on individuals who have completed a four-year college degree. To ensure all variables are mutually exclusive and all inclusive, dummy variables were created like midwest, northeast, south, and west to accompany the different regions in the USA. Additionally, our key variable of interest is the field of study (major) for respondents' bachelor's degrees, categorized into 11 groups. These categories are based on the ten most frequently reported majors, with an additional category for all other fields. Below is a detailed description of the selected data and the process we used to construct the dummy variables.

Citation is as follows from Version 15.0 from IPUMS

Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. 2023. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024. https://doi.org/10.18128/D010.V15.0

---

***Variables extracted from Census Year: 2023***

| **Continuous Variables** | **Description** |
|---|---|
| *earnings (Change from incwage)* | Respondent's total pre-tax wage and salary income for the previous year, expressed in contemporary dollars. |
| *hrswork (Change from uhrswork)* | The number of hours per week that the respondent usually worked, if the person worked during the previous year. |
| *age* | Reports the person's age in years as of the last birthday. |

| **Categorical Variables** | **Description** | **Dummy Variables Created** |
|---|---|---|
| *region* | Identifies the region and division where the housing unit was located. | 1 = midwest, 0 otherwise<br>1= northeast, 0 otherwise<br>1= south, 0 otherwise<br>1=west, 0 otherwise |
| *degfield* | The field in which the person received a Bachelor's degree, if the person holds a Bachelor's degree. | 1= communications, 0 otherwise<br>1= education, 0 otherwise<br>1= biology, 0 otherwise<br>1= english, 0 otherwise<br>1= psychology, 0 otherwise<br>1= socialsciences, 0 otherwise<br>1= engineering, 0 otherwise<br>1= finearts, 0 otherwise<br>1= medical, 0 otherwise<br>1= business, 0 otherwise<br>1= other, 0 otherwise<br>(other is classified as Agriculture, environment and natural resources, architecture, area, ethnic, and civilization studies, communication technologies, computer and information sciences, cosmetology services and culinary arts, engineering technologies, linguistics and foreign languages, family and consumer sciences, law, liberal arts and humanities, library science, mathematics and statistics, military technologies, interdisciplinary and multidisciplinary studies, physical fitness, parks, recreation, philosophy and religious studies, theology and religious vocations, physical sciences, nuclear, industrial radiology, and biological technologies, criminal justice and fire protection, public affairs, policy, and social work, construction services, electrical and mechanic repairs and technologies, transportation sciences and technology, and history). |
| *nchild* | Counts the number of own children residing with each individual. | 1= child (at least one), 0 otherwise |
| *sex* | Reports whether the person was male or female. | 1=male<br>0=female |
| *race & hispan* | RACE categories consist of national origin groups, allowing respondents to report as many races as they | 1=white_nh, 0 otherwise<br>1=black_nh, 0 otherwise<br>1=other_nh, 0 otherwise<br>1=hispanic, 0 otherwise |

| | | |
|---|---|---|
| | felt necessary to describe themselves.<br><br>HISPAN identifies persons of Hispanic/Spanish/Latino origin and classifies them according to their country of origin when possible. | |
| *educ* | Indicates respondents' educational attainment, as measured by the highest year of school or degree completed. | 1=post_college if 5+ years of college, 0 otherwise |
| *marst* | Gives each person's current marital status | 1=married, 0 otherwise |