

Examining the Impact of SNAP Policy Choices on Poverty Levels: A Comprehensive Analysis

Lauren Murphy

MA346 Data Science Final Project

This research paper is a continuation of a previous study examining the relationship between state-level policies of the Supplemental Nutrition Assistance Program (SNAP) and state-level poverty in United States. The Supplemental Nutrition Assistance Program (SNAP) is a government-sponsored initiative that offers nutritional aid to low-income individuals and families who meet the eligibility criteria. This final project incorporates additional statistical techniques and analysis to refine understanding of how specific SNAP policies impact poverty.

Exploring beyond the initial focus of the ten highest and lowest poverty states, this research searches for correlation between inclusive SNAP policy choices and lower poverty levels in all fifty states,. This study aims to identify whether policies such as simplified reporting options, lower average certification periods, online application availability, and call center availability correlate with lower poverty levels. Findings reveal there is a correlation between poverty and certification period, no correlation between poverty and call center availability, and uncertainty with correlation between online applications and poverty level.

Methodology / Data:

Datasets:

SNAP Policy Choices: This dataset, provided by John Snow Labs, outlines various state-level SNAP policies from 2003 to 2016, including certification periods, online application availability, and call center availability.

US County Level Poverty & Median Household Income: This dataset from the U.S. Census Bureau provides insights into poverty rates and household incomes across U.S. counties over the same period.

Data Cleaning and Preparation:

Selection of relevant columns: For analysis relevant to the research thesis, only specific columns were pulled from both datasets. In SNAP Policy Choices dataset: *Observation_Year*, *State_Abbreviation*, *Average_Certification_Period*, *Call_Center_Availability*, *Is_Simplified_Reporting_Option_Available*, *Online_Snap_Submission*, *Units_1_to_3_Months_Recertification*, *Units_4_to_6_Months_Recertification*, *Units_7_to_12_Months_Recertification*, *Units_13_Plus_Months_Recertification*.

In US County Level Poverty & Mean Household Income dataset: *Month*, *Year_Only Date_Re*, *Year*, *% Poverty Estimate (All Ages)*.

Time Frame Adjustment: Both datasets were trimmed to ensure consistency, focusing on the years 2003 to 2016. Moreover, the study concentrates on end-of-year figures, specifically for the month of December.

Handling Missing Data: Missing values (NaNs) were filled with 0, under the assumption that no data could be collected for those instances. For ‘simplified reporting’ cleaning, Nan values were given a 0.5 value in order to not skew the 0-1 rule.

Data Synchronization: For simplicity of date conversion, three columns were added to the excel dataset US County Level Poverty & Mean Household Income: *Month*, *Year_Only* and *Date_Re*, to split up the original *Year* column.

Column Renaming: For ease of merging datasets, the *State* column was standardized to *State_Abbreviation* across both datasets.

Certification cleaning: For simplicity, noise elimination models dropped x values less than 3%. While a larger minimal x value would be more appropriate, due to limited datapoints, this figure was shrunk to capture a more accurate OLS regression.

Statistical Techniques:

Descriptive Statistics and Exploratory Data Analysis were used to summarize the data and understand distribution patterns. This was done through quartile analysis and graphic visualizations.

Correlation Analysis and Regression Models were applied to explore the relationships between SNAP policies and poverty levels.

Data Rules:

Simplified Reporting: Responses from states are recorded monthly, and while some states may not have had simplified reporting at first, they may have switched to it within the same year. As a solution, states that have made simplified reporting available for at least 6 months in a year are given a value of 1 for that year to indicate their online reporting availability. This approach was implemented to establish a majority rule for determining simplified reporting score. Furthermore,

this methodology falls assumes a progression of technological advancement, where a state would not switch from 'yes' to 'no'.

Online Availability: Online availability can be classified into one of three responses - yes, no, or in select areas. To measure the level of availability in each state, a point system was used. For each month in each year, a state was assigned +1 point for "yes", -1 point for "no", and 0 points for "select areas". The scores were then compiled for each year, with a minimum score of -12 and a maximum score of 12.

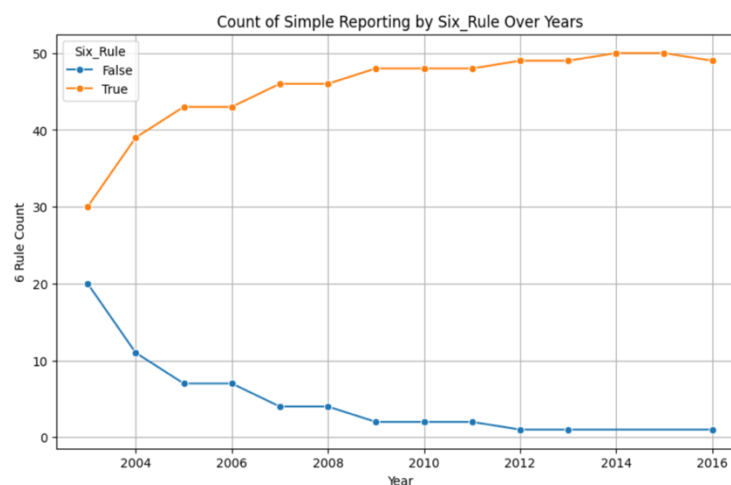
Call Center Availability: Call center availability follows the same scoring system as online availability, with no -1, yes 1, and in some areas 0.

Results:

Simplified Reporting:

Simplified reporting is a certification type that allows for reduced reporting requirements for recipients.

From 2003 to 2016, the number of states with simplified reporting increased from 30 to 49, while those without decreased from 20 to 1. Most states provided simplified reporting past 2003.

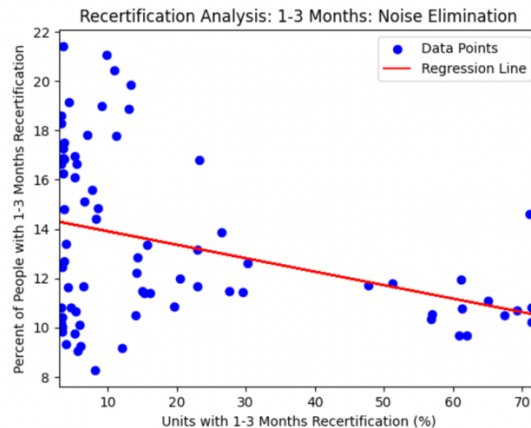
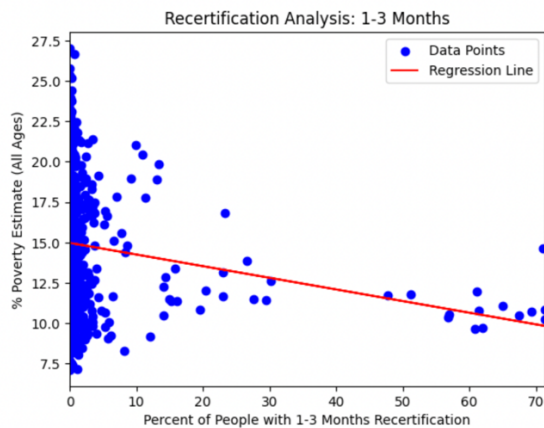


Certification Periods

Certification Periods refer to the process of determining an individual or household's eligibility for SNAP benefits.

1-3 Months (Fast Rate):

The original regression has a -0.0721 slope and P value of 0. When omitting outliers, the slope is -0.0547, with a P value of 0.02, proving a statistically significant relationship between poverty levels and 1–3-month recertification percentages.

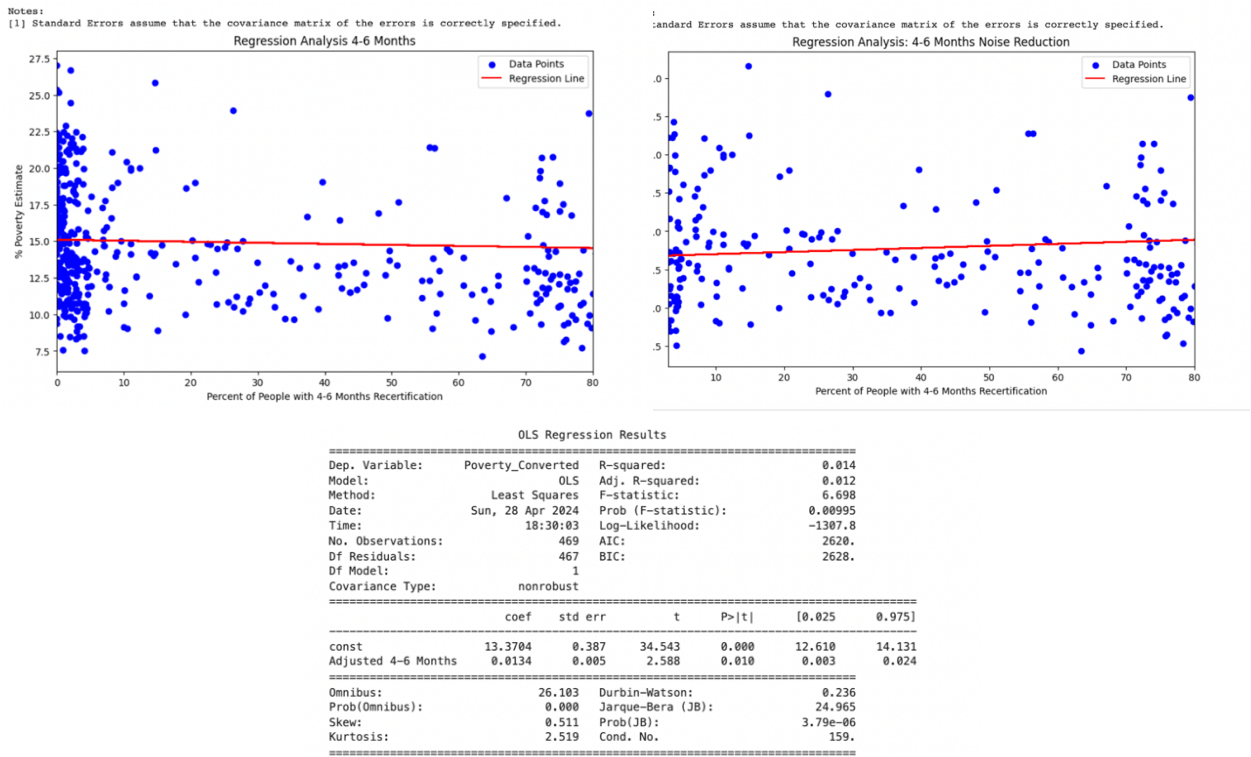


OLS Regression Results						
Dep. Variable:	Poverty_Converted	R-squared:	0.124			
Model:	OLS	Adj. R-squared:	0.112			
Method:	Least Squares	F-statistic:	10.44			
Date:	Wed, 24 Apr 2024	Prob (F-statistic):	0.00184			
Time:	17:16:49	Log-Likelihood:	-195.62			
No. Observations:	76	AIC:	395.2			
Df Residuals:	74	BIC:	399.9			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	14.4551	0.493	29.343	0.000	13.474	15.437
1_3_Converted	-0.0547	0.017	-3.231	0.002	-0.088	-0.021
Omnibus:	3.155	Durbin-Watson:	0.995			
Prob(Omnibus):	0.207	Jarque-Bera (JB):	2.624			
Skew:	0.339	Prob(JB):	0.269			
Kurtosis:	2.392	Cond. No.	38.9			

Regression Results for Noise Reduction Regression

4-6 Months (Moderate Rate):

The original regression has a -0.0721 slope and P value of 0. When omitting outliers, the slope is 0.0134, with a P value of 0.010, proving a statistically significant relationship between poverty levels and 4-6 month recertification percentages.

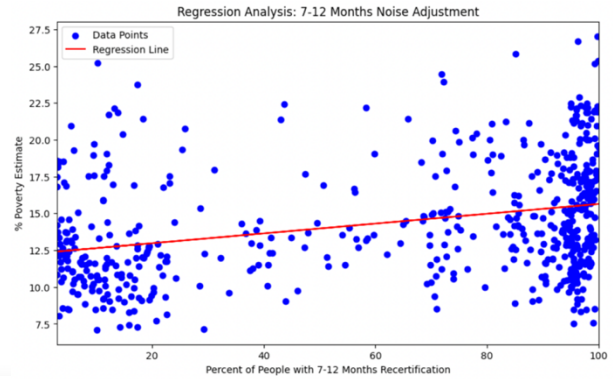
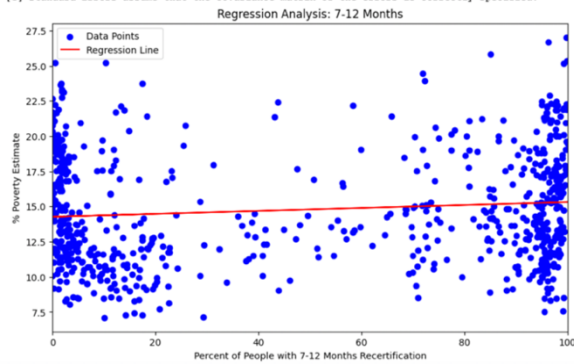


Regression Results for Noise Reduction Regression

7-12 Months (Slow Rate):

The original regression has a 0.0104 slope and P value of 0.004. When omitting outliers, the slope is 0.0333, with a P value of 0, proving a statistically significant relationship between poverty levels and 7–12-month recertification percentages.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

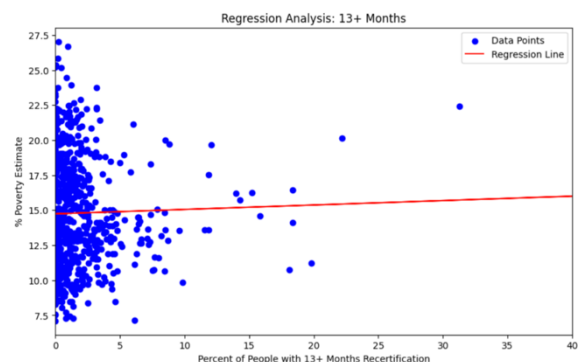
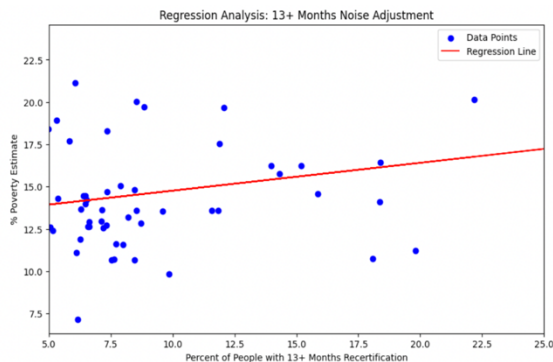


OLS Regression Results						
Dep. Variable:	Poverty_Converted	R-squared:	0.099			
Model:	OLS	Adj. R-squared:	0.097			
Method:	Least Squares	F-statistic:	61.84			
Date:	Mon, 29 Apr 2024	Prob (F-statistic):	1.91e-14			
Time:	12:56:34	Log-Likelihood:	-1550.2			
No. Observations:	565	AIC:	3104.			
Df Residuals:	563	BIC:	3113.			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	12.3084	0.311	39.550	0.000	11.697	12.920
7-12 adjusted	0.0333	0.004	7.864	0.000	0.025	0.042
Omnibus:	32.188	Durbin-Watson:	0.278			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	36.726			
Skew:	0.624	Prob(JB):	1.06e-08			
Kurtosis:	2.988	Cond. No.	144.			

Regression Results for Noise Reduction Regression

13 + Months (Slowest Rate):

The original regression has a 0.0316 slope and P value of 0.497. When omitting outliers, the slope is 0.1646, with a P value of 0.03, proving a statistically significant relationship between poverty levels and 13+ month recertification percentages.



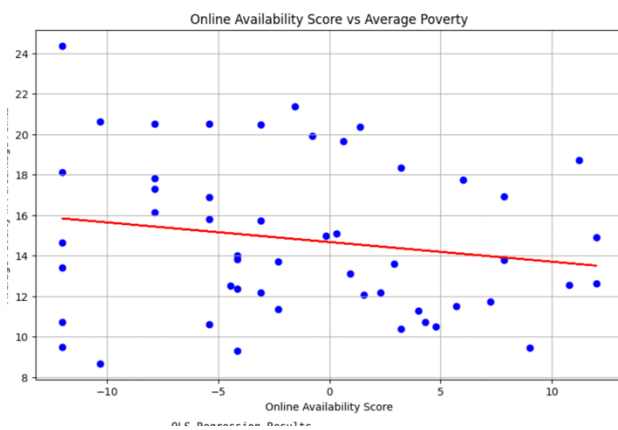
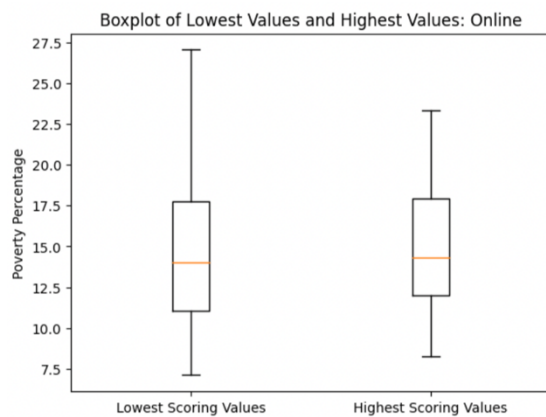
OLS Regression Results						
Dep. Variable:	poverty_adjusted		R-squared:	0.076		
Model:	OLS		Adj. R-squared:	0.068		
Method:	Least Squares		F-statistic:	9.515		
Date:	Sun, 28 Apr 2024		Prob (F-statistic):	0.00255		
Time:	19:12:49		Log-Likelihood:	-304.63		
No. Observations:	118		AIC:	613.3		
Df Residuals:	116		BIC:	618.8		
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	13.1211	0.465	28.231	0.000	12.201	14.042
13_adjusted	0.1646	0.053	3.085	0.003	0.059	0.270
Omnibus:	14.564		Durbin-Watson:		0.921	
Prob(Omnibus):	0.001		Jarque-Bera (JB):		15.890	
Skew:	0.835		Prob(JB):		0.000354	
Kurtosis:	3.665		Cond. No.		13.7	

Regression Results for Noise Reduction Regression

Online Availability

Online availability indicates if a state offers online application submissions.

States with the lowest score for online availability had a larger range of poverty values (7.12%-27.02%) compared to best scoring states (8.26%-23.30%). For states scoring between -12 and 12, no trend or correlation is present. When merging every year into one value to analyze the overall performance over 2003-2016, there is a weak negative correlation between a high score and poverty, with a coefficient of -0.0971 and a p value of 0.214.



Call Center

Call center indicates if a state offers call centers.

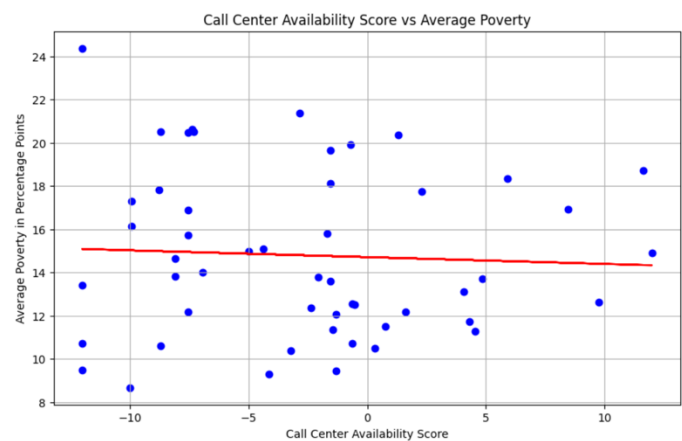
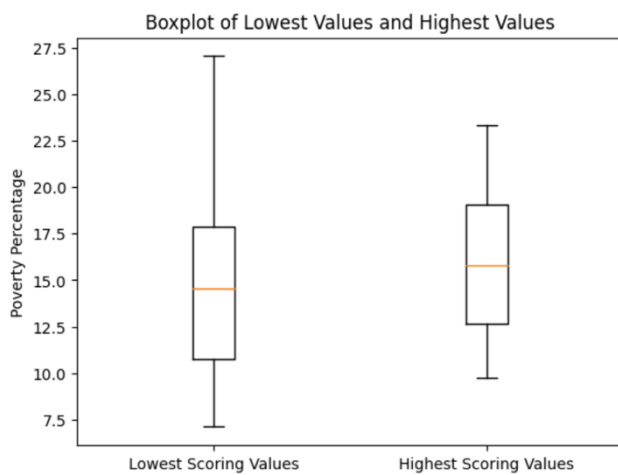
States with the lowest score for call center availability had a larger range of poverty values

(7.10%-27.02%). compared to best scoring states (9.82%-23.20). When merging every year into

one value to analyze the overall performance over 2003-2016, there is a weak negative

correlation between a high score and poverty, with a coefficient of -0.0314 and a p value of

0.723.



Discussion:

Simplified Reporting:

Null Hypothesis (H0): The use of simplified reporting methods has no effect on poverty levels. Increases or decreases in the use of simplified reporting methods do not affect poverty levels.

Alternative Hypothesis (H1): Poverty levels decrease as the use of simplified reporting methods increases. This suggests a negative correlation between the use of simplified reporting methods and poverty levels.

In the previous research, the mean of simplified reporting for the highest and lowest poverty states from 2003-2016 are 12.6 and 11.1 years. Additionally, the states above the poverty mean that offered simplified applications for 2003-2016 were 12.13 years, and the lack thereof was 1.87 years. For states below mean poverty, the offering and lack of simplified applications was 12 years and 2 years.

While my previous research contradicting the initial hypothesis that simplified reporting would correlate with lower poverty levels. This finding led to a more detailed analysis using the "six rule" for data aggregation, which provided insights into the widespread adoption of simplified reporting. Here, it is evident by 2006 most states adopted the simplified reporting rule. Therefore, it would be insufficient to collect poverty comparisons between these two subgroups, as the 'no' group is representative of a small fraction of cumulative states.

Simplified reporting being universal could indicate that expediting processes to receive SNAP is beneficial in more ways than one but would require further data to conclude why.

Average Certification Period:

Null Hypothesis (H0): There is no relationship between the certification period length and poverty levels. This means that changes in the certification period length do not affect poverty levels.

Alternative Hypothesis (H1): Poverty levels increase as the certification period length increases. This hypothesis claims that there is a positive correlation between the length of the certification period and poverty levels.

1-3 Months (Fast Rate): The higher the percentage of people in each state who receive food stamps quickly, the lower the poverty levels should be, as food stamps are more easily accessible.

On the other hand, when a lower percentage of people receive food stamps at a fast rate, they are either from the medium or long portion periods. The following reasons can explain the negative relationship between the two sections is represented by a slope of -0.0547, which is statistically significant with a P value of 0.02, after removing outliers. Vertical scatter on the left-hand side. The two sections have a negative slope, with a statistically significant value of -0.0547 ($P = 0.02$), even when outliers are excluded.

4-6 months (Moderate Rate): The higher the percentage of people in each state that receive food stamps at a moderate rate, the poverty levels will stay relatively the same, as it does not become a defining factor. On the other hand, when a lower percentage of people receive food stamps at a moderate rate, they are either from the short or long portion periods. The following reasons can explain the vertical scatter on the left-hand side. The relationship between these two sections is represented through a minimally positive slope.

7-12 months (Slow Rate): The higher percentage of people in each state that receive food stamps at a slow rate, the higher poverty levels should be, as it is less accessible.

On the other hand, when a lower percentage of people receive food stamps at a slow rate, they are either from the medium or short portion periods. The following reasons can explain the vertical scatter on the left-hand side. Excluding outliers, the relationship between these two

sections is represented through a positive slope of 0.033 and a p-value of 0. This indicates poverty increases for states in groups 7-12 compared to 4-6 months.

13+ months (Slowest Rate): The higher the percentage of people in each state who receive food stamps at the slowest rates, the higher the poverty levels should be, as food stamps are less accessible.

On the other hand, when a lower percentage of people receive food stamps at the slowest rate, they are either from the medium or short portion periods. The following reasons can explain the vertical scatter on the left-hand side. Excluding outliers, a positive relationship exists between these two sections, with a slope of 0.0316 and a p-value of 0.03. Similar to 7-12 months, poverty does not increase worse for the 13+ group than the 7-12 months.

The nature of poverty levels explains the low r^2 in each of these graphs. It would be inaccurate to expect that individual SNAP policies would significantly explain the variability in poverty levels, as SNAP is one of a multitude of financial assistance programs used in the United States.

Online availability/Call Center

Null Hypothesis (H_0): The number of call centers has no effect on poverty levels. Increases or decreases in the number of call centers do not affect poverty levels.

Alternative Hypothesis (H_1): Poverty levels decrease as the number of call centers increases.

This suggests a negative correlation between

Large P-values indicates that implemented scoring system fails to reject the null hypothesis for both parameters. Had there been a relation between poverty and scores that

support ease of access, a negative linear relationship should have been statistically significant between the independent and dependent variables for both call centers and online availability.

Notably, the availability of online and call center services has a mean of higher poverty than the lack thereof. While quartile distributions contradict the alternative hypothesis, there is a possibility of an overlap between states that fall within both categories. Further analysis could unravel why both see less concentrated poverty distributions.

Furthermore, the online availability's p-value of 0.021% may have statistical significance. Further testing could indicate a relationship between poverty and online applications. The call center's p-value of .723 alludes that a correlation with poverty is less likely.

Conclusion:

The findings of this extended study statistically prove that a correlation exists between certification times and poverty. A higher certification delay indicates increasing poverty levels in a state. Studying the implementation of policies to reduce certification perithe ods could improve the qualityof life for Americans by decreasing poverty levels.

While simplified reporting initially failed to reject the null hypothesis, further investigation concluded that the data was not an accurate representation due to small sampling size. Different datasets should be analyzed to investigate the reasons for the popularity of simplified reporting.

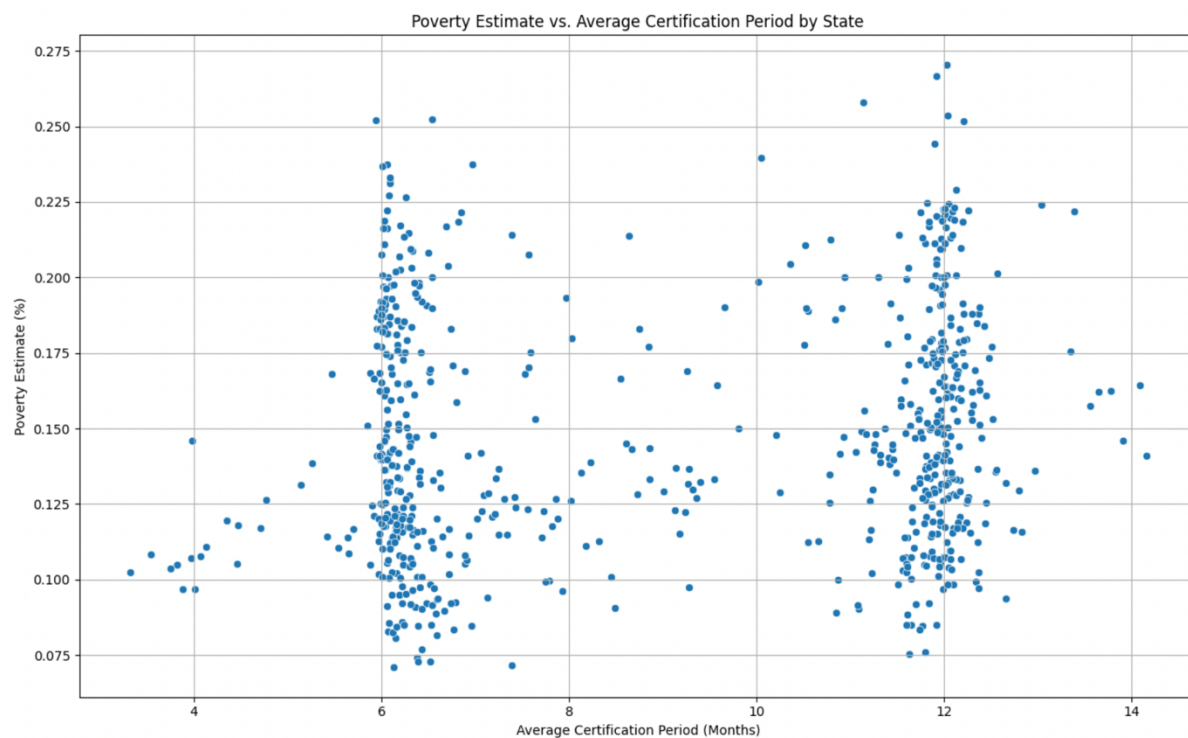
The interpretation of the correlation between poverty and online applications is subjective. Although a positive slope suggests that an increase in 'score' corresponds to a reduction in poverty, the significance of a p-value of 0.21 may vary depending on the evaluator. Different analysts might disagree on whether this p-value constitutes statistically significant

evidence. Finally, there could be no correlation between poverty and call centers, as the p-value was 0.723.

These findings provide valuable insights into the role of policy in shaping the accessibility of SNAP (Supplemental Nutrition Assistance Program). Policy research is significant and can provide evidence to advocate for better economic outcomes for the United States.

Appendix:

Analysis of average certification period versus poverty levels further proves higher certification times are correlated with higher poverty levels.



	index	pov_est_under9	pov_est_over9	poverty_comparison
1	mean	0.142488	0.153126	1.063875
2	std	0.040696	0.038915	-0.178145
3	min	0.071091	0.075333	0.424242
4	25%	0.111958	0.125686	1.372752
5	50%	0.133113	0.148068	1.495409
6	75%	0.175431	0.178177	0.274585
7	max	0.252217	0.270289	1.807229