Final Project
Group D
Sultan Ali Alharbi, Faisal Alharbi, Farouk Shami, and Shawn Firth

Introduction

The purpose of our research is to study poverty in each county in the state of Utah and compare it with poverty in the United States. In our data set, which is the US Census, we considered different variables such as: Gender, Race, Average Income, Employment Rate, Self-Employment, and Poverty. We chose the variable of Poverty from the dataset to see where the Utah poverty rate falls in comparison with the national average; to compare the poverty rates between counties in the state; to look for poverty trends based on location in Utah; and to assess if poverty trends match the national poverty trends. We will be doing descriptive statistics, summary tables, and histograms in order to reveal outliers, study strengths and weaknesses in the dataset, examine the correlation between the different variables, and understand the regression analysis. Our questions are: What is the correlation between poverty and income within the state of Utah? How does the rate of poverty in the state of Utah compare to the national poverty level?

Hypothesis

Up until last week in class, based on our learning, we wanted to study race and poverty. Our null hypothesis was there is a strong correlation between poverty and race. Our alternative hypothesis was that there is no correlation between poverty and race. After we studied linear regression, we realized we wanted to change our study to poverty and income. This is because there are two clear variables that display whether poverty, the dependent variable, is dependent on income, the independent variable. Our null hypothesis is that there is a strong correlation between poverty and income. Our alternative hypothesis is that there is a weak correlation between poverty and income.

We assume that there will be a strong correlation between poverty and income. And we assume that the poverty rate in Utah is better than the national average, considering the unemployment rate in Utah is known to be less than the national average. We also assume however, that any income from employment does not always guarantee less poverty.

Data analysis

Our source of data is the US Census Demographic Data. Using the data from the State of Utah, we will show and compare the poverty level in relation to the variables of self-employed, family work, employed, unemployed and how they interact with income. We will compare the poverty and income levels for Utah and compare it with the national poverty and income levels.

The first type of graph we used is column/bar graphs, which gives us a simple way to understand the numbers from our data and to show the outliers, if any. The second type is a scatter graph. This shows the regression statistics, the correlations, and the outliers. The third type we used is a histogram shows how many of the same type of data occur in the same range as well as how frequently that data occurs. We also used a chart to display the numerical findings of the data, such as mean, median, and standard deviation.

The following is a numerical summary of the data acquired from the US Census.

Utah
Population = 2903379 people
Males = 1459229
Females = 1444150
Percentage of White = 85.62%
Percentage of Hispanic = 8.71%
Percentage of Black = 0.48%
Percentage of Native = 2.64%

Average Income = $54687.03
Percentage of Poverty = 13.27%
Percentage of Self-Employed = 5.7%
Percentage of Family Work = 0.3%
Percentage of Employed = 42.1%
Percentage of Un-Employed is 6.0%

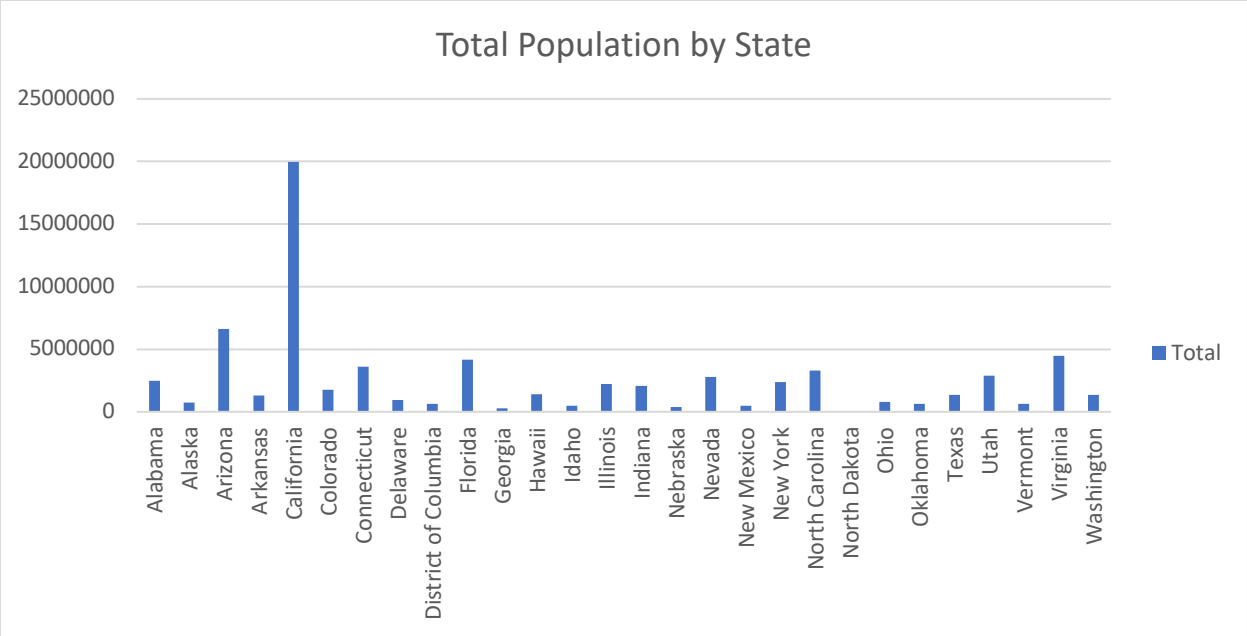
United States
Population = 70123781
Males = 34654052
Females = 35469739
Percentage of White = 72.01%
Percentage of Hispanic = 11.2%
Percentage of Black = 8.7%
Percentage of Native = 3.5%

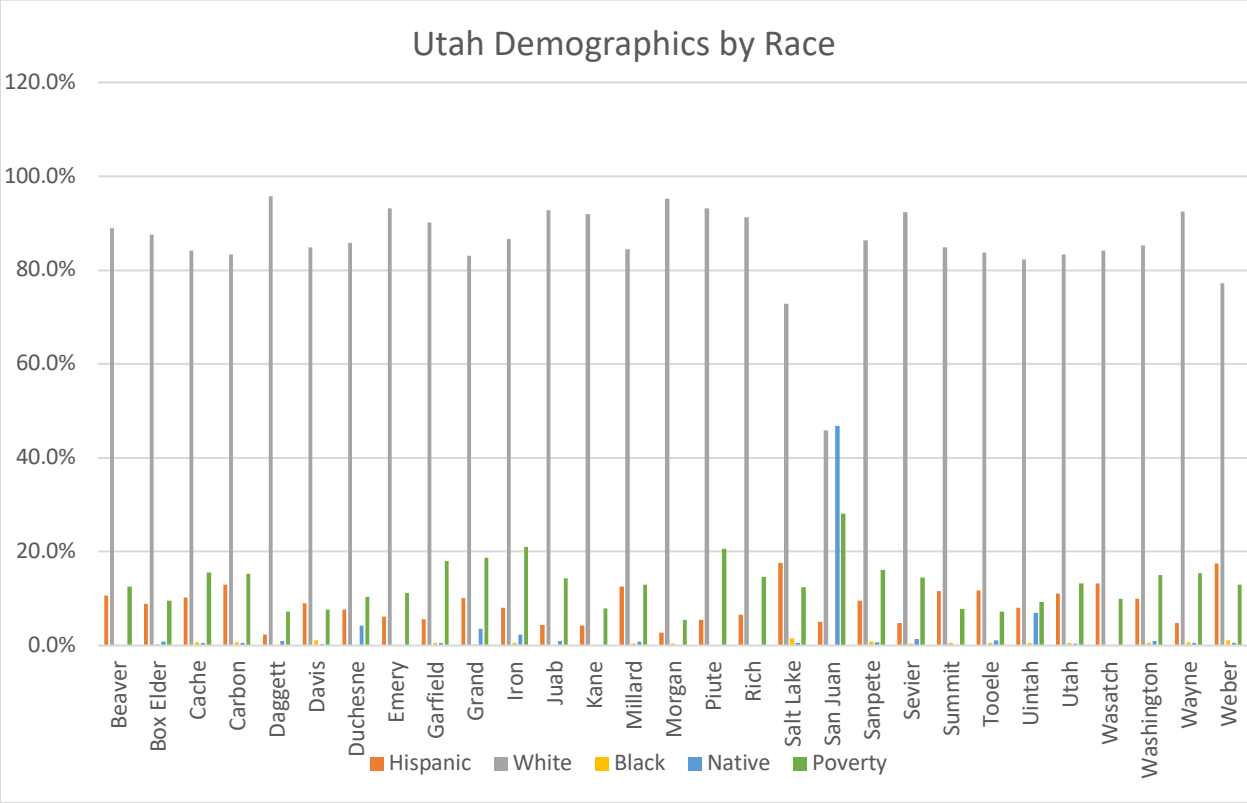
Average Income = $49026.78
Percentage of Poverty = 16.42%
Percentage of Self-Employed = 7.4%
Percentage of Family Work = 0.3%
Percentage of Employed = 43.1%
Percentage of Unemployed = 8.2%

The data is divided into the following variables: State, County, Total Population, Men, Women, Hispanic, White, Black, Native, Income, Poverty, Employed, Self-Employed, Family Work, Unemployment

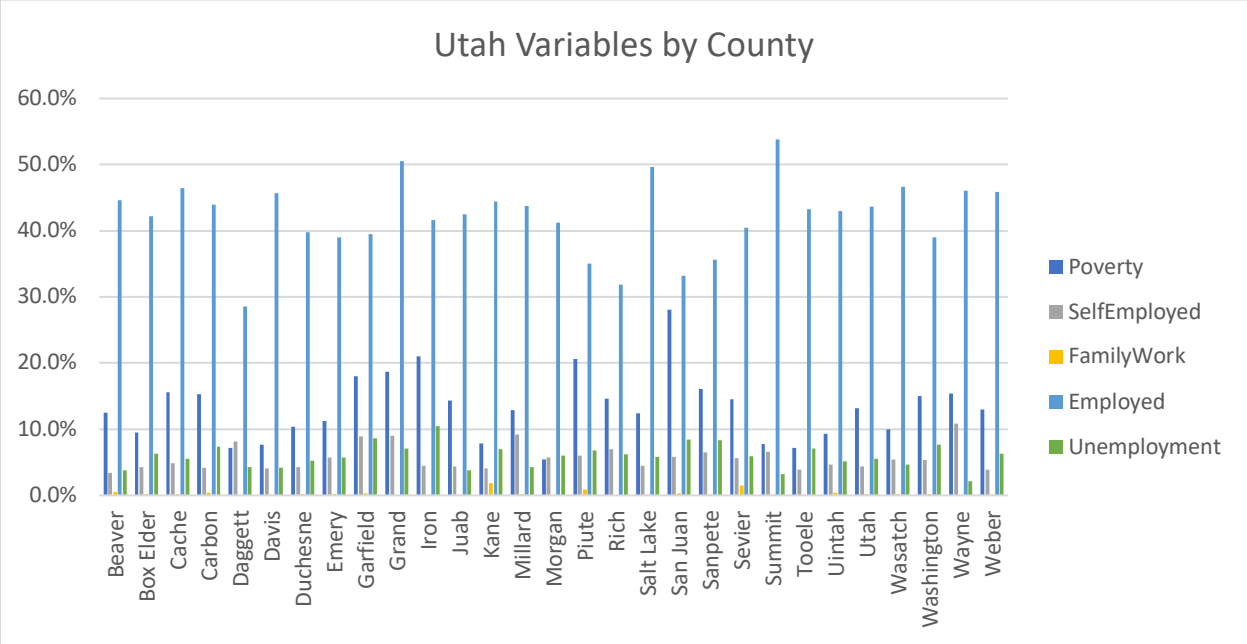
The original data is presented in real numbers (the total population, men, women, income, employed); and percentages (Hispanic, White, Black, Native, poverty, self-employed, family work, unemployment)

The graphs on population, gender, and race are for information purposes only, to understand the population we are studying.


Total Population by State
25000000
20000000
15000000
10000000
5000000
0
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Nebraska
Nevada
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Texas
Utah
Vermont
Virginia
Washington
Total


Utah Demographics by Race
120.0%
100.0%
80.0%
60.0%
40.0%
20.0%
0.0%
Beaver
Box Elder
Cache
Carbon
Daggett
Davis
Duchesne
Emery
Garfield
Grand
Iron
Juab
Kane
Millard
Morgan
Piute
Rich
Salt Lake
San Juan
Sanpete
Sevier
Summit
Tooele
Uintah
Utah
Wasatch
Washington
Wayne
Weber
Hispanic
White
Black
Native
Poverty

The graph comparing the counties in the state of Utah on variables of poverty, self-employed, family work, employed, unemployed clearly show the relationship between the variables. The counties with the highest employment rate have the lower relative poverty, with Summit County in the lead, then Grand, Salt Lake, and Wasatch. Counties with high unemployment rates also have significant poverty rates, namely Iron, Garfield, and Grand. Grand county has the second highest employment rate but also a high poverty rate. San Juan is the outlier with poverty almost equal to the employment rate but also a relatively low unemployment rate.



Strengths and weaknesses.

The data is thorough and allows researchers to study and compare many different factors. Based on the information provided in the data, the study can be broad (Ex: only comparing one factor in the entire nation) or narrow (Ex: compare the variables in one county or compare two variables).

The weakness of using data from the census is we have to take the numbers at face value and there is no examination into real issues about real people. So it is difficult to anticipate what the recommendations will be without knowing the root of poverty for a particular area in the nation or state.

Probability questions.

Awaiting questions from professor.

Correlations.

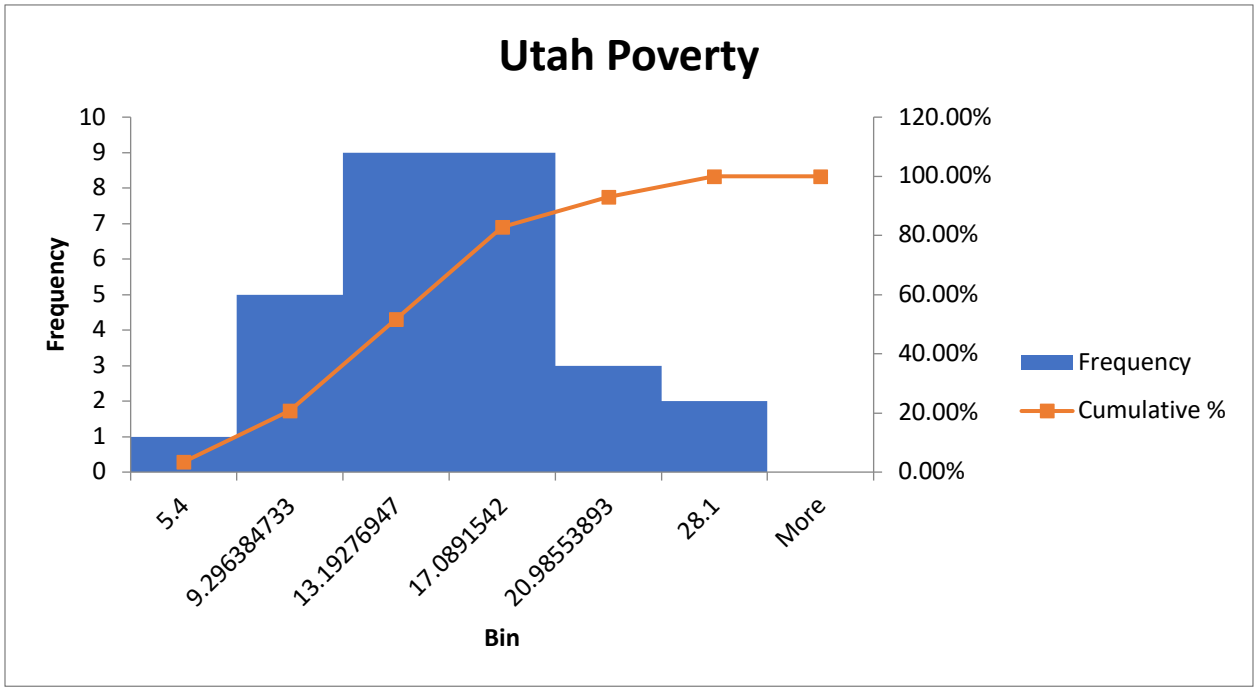
When comparing poverty and income across counties in the state of Utah, the correlation is -0.7367, which is a very weak correlation. When comparing poverty and income across in the United States, the correlation is -0.7523, which is also very weak.

Correlations between poverty and the other variables from the data are listed below:

Utah
Poverty & Un-Employment = 0.5158
Poverty & Employed = -0.08756
Poverty & Self-Employed = 0.2297
Poverty & Family Work = -0.0044

It is interesting to note that Poverty and Un-Employment have a very strong correlation and therefore, should be considered in future studies when trying to resolve issues related to poverty. Poverty & Employed have a slightly weak correlation. Poverty & Self-Employed have a slightly strong correlation. And Poverty & Family Work have a slightly weak correlation.

The chart below shows the frequency that the poverty is occurring based on the number of people in each range. (X-axis represents income and Y-axis represents poverty)



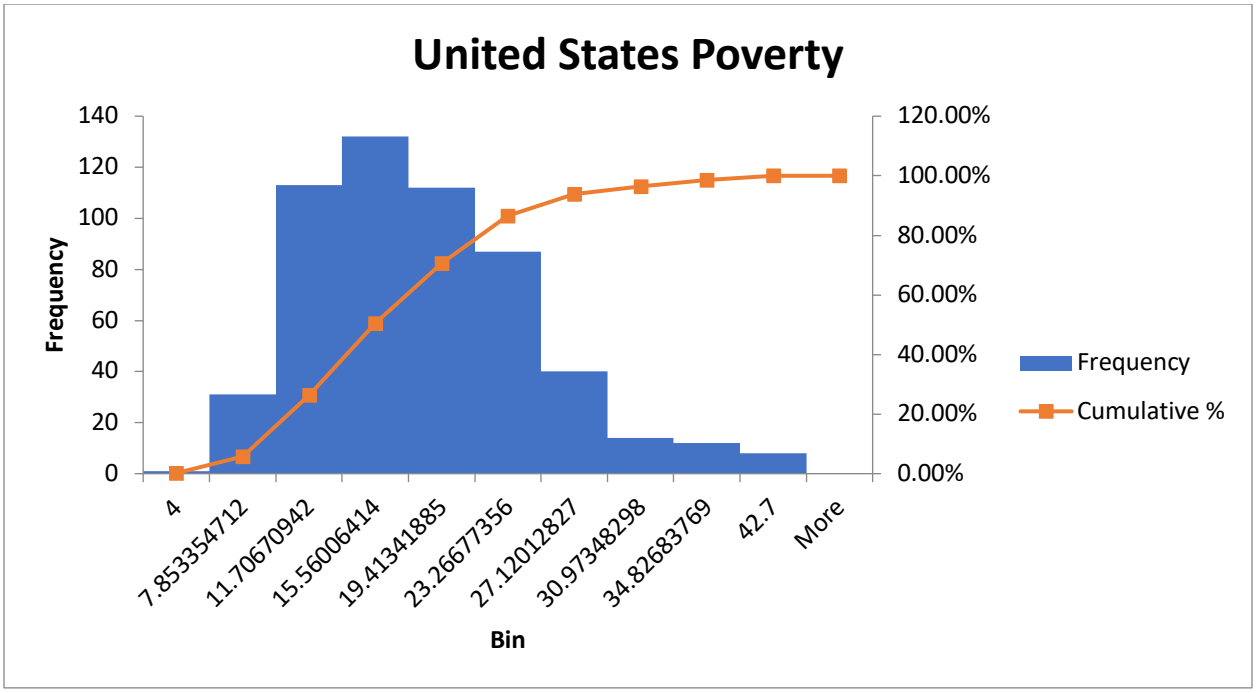
Poverty In Utah

Mean	0.132689655
Standard Error	0.00925063
Median	0.13
Mode	0.072
Standard Deviation	0.049816165
Sample Variance	0.00248165
Kurtosis	1.410679033
Skewness	0.880689278
Range	0.227
Minimum	0.054
Maximum	0.281
Sum	3.848
Count	29

US
Poverty & Un-Employment = 0.5951
Poverty & Employed = -0.03228
Poverty & self-Employed = -0.1187
Poverty & Family Work = -0.0205

It is interesting to note that Poverty and Un-Employment have a very strong correlation and therefore, should be considered in future studies when trying to resolve issues related to poverty. Poverty & Employed have a slightly weak correlation. Poverty & Self-Employed have a weak correlation, which is opposite of the results for Utah. And Poverty & Family Work have a slightly weak correlation.

The chart below shows the frequency that the poverty is occurring based on the number of people in each range. (X-axis represents income and Y-axis represents poverty)



Poverty in the US	
Mean	16.42018182
Standard Error	0.278694506
Median	15.45
Mode	12.9
Standard Deviation	6.535965509
Sample Variance	42.71884514
Kurtosis	0.903438137
Skewness	0.837722097
Range	38.7
Minimum	4
Maximum	42.7
Sum	9031.1
Count	550

Hypothesis testing.

We tested out hypothesis by inputting the information from the Census data set into Excel. We then used the pre-set equations to gather the regression analysis, correlation coefficients, and other statistics such as mean, standard error, median, mode, etc. Excel was also a valuable tool to display the data in graphs and charts.

Using the information from the regression analysis (R-square and P-value) and the correlation coefficients, it became clear to us that we needed to reject our null hypothesis in favor of our alternative hypothesis. The results also showed us information regarding poverty in relation to the other variables and helped us realize that there was valuable information that we needed to consider as well.

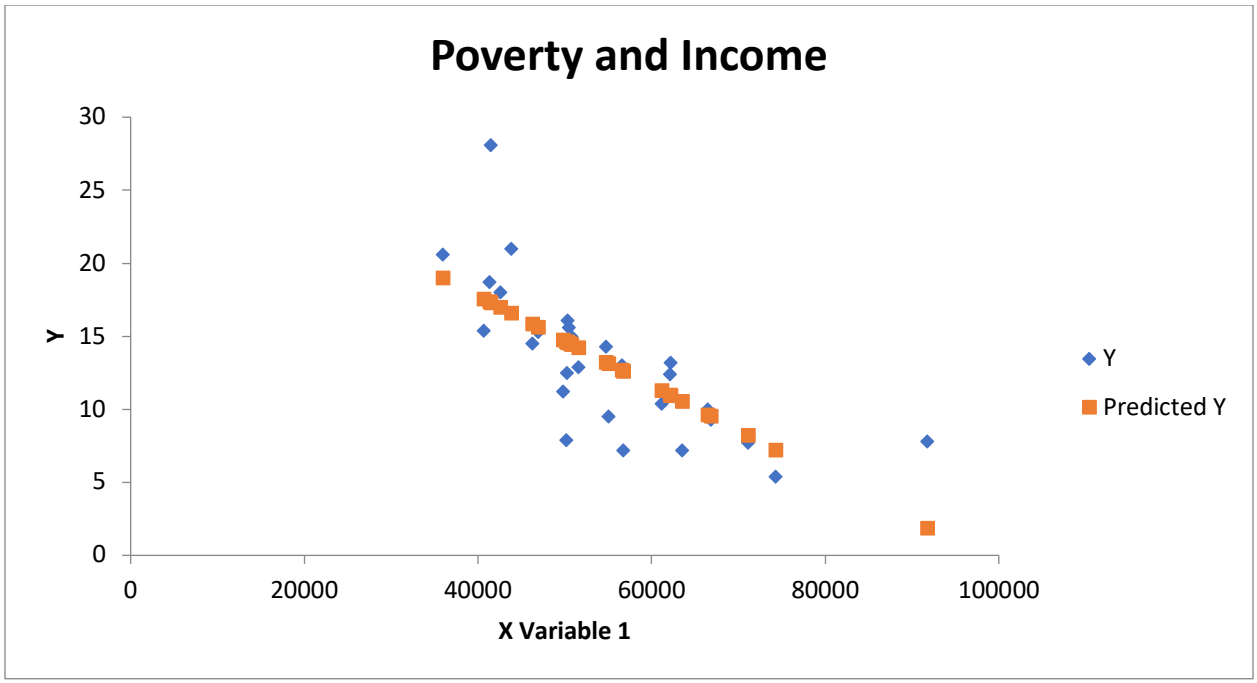
Regression analysis.

The R square results is above 50%, at 54%, meaning there is a relationship between the variables of poverty and income. The coefficients deal with the correlation, and -0.0003073 indicates that the variables move in opposite directions. The number -0.0003073 is close to zero and therefore shows a slightly weak correlation between the variables.

The p-value tests the hypotheses. From our data, the p-value 0.00000519384 (from the chart 5.1938E-06) shows strong evidence against the null hypothesis, so we reject the null hypothesis that there is a strong correlation between poverty and income. As we previously mentioned, the correlations support this rejection because there is a very weak correlation between poverty and income.

Regression Statistics	
Multiple R	0.736698322
R Square	0.542724418
Adjusted R Square	0.525788285
Standard Error	3.430492122
Observations	29

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	30.07945544	3.037156741	9.90382058	1.7463E-10
X Variable 1	-0.000307394	5.43017E-05	-5.6608619	5.1938E-06
	Lower 95%	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	23.84772456	36.31118633	23.8477246	36.3111863
X Variable 1	-0.000418812	-0.000195977	-0.0004188	-0.000196



Summary

Our null hypothesis is that there is a strong correlation between poverty and income. Our alternative hypothesis is that there is no relationship between poverty and income. Based on the information from our p-value and correlation coefficients, we reject the null hypothesis and accept the alternative hypothesis, which states that there is a weak correlation between poverty and income.

Research questions.

What is the correlation between poverty and all other variables in the dataset?

See the discussion on correlation above.

Which variable is considered a strong predictor for poverty (Un-Employment or self-employed)?

Unemployed

Strengths and weaknesses of project.

The strengths of our project are as follows:

- Data is thorough
- Data is generally reliable
- When plotting and graphing the data, trends are clear
- The data allows us to accurately assess our hypothesis

The weaknesses of our project are as follows:

- The data set reflects only one point in time, when the survey was produced

- The results only reflect the data from one point in time, so they are not as significant as comparing historical data with current data
- If we compared historical data with the current data, we could make suggestions and predictions for future changes, including growth and decline for stronger poverty correlations

Conclusion

While creating the data, we were surprised to see that there was a consistent correlation that was weak or very weak between poverty and income, since poverty is a variable that is dependent on income. We thought the correlation would be much stronger.

Considering the relationship between poverty and the other variables in the data set, it is now obvious to us that the strongest correlation is between poverty and unemployment. The data between these two variables tells us that there is work to be done in the state of Utah to decrease the rate of unemployment.

Recommendation

Based on our data and findings, we recommend that the state of Utah utilize resources to improve the rate of unemployment but to also promote the success of individuals who are self-employed.

On the national level, we propose a focus on un-employment in order to improve poverty rates or eliminate poverty.

Worksheet for Original Hypothesis

Our null hypothesis was there is a strong correlation between poverty and race. Our alternative hypothesis is that there is a weak correlation between poverty and race.

When comparing race and poverty across counties in the state of Utah, the correlations are as follows:

There is a correlation across all counties for poverty and race besides San Juan, which is the outlier.

For Whites and Poverty, the correlation is -0.4885, meaning there is a medium-strong correlation

For Hispanics and Poverty, the correlation is -0.0675, meaning a weak negative correlation

For Blacks and Poverty, the correlation is -0.0203, which is also a weak negative correlation

For Natives and Poverty, the correlation is 0.5669, meaning there is a medium-strong

correlation.

Previous Conclusion

- The data shows that minority group particularly Latinos are subjected to poverty. This might be attributed by the rate of discrimination against Latino. Hence there is a significant relationship between race and poverty.
- The level of economic inequality among the minority groups in the United States has become severe, immediate measures should be taken to change this situation.

Previous Recommendation

The government and other stakeholders should come up with measure of empowering Latinos.

- Promoting education in this group is another factor that will promote economic equality hence eliminating poverty.