Exercise 5: Census Exercise 2

For the exercise:

1. Conduct a get\_pums( ) request with at least 10 variables (including PUMA) for any ACS   
   (1 or 5) for any available year, weighted by person)

Example: get\_pums(variables = c("PUMA", "SEX", "MIG", "AGEP", "SCHL", "WAGP", "COW", "ESR", "MAR", "NOC"), state = "WV", survey = "acs5", year = 2020, rep\_weights = "person")

My get\_pums() request:

ten\_var <- c('PUMA', 'AGEP', 'SCHL', 'WKL', 'HISP', 'NATIVITY', 'ESR', 'NOC', 'HICOV', 'POVPIP')

pums\_df <- get\_pums(variables = ten\_var, state = 'AZ', year = 2018,

survey = 'acs1',

variables\_filter = list(

SEX = 2, MAR = 1

), rep\_weights = "person")

1. **Save that request as a csv (and submit that file with your assignment)**
2. Compose a hypothesis using the variables you included. You can replicate the examples from the lecture or book using another state or generate an original hypothesis.

I recoded most of the categorical variables and the poverty ratio variable for the logistic regression model. Details on the data preparation and recoding are in the included .rmd file.

1. Conduct a multivariate analysis of the hypothesis.
2. Present that analysis in a table (using stargazer or similar)
3. **Submit the table and a 2-3 paragraph discussion of the findings as your submission.**
4. Extra: create a plot of one of the findings or variables using the PUMAs

This study examines the relationship between childbirth and sociodemographic factors among married women in Arizona using 2018 individual-level one-year PUMS data. The decision to have children is influenced by a range of economic, personal, and cultural factors, including financial stability (poverty ratio), personal resources (education, employment, health insurance, recent work), geographic location (PUMA), biological constraints (age), and cultural norms surrounding family formation (Hispanic origin, nativity). To test whether the probability of having at least one child varies significantly across these dimensions, I estimated a survey-weighted logistic regression model with childbirth as the dependent variable and college completion, Hispanic origin, nativity, PUMA, employment status, having worked in the past 12 months, being above/below the poverty line, health insurance coverage, and age as predictors.

Table 1: Logistic Regression Results

| Variable | **OR** | **95% CI** | **p-value** |
| --- | --- | --- | --- |
| **College** |  |  | **<0.001** |
| *yes* | 1.434 | 1.247, 1.650 |  |
| **Hispanic Origin** |  |  | **0.002** |
| *yes* | 1.296 | 1.090, 1.541 |  |
| **Health Insurance** |  |  | 0.95 |
| *yes* | 1.006 | 0.809, 1.251 |  |
| **Employed** |  |  | 0.71 |
| *yes* | 1.049 | 0.802, 1.373 |  |
| **U.S. Native** |  |  | **<0.001** |
| *yes* | 0.603 | 0.505, 0.720 |  |
| **Age** | 0.893 | 0.888, 0.898 | **<0.001** |
| **Worked in past 12 months** |  |  | **0.016** |
| *yes* | 0.737 | 0.564, 0.963 |  |
| **At/Below Poverty Line** |  |  | **0.011** |
| *yes* | 1.308 | 1.049, 1.631 |  |
| **PUMA** |  |  | **<0.001** |
| Abbreviations: CI = Confidence Interval, OR = Odds Ratio | | | | |

The logistic regression results reveal statistically significant associations between childbirth and several key sociodemographic factors at the α = 0.05 level. Married women with at least an associate degree are 43% more likely (CI: 25%, 65%) to have at least one child compared to those without a college education, controlling for other variables. The likelihood of having one or more children increases 30% for Hispanic women (CI: 9%, 54%), while U.S.-born women exhibit a 40% lower probability (CI: 28%, 49%) of childbirth compared to foreign-born counterparts. Recent work status also plays a significant role, with married women that have worked in the past 12 months being 26% less likely to have a child (CI: 4%, 44%). Age is negatively associated with childbirth, such that each additional year reduces the likelihood by 10.7%, all else held constant (CI: 10.2%, 11.2%). Married women at or below the federal poverty line are 31% more likely to have at least one child (CI: 5%, 63%). Geographic variation, captured through PUMA codes, is largely negatively associated with childbirth, suggesting potential regional disparities in fertility patterns across Arizona’s urban and rural areas (I included the full regression table with PUMA codes in a separate document). Because the original dataset contains more than 50 unique PUMA codes for Arizona, I ran the regression again for increased interpretability of the relationship between geographic location and childbirth probability, all other variables held constant. I compressed the PUMA codes into major counties such as Phoenix, Maricopa, Tucson, Pima, and other districts using a reference table for 2010 PUMA codes found here. This time, living in Phoenix, Tucson or other area codes outside of Maricopa or Pima is significantly negatively associated with the likelihood of childbirth for married women compared to those who do not, with the statistical significance and direction of all other coefficients being transferable from the first regression model.

Table 2: Recoded PUMA Regression Results

| **Variable** | **OR** | **95% CI** | **p-value** |
| --- | --- | --- | --- |
| **College** |  |  |  |
| *yes* | 1.446 | 1.276, 1.640 | **<0.001** |
| **Hispanic Origin** |  |  |  |
| *yes* | 1.295 | 1.103, 1.521 | **0.002** |
| **Health Insurance** |  |  |  |
| *yes* | 1.014 | 0.828, 1.240 | 0.89 |
| **Employed** |  |  |  |
| *yes* | 1.064 | 0.815, 1.388 | 0.64 |
| **U.S. Native** |  |  |  |
| *yes* | 0.612 | 0.522, 0.718 | **<0.001** |
| **Age** | 0.895 | 0.890, 0.899 | **<0.001** |
| **Worked in past 12 months** |  |  |  |
| *yes* | 0.730 | 0.565, 0.944 | **0.017** |
| **At/Below Poverty Line** |  |  |  |
| *yes* | 1.288 | 1.056, 1.573 | **0.013** |
| **Arizona County** |  |  |  |
| *Maricopa* | — | — |  |
| *Other* | 0.824 | 0.712, 0.954 | **0.010** |
| *Phoenix* | 0.826 | 0.706, 0.967 | **0.018** |
| *Pima* | 0.944 | 0.778, 1.144 | 0.55 |
| *Tucson* | 0.666 | 0.511, 0.870 | **0.003** |
| Abbreviations: CI = Confidence Interval, OR = Odds Ratio | | | |

These findings highlight how educational attainment, employment, cultural background, and economic standing intersect to shape reproductive decisions for married women in Arizona. The higher fertility rates among Hispanic and foreign-born women align with broader demographic trends, where cultural norms and extended family structures may encourage earlier or higher fertility. Findings that childbirth is positively associated with completion of college yet negatively associated with current employment may reflect the underlying influence of economic stability on family outcomes. Having at least an associate’s degree and not needing to participate in the workforce both indicate a consistent source of income that may better allow married women to raise families. Alternatively, the strong correlation between poverty and childbirth raises questions about access to family planning resources in lower-income households. The observed regional differences in fertility further suggest that local economic conditions, housing costs, and healthcare access may influence childbirth rates across Arizona. These insights contribute to a deeper understanding of how economic and cultural forces drive fertility patterns and may inform policy discussions on statewide family support programs and workforce participation.