Impoverishment & Aid in the United States

By: Michael Bemus

Background

- The Supplemental Nutrition Assistance Program (SNAP) provides food subsidies to those earning income at or below 130% of the federal poverty line.
- In 2020 was reported to have:
  - Provided benefits to ~39.9 million people.
    - Assisted ~20.5 million households.
  - Spent $74.2 billion in direct aid.
- Of those assisted:
  - 81% had income below the federal poverty line (pre-pandemic).
  - 36% had income less than half the federal poverty line
    - 54% of spending went to these individuals.

One of many types of subsidies available to struggling families in the US.

Poverty Line

How is it defined?

Federal Poverty Line:

- Single Individual: $16,770
- Per additional person: +$5,910
  - In 2022, ~37.9 million Americans estimated to be impoverished.

Supplemental Poverty Threshold:

- Looks at income, subsidies, taxes, and other expenses.
- Household of 4:
  - Renting: $34,518
  - Mortgage: $34,235
  - Owning w/o Mortgage: $28,909

Effects of Poverty

According to the US Office of Disease Prevention and Health Promotion (OASH):

- Impoverished individuals have an increased risk of developing mental illness or chronic disease.
  - Reduced life expectancy.
- Poverty is often area-based and out of individual control.
- Represents systemic issues:
  - Discrimination
  - Sustainable housing
  - Nutrition
  - Neighborhood safety
  - Education
  - Employment

2021 American Community Survey

Data collected by US Census Bureau from 3,092,744 individuals selected by random household sampling.

- We reduced this to 2,137,708 observations of 34 variables.

age	education	income	state tax	Supplemental health insurance premium	medicare part b	child care expenses	number of kids	SNAP	WIC subsidy
sex	marriage	agi	eitc	medical expenses	capped total expenses	mortgage status	cohabiting couples?	school lunch subsidy	housing subsidy
race	spm threshold	federal tax	fica	medical out of pocket	work expenses	number of adults	unidentified kids?	energy subsidy	group
hispanic?	spm resources	tax before credit	health insurance premium						

Group

Target Variable

Poor Group: 261,978 Observations

Aid Group: 427,112 Observations

No Group: 1,448,618 Observations

- Poor
  - Any individual classified as poor in data set.
    - Both federal poverty and SPM poverty.
- Aid
  - Any individual with nonzero SNAP, slunch, energy, house_sub, or WIC not in poverty group.
- No
  - All other observations.

Sample

- 50-30-20 Split

Train:

- Variable analysis
- Misclassification

Test:

- Retrain to test variable validity

Validation:

- Final misclassification rates.

Sample	Count
Train	1,068,854
Test	641,312
Validation	427,543

Methods

Principal Components, Discriminant Analysis, Logistic Regression

Principal Components

Unsupervised Learning Method

Analyze Group Variance Explanation

Compute eigenvectors of data correlation matrix and use as weights for creating new variables.

- Explain amount of variance equal to eigenvalue.
- Typically used to create variable combinations for dimension reduction.

Discriminant Analysis

Supervised Classification

Geometric Division

Create classification regions for which probability of a class is maximized based on prior probabilities.

- Assumes normality.

Linear:

- Assumes equal covariance matrices among groups.

Quadratic:

- Does not assume equal covariance matrices.

Logistic Regression

Supervised Classification

—

Classification by Odds

Optimize *k*-1 logit functions and assigns observations to group of maximum probability.

- Utilizes Bernoulli probability distribution for groups.

Metrics

Principal Components:

- Component Matching
  - Variables of coefficient greater than 1/sqrt(k) are considered significant.
  - **Perfect Match (PM):** Match component and position.
  - **Component Match (CM):** Match component.
- Cosine Similarity
  - Vector dot product divided by product of lengths.
  - **CS:** Base cosine similarity.
  - **aCS:** Cosine similarity of absolute-value vectors.
  - **wCS:** Weighted similarity based on component contribution.

Classification:

- **Accuracy:** Number correct divided by total observances.
- **Recall:** Number correct in Aid group divided by total Aid group in test.
- **Precision:** Number correct in Aid group divided by total predicted in Aid group.
- **F1:** 2*(Recall * Precision)/(Recall + Precision)
- Misclassification Analysis

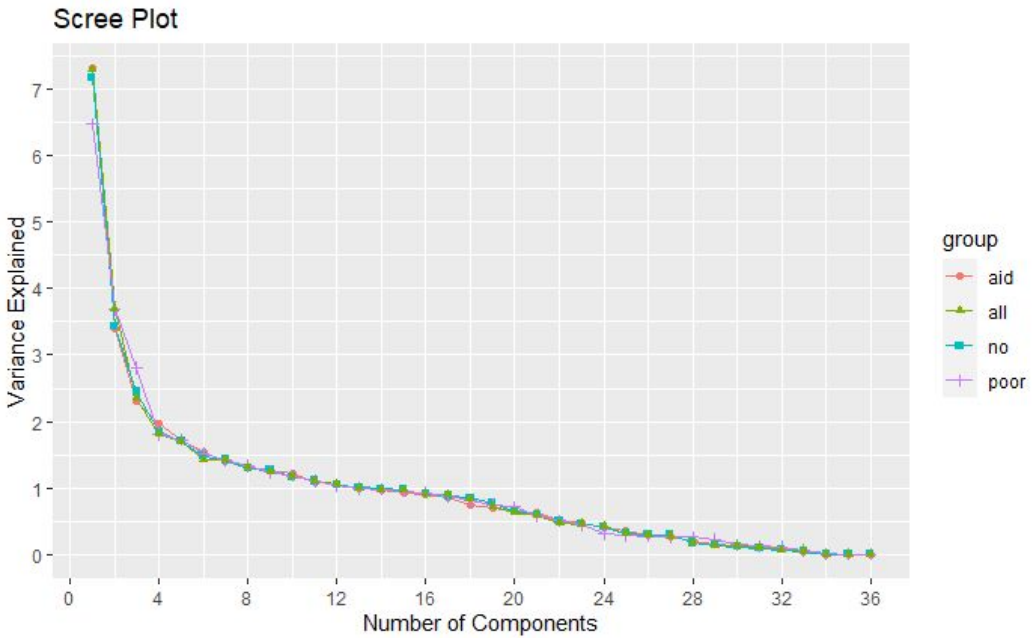
Results

Principal Components - Variance Explanation

Similar rates of explanation.

- PC 1: Income, Tax, Expenses
- PC 2: Household, Expenses, Tax
- PC 3: Medical Expenses
- PC 4: Marriage, Housing, Health Insurance

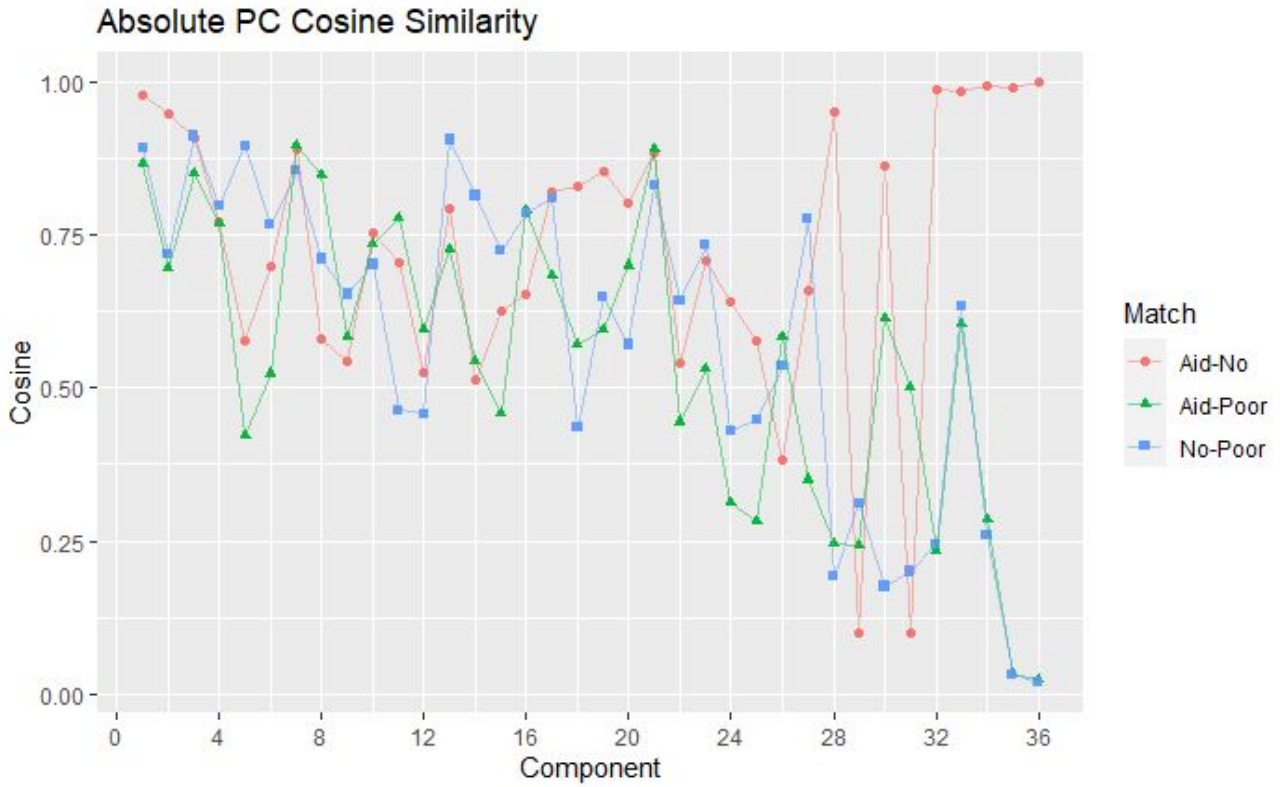
First 4 explain ~41% of total variance.



Principal Components - Similarity

Metric	Aid-No	Aid-Poor	No-Poor	Aid-Aid	No-No	Poor-Poor
PM	0.14634	0.06130	0.06654	0.78210	0.72793	0.51224
CM	0.54409	0.35632	0.38817	0.96109	0.94775	0.84746
CS	-0.12623	0.02001	-0.02923	-0.66538	0.94101	0.81058
aCS	0.72648	0.55025	0.58295	0.99745	0.99558	0.93428
wCS	0.78652	0.70013	0.75263	0.99794	1.00000	1.00000

Principal Components - Absolute Cosine Similarity



Discriminant Analysis - Confusion Matrices

LD		Predicted		
		Aid	No	Poor
Observed	Aid	38142	39523	7757
	No	10699	275555	3470
	Poor	3193	11143	38060

QD		Predicted		
		Aid	No	Poor
Observed	Aid	47582	20367	17473
	No	18797	253435	17492
	Poor	1593	3307	47496

LD-B		Predicted		
		Aid	No	Poor
Observed	Aid	52616	18007	14799
	No	23738	247281	18705
	Poor	2488	1869	48039

QD-B		Predicted		
		Aid	No	Poor
Observed	Aid	37986	28127	19309
	No	30841	226660	32223
	Poor	1947	2861	47588

Discriminant Analysis - Metrics

Highest Accuracy: LD

Highest Recall: LD-B

Highest Precision: LD

Highest F1: LD-B

- LD likely leverages size of No Group.
- QD assumptions best fit scenario.

	LD	LD-B	QD	QD-B
Accuracy	0.82274	0.81381	0.81516	0.73030
Recall	0.44651	0.61595	0.55702	0.44469
Precision	0.73302	0.66736	0.70002	0.53672
F1	0.55497	0.64043	0.62039	0.48639

Logistic Regression

Metrics:

- Accuracy: 0.91860
- Recall: 0.72552
- Precision: 0.93683
- F1: 0.81774

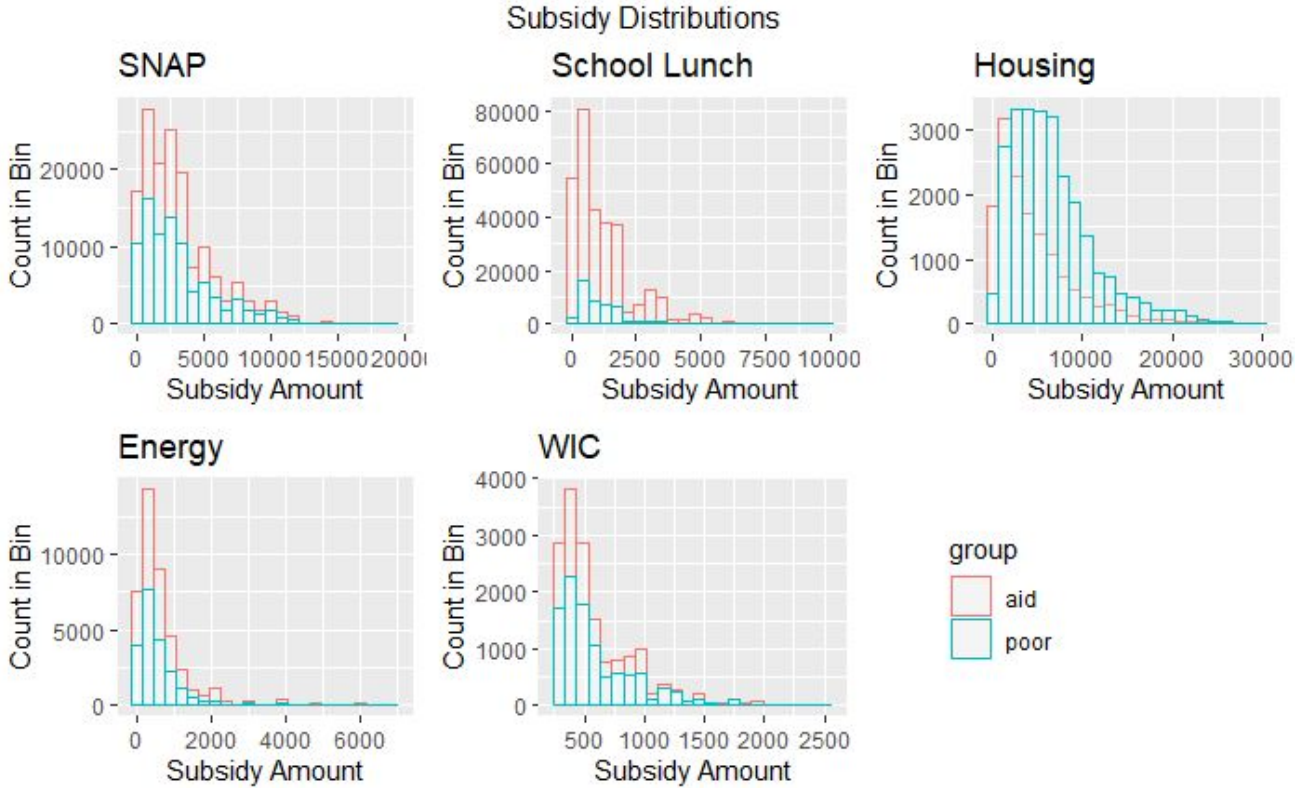
Highest performance on all metrics.

Tending to misclassify Aid Group into No Group.

Logistic		Predicted		
		Aid	No	Poor
Observed	Aid	61975	18064	5383
	No	2236	285668	1820
	Poor	1943	5357	45096

Conclusion

A Note From Exploratory Analysis



Conclusion

All modeling approaches implied the Aid Group was most similar to the No Group.

Limitations:

- School Lunch Subsidy inclusion
- Multicollinearity
- Normality
  - Possible Transformations
- Not statistically rigorous

Implications:

- No action should be taken.
- Cannot justify poor classification for aid group.

Further Research:

- Subsidy Analysis
- Aid Group Prediction

Thank You!

References

Benson, C. (2022). Poverty 2018 and 2021. United States Census Bureau, 1-20, https://statistical-proquestcom.ezproxy.library.pfw.edu/statisticalinsight/result/pqpresultpage.previewtitle?docType=PQSI&titleUri=%2Fcontent%2F2023%2F2316-15.76534.xml&accountid=11649

Bureau of Labor Statistics (2022). 2022 Research Supplemental Poverty Measure Thresholds. https://www.bls.gov/pir/spm/spm_thresholds_2022.htm#:~:text=As%20seen%20in%20C hart%201,for%20two%20adults%20two%20children).

Center for American Progress (2022). Data on Poverty in the United States. https://www.americanprogress.org/data-view/poverty-data/#tab1

Cronquist, K. & Eiffes, B. (2020). Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2020. Food and Nutrition Service (USDA), xvii-xxi, https://statistical-proquestcom.ezproxy.library.pfw.edu/statisticalinsight/result/pqpresultpage.previewtitle?docType=PQSI&titleUri=%2Fcontent%2F2023%2F1364-8.xml&accountid=11649

Richard, J. & Wichern D. (2007). Applied Multivariate Statistical Analysis (6th Edition). Pearson Education Inc.. Mowafi, M. & Khawaja, M. (2005). Poverty. Journal of Epidemiology and Community Health (1979-), 59(4), 260–264. http://www.jstor.org/stable/25570683 Office of the Assistant Secretary for

Planning and Evaluation (2023). Poverty Guidelines. ASPE. https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines

Office of Disease Prevention and Health Promotion (n.d.). Poverty. Healthy People 2030. https://health.gov/healthypeople/priority-areas/social-determinants-health/literaturesummaries/poverty

Romig, K. (2020). Social Security Lifts More People Above the Poverty Line Than Any Other Program. Center on Budget and Policy Priorities, 1-13, https://statistical-proquestcom.ezproxy.library.pfw.edu/statisticalinsight/result/pqpresultpage.previewtitle?docType=PQSI&titleUri=%2Fcontent%2F2022%2FR3834-37.xml&accountid=11649

United States Census Bureau (2023a). About the American Community Survey. https://www.census.gov/programs-surveys/acs/about.html

United States Census Bureau (2023b). ACS Supplemental Poverty Measures (SPM) Research Files: 2009 to 2019, 2021. https://www.census.gov/data/datasets/timeseries/demo/supplemental-poverty-measure/acs-research-files.html United States Census Bureau (2023c). Top

Questions About the Survey. https://www.census.gov/programs-surveys/acs/about/top-questions-about-the-survey.html