Michelle Wang

Prof. Raman

ECON 255

Food Spending and Income

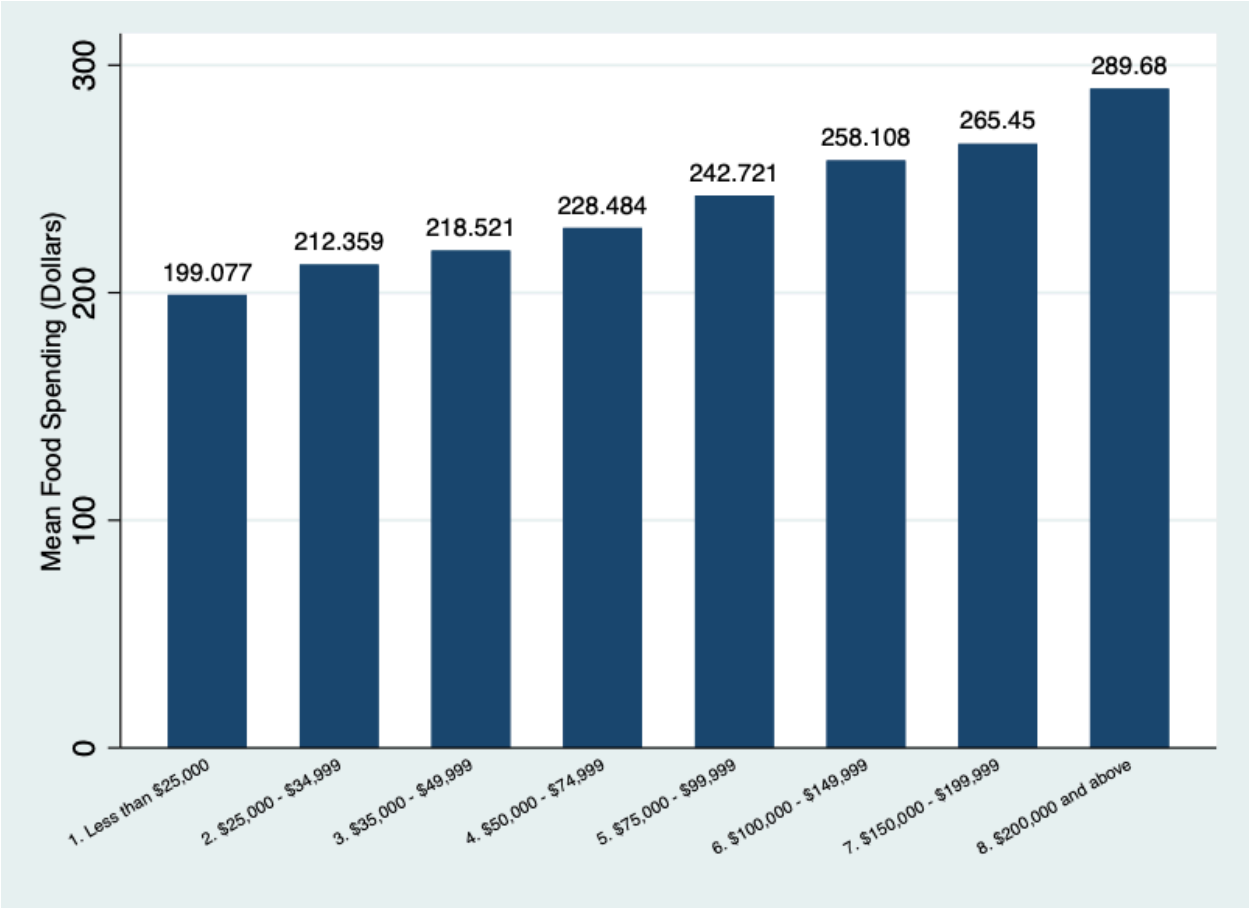
Introduction

This report explores variation in household spending on food and groceries along a series of explanatory variables. The primary relationship of interest is between household spending on food and groceries and household income. This relationship is important because we want to know if income is related to how much people spend on food, and indirectly, their food security.

We will also be exploring how 3 additional variables – the total number of people in the household, employment status, and housing tenure – are associated with food spending. The total household size is relevant because the more people there are, the more food they need to be fed and hence, more necessary food spending. Employment status is relevant because if someone in the household has a job, then they have a source of income, which may lead them to be able to afford and spend more on food and groceries. Housing tenure is relevant because the way that people are paying for their residence may affect how they spend on food and groceries. These relationships are important because they allow us to understand how household spending, financial well-being, and food security are related.

Existing literature has found that during COVID-19, people shifted away from full-service restaurant meals towards purchasing food and groceries. However, the increase in food spending was uneven for subpopulations (Okrent and Zeballos, 2022). Therefore, we are interested in exploring how food spending varies across different income groups. Additionally, prior studies found that greater spending on food is not associated with better health, controlling for socioeconomic status, which motivated our exploration of income as an explanatory variable (Tumin, 2022). Furthermore, larger households are associated with greater food insecurity in Nigeria, which motivated the inclusion of household size as an additional explanatory variable in our regression model for United States data (Adeoye, Afolaranmi, Ofili, Chirdan, Agbo, Adeoye, and Su, 2022). Finally, studies have shown that household food insecurity was higher among those at an economic disadvantage in Canada (Liu, Urquia, and Tarasuk, 2022). We are curious about this relationship in the United States, which is where the dataset is drawn from in this report.

Figure 1. Food Spending and Income



Note: Figure 1 is a bar graph of mean household spending on food and groceries in the past week measured in dollars on the y-axis and household income in 2021 measured in 8 groups on the x-axis. This figure shows our primary relationship of interest between mean food spending and our primary explanatory variable, income. Figure 1 shows a clear positive linear relationship between food spending and income, which matches our conditional expectation that food spending increases as income increases.

Data source: United States Census Bureau, Household Pulse Survey.

This dataset is appropriate for investigating this relationship because it covers a large geographical area and contains data from all over the country about food spending, income,

number of people in a household, etc. The breadth of this data makes it trustworthy. It was also compiled during COVID-19, which makes it relevant to the reasoning behind our investigation of our research question. Other papers have used data from the U.S. Bureau of Labor Statistics Consumer Expenditure Diary Survey, Survey of Consumer Finances (SCF), and the National Health and Nutrition Examination Survey (NHANES).

We will be running OLS regressions to analyze the relationship between mean food spending and our primary explanatory variable, income, as well as additional variables like the total household size, employment status, and housing tenure.

This data brief proceeds as follows. Section 2 describes the data used for evaluating the relationship of interest. Section 3 presents descriptive statistics and regression results. Section 4 concludes.

Data

The dataset that we are working with is from the latest fielding of the Census's Household Pulse Survey – started in April 2020 – which is designed to collect data on the social and economic impacts of COVID-19 on American households in all 50 US states and the District of Columbia. The survey records data on the following topics:

- Employment status
- Food security
- Housing security
- Physical and mental health
- Access to healthcare
- Education disruptions

- Household spending
- Various other measures of government interaction

The outcome variable that we will use for analysis is the household spending on food and groceries in a week in dollars, which we will refer to as food spending for short. Our primary explanatory variable is the total household income in 2021 before taxes in dollars, which we will refer to as income for short. Our 3 additional explanatory variables are the total number of people in the household, which we will refer to as household size for short, employment status within the past week as a binary variable, and housing tenure as a categorical variable.

Some data hazards that this dataset may be associated with are reinforcing existing bias and being difficult to understand. It may reinforce existing bias such as the stereotype that low-income households are food insecure. The dataset may also be "difficult to understand" because many of the variables don't have representative names, such as the food spending that we use for analysis in this report, which is represented as "tspndfood". Although there is documentation to help identify variables, this nonetheless makes the dataset confusing and difficult to work with.

Results

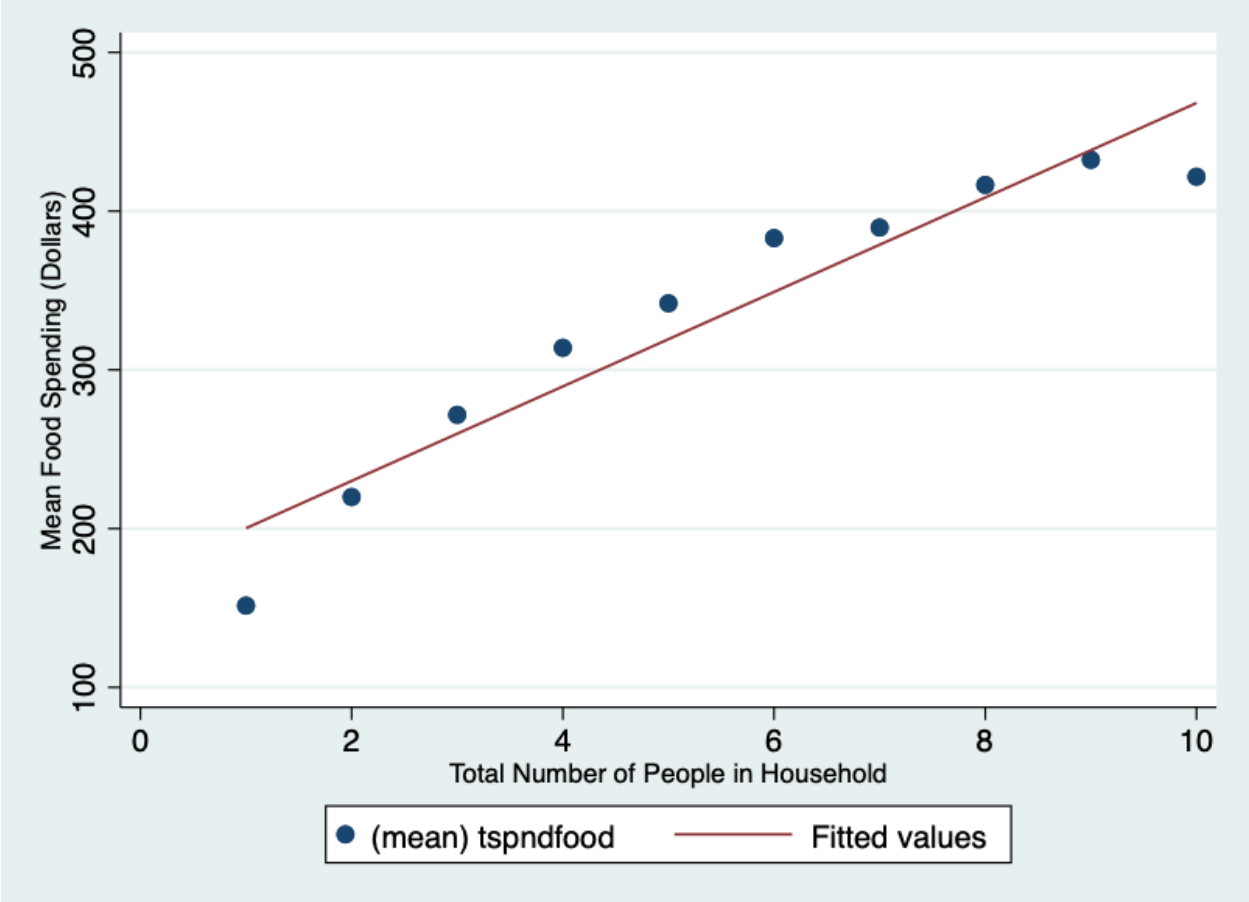
We ran a multivariate OLS regression of our outcome variable, food spending, on our primary explanatory variable, income, and our additional explanatory variables, household size, employment status, and housing tenure in order to explore how variation in food spending can be explained by these explanatory variables. The inclusion of these additional variables helps reduce omitted variable bias, in that they explain some of the variation in food spending that

income cannot explain. However, OLS only examines correlation – we cannot draw causal conclusions from this model.

Descriptive Statistics

The following figures show the relationships between mean food spending and our 3 additional explanatory variables: the total household size, employment status for the last 7 days, and housing tenure (omitting non-responses). These demonstrate that our additional variables are correlated with food spending and can explain variability in our conditional expectations of food spending that our primary explanatory variable, income, cannot.

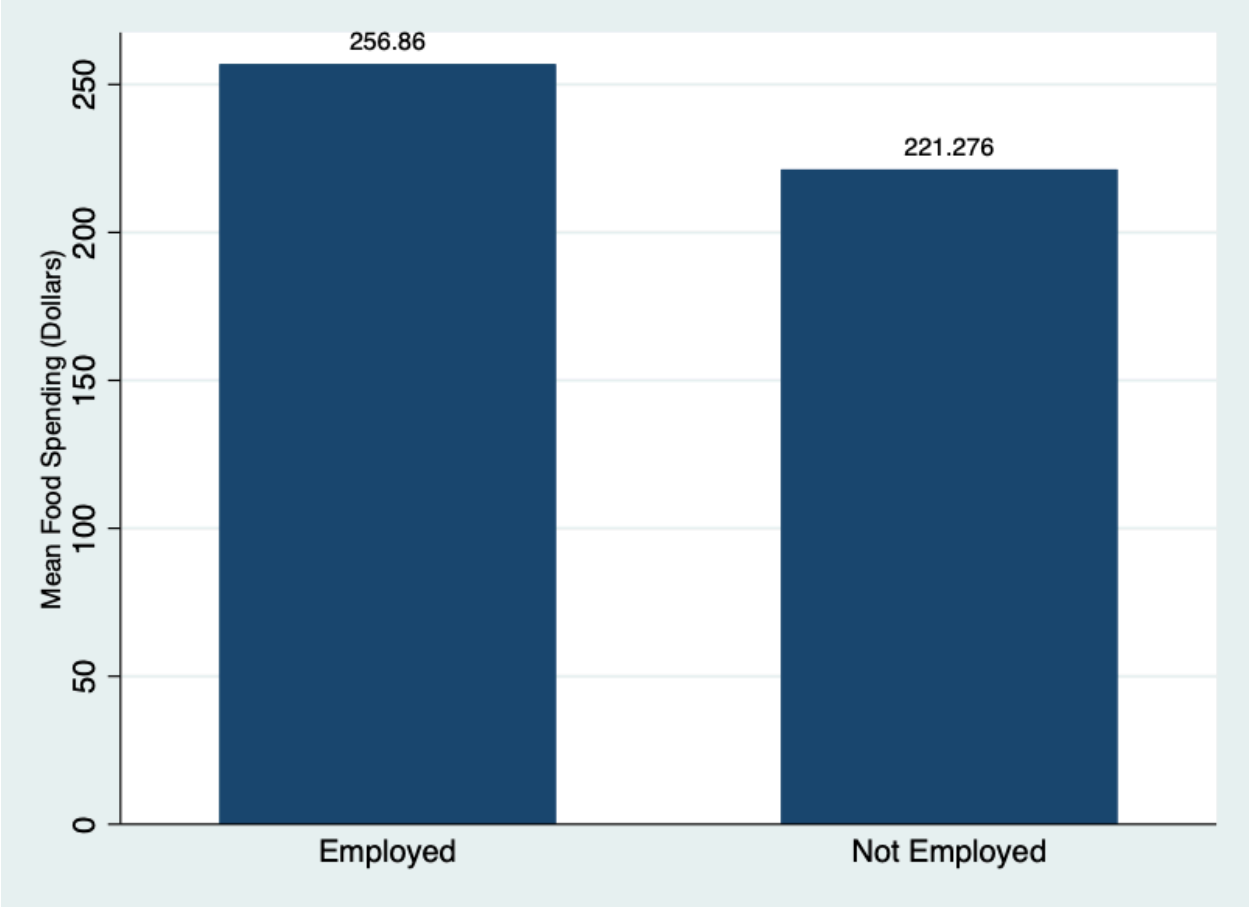
Figure 2: Food Spending and household size



Note: Figure 2 is a scatterplot of mean household spending on food and groceries in the past week measured in dollars on the y-axis and the total household size on the x-axis. Figure 2 shows a positive linear relationship between mean food spending and total household size.

Data source: United States Census Bureau, Household Pulse Survey.

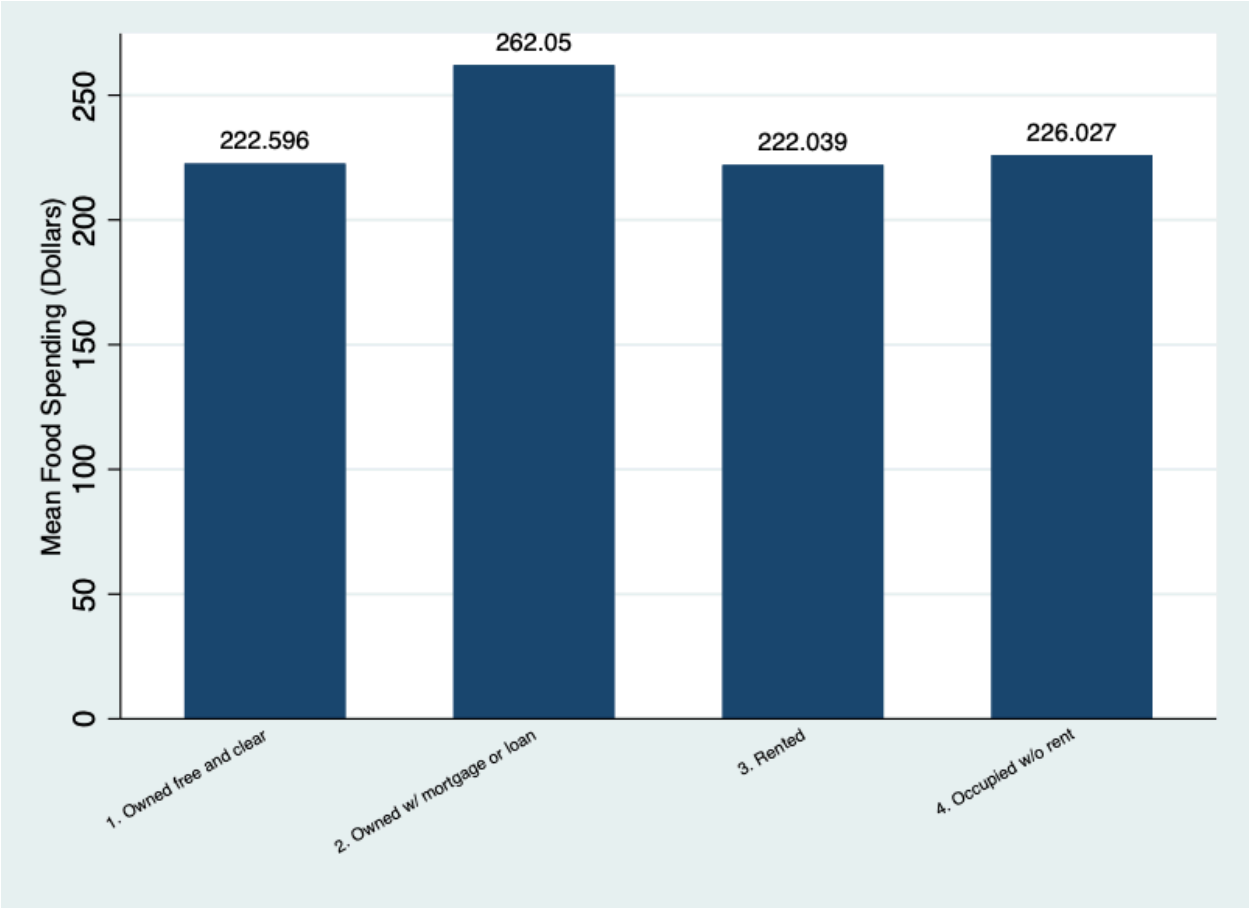
Figure 3: Food Spending and Employment Status



Note: Figure 3 is a bar graph of mean household spending on food and groceries in the past week measured in dollars on the y-axis and employment status in the last 7 days on the x-axis. Figure 3 shows that employed households spend more on food than their unemployed counterparts.

Data source: United States Census Bureau, Household Pulse Survey.

Figure 4: Food Spending and Housing Tenure



Note: Figure 4 is a bar graph of mean household spending on food and groceries in the past week measured in dollars on the y-axis and housing tenure on the x-axis. Figure 4 shows that people who rely on mortgages or loans to own their place of residence spend more money on food and groceries than people who own, rent, or occupy without rent in their place of residence.

Data source: United States Census Bureau, Household Pulse Survey.

Regression Results

Table 1 provides the results of our OLS regression analysis. In the section below, we will be discussing the results from Table 1.

Table 1: OLS Regression of Food Spending on Income, Number of People in

Food Spending				
	Column 1	**Column 2**	**Column 3**	**Column 4**
$25,000 - $34,999	13.28***	5.518	4.511	4.501
	(3.83)	(1.71)	(1.39)	(1.38)
$35,000 - $49,999	19.44***	10.93***	9.289**	8.780**
	(6.06)	(3.64)	(3.07)	(2.88)
$50,000 - $74,999	29.41***	17.26***	15.42***	14.82***
	(10.11)	(6.34)	(5.58)	(5.28)
$75,000 - $99,999	43.64***	26.98***	24.83***	23.88***
	(14.43)	(9.52)	(8.64)	(8.11)
$100,000 - $149,999	59.03***	35.19***	32.76***	31.52***
	(20.46)	(12.86)	(11.70)	(10.92)
$150,000 - $199,999	66.37***	39.62***	36.83***	35.44***
	(20.47)	(12.87)	(11.69)	(10.99)
$200,000 and above	90.60***	58.12***	55.04***	53.69***
	(28.07)	(18.87)	(17.48)	(16.54)
Household Size		42.01***	41.73***	41.43***
		(70.54)	(69.42)	(67.75)
Employed in Past Week (0/1)			6.381***	6.379***
			(4.38)	(4.37)
Owned with Mortgage or Loan				7.913***
				(4.69)
Rented				3.400
				(1.74)
Occupied without rent				10.22
				(1.47)
Constant	199.1***	105.4***	110.5***	106.7***
	(83.90)	(44.61)	(41.70)	(34.88)
N	56940	56940	56726	56533
se				
r2	0.0241	0.143	0.144	0.145

Note: Table 1 displays the results of our OLS regression of our outcome variable, household spending on food and groceries, on our explanatory variables, income, household size, employment status, and housing tenure. Column 1 displays the regression coefficient estimates of the regression of food spending on income, relative to the base category of less than $25,000. However, income may not explain all of the variation in food spending, as there may be other explanatory variables. Column 2 additionally accounts for the household size. Column 3 additionally accounts employment status. Column 4 additionally accounts for housing tenure, relative to the base category of "Owned free and clear". * Significant at 5% ** Significant at 1% *** Significant at 0.1%

Data source: United States Census Bureau, Household Pulse Survey.

1. Analysis of Income Coefficients

The results from our table show that on average, for respondents in an income category greater than or equal to $35,000, those in a higher income group are associated with higher food spending than those in the base income group, holding constant the household size, employment status, and housing tenure, relative to the base income group of those earning less than $25k.

In Column 1, the coefficients for the 7 income categories show the differences in dollar amount of food spending for each income category relative to the base category, holding the household size constant. The coefficients in Column 2 show these differences in food spending when holding employment status constant in addition to the household size. The coefficients in Column 3 show these differences in food spending when holding housing tenure constant in addition to employment status and the household size. We can statistically distinguish these coefficient estimates from 0 because the p-value produced from the t-statistic is less than our

significance level of 0.001. The significance level and sign do not change when holding constant each additional variable, which supports the idea that those in a higher income group are associated with higher food spending than those in the base income group.

However, this is not true for the income group $25,000 - $34,999, in which we found that food spending is not significantly different than the base income group of less than $25k when controlling for the additional variables. When we only look at the relationship between food spending and income, we see that on average, those in this income group of $25,000 - $34,999 are associated with $13.28 more in food spending than those in the income group under $25,000, and we can statistically distinguish this coefficient estimate from 0 because the p-value produced from the t-statistic is less than our significance level of 0.001. However, when we keep the additional explanatory variables constant, our coefficient estimates drop in both magnitude and significance level, meaning that those explanatory variables explained more of the variation in food spending than income could in this income group. When holding the household size constant, we see that on average, those in a higher income group are associated with $5.52 more in food spending than those in the base group. When holding the household size and employment status constant, we see that on average, those in a higher income group are associated with $4.51 more in food spending than those in the base group. When holding the household size, employment status, and housing tenure constant, we see that on average, those in a higher income group are associated with $4.50 more in food spending than those in the base group. Despite the fact that these are positive coefficients, we cannot statistically distinguish these coefficient estimates from 0 because the p-value produced from the t-statistic is greater than all of our significance levels.

2. Analysis of household size Coefficients

On average, every additional person in the household is associated with around $42 more in food spending, holding constant income in Column 2, income and employment status in Column 3, and income, employment status, and housing tenure in Column 4. We can statistically distinguish these coefficient estimates from 0 because the p-value produced from the t-statistic is less than our significance level of 0.001. The significance level and sign do not change when controlling for each additional variable, which supports the idea that on average, every additional person in the household is associated with higher food spending.

3. Analysis of Employment Status Coefficients

On average, households with employment in the last 7 days are associated with higher food spending relative to households without employment in the last 7 days, holding constant income and the household size in Column 3 and income, the household size, and housing tenure in Column 4. We can statistically distinguish these coefficient estimates from 0 because the p-value produced from the t-statistic is less than our significance level of 0.001. The significance level and sign do not change when controlling for each additional variable, which supports the idea that on average, unemployed households are associated with lower food spending relative to employed households.

4. Analysis of Housing Tenure Coefficients

On average, households living in housing that is owned with a mortgage or loan are associated with $7.91 more in food spending relative to the base group of households owned free and clear, holding income, the household size, and employment status constant in Column 4. We

can statistically distinguish this coefficient estimate from 0 because the p-value produced from the t-statistic is less than our significance level of 0.001, which supports the idea that on average, households living in housing that is owned with a mortgage or loan are associated with higher food spending relative to the base group of households owned free and clear.

However, households that rent or occupy without rent in their place of residence are associated with a difference in food spending relative to the base group that is not differentiable from 0. This means that there is not a significant difference in food spending between households that own housing free and clear and households that rent or occupy without rent in their place of residence.

Further Discussion

There is reason to believe that the additional variables of the total household size, employment status for the last 7 days, and housing tenure are relevant to food spending because the R-squared value of the regression increased with each additional variable, meaning that more of the variability in our predicted values of food spending was explained by our model with the addition of each variable.

Conclusion

Our analysis is limited by the fact that there are other variables that create variation in food spending that is not included in the dataset we used. Although our model was able to show that income, household size, employment status, and housing tenure can explain some of the variation in food spending, there are other factors that contribute as well, which is shown by

existing literature mentioned previously. Also, there are limitations to the dataset that we used – for example, our income variable was measured in categories instead of exact values, which may have caused inaccuracy in our model.

Finally, our data limits us from making causal claims about the relationship we've identified due to the dataset that was created from an observational study and not an experimental one; there was no experiment or treatment administered to the subjects being studied, which would have allowed us to quantify how food spending would change if the individual received a different treatment. Had there been a treatment administered, we would then have been able to observe differences between a control group and the treatment group to make causal claims about food spending.

References

Adeoye, P. A., Afolaranmi, T. O., Ofili, A. N., Chirdan, O. O., Agbo, H. A., Adeoye, L. T., & Su, T. T. (2022). Socio-demographic predictors of food security among rural households in Langai district in Plateau-Nigeria: A cross-sectional study. *The Pan African Medical Journal*, *43*, 36. https://doi.org/10.11604/pamj.2022.43.36.32704

Liu, R., Urquia, M. L., & Tarasuk, V. (2023). The prevalence and predictors of household food insecurity among adolescents in Canada. *Canadian Journal of Public Health = Revue Canadienne de Santé Publique*, *114*(3), 453–463. https://doi.org/10.17269/s41997-022-00737-2

Okrent, A. (n.d.). *COVID-19 Working Paper: Consumer Food Spending Changes During the COVID-19 Pandemic*.

Tumin, D. (n.d.). *Food spending and household health*.