Assignment #1: Module 2 Problem Set

Charles Titus

Grand Canyon University: RES – 866

September 11, 2019

**Module 2 Problem Set**

**2.11 Learning Activity**

1. Create standardized scores for all scale variables (price through alcohol).

DESCRIPTIVES VARIABLES=price cost calories sodium alcohol

/STATISTICS=MEAN STDDEV MIN MAX.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| Price per 6-pack | 35 | 1.59 | 7.19 | 3.0274 | 1.12343 |
| Cost per 12 Fluid Ounces | 35 | .27 | 1.20 | .5057 | .18732 |
| Calories per 12 Fluid Ounces | 35 | 68 | 175 | 139.77 | 24.447 |
| Sodium per 12 Fluid Ounces in mg | 35 | 6 | 27 | 14.66 | 6.145 |
| Alcohol by Volume (in %) | 35 | 2.30 | 5.50 | 4.5771 | .60298 |
| Valid N (listwise) | 35 |  |  |  |  |

Which beverages have positive standardized scores on every variable?

From what the data shows there were 5 drinks that had a positive standardized score on every variable. Those five were UA, UH, UL, UR, and SA.

What does this mean?

This means that the raw scores for the drinks are above or greater than the mean.

2. What is the most extreme z-score on each variable?

Price 3.70524 (SA)

Cost 3.70636 (SA)

Calories -2.93581 (UNR)

Sodium 2.00859 (PF and PJ)

Alcohol -3.77651 (UNR)

What is the most extreme z-score across all variables?

The most extreme was:

Alcohol -3.77651 (UNR)

3. What beverage is most typical of all beverages, that is, has z-score values closest to 0 for these variables?

The drink that is the most typical is UIR because the z scores for this drink is -.03332, -.03050, - .19518, -.59514, -.62547. These were the closet values to 0 of all of the drinks. As a result of this it is the most typical beverage.

**3.9 Learning Activity**

1. Run the Frequencies procedure on the following variables: sex, wrkstat (Labor Force Status), paeduc (Father’s highest degree), and satjob (Job or Housework). What is the scale of measurement for each?

The scale of measure for each variable is as follows:

* Sex – nominal
* Wrkstat (Labor Force Status) – nominal
* Paeduc (Father’s highest degree) – nominal
* Satjob (Job or Housework) – nominal

Request appropriate summary statistics and charts:

**Frequencies:**

[DataSet1] C:\Users\ctitus\Downloads\Census (1).sav

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Statistics** | | | | | |
|  | | RESPONDENTS SEX | LABOR FRCE STATUS | FATHERS HIGHEST DEGREE | JOB OR HOUSEWORK |
| N | Valid | 2023 | 2021 | 1553 | 1536 |
| Missing | 0 | 2 | 470 | 487 |

**Frequency Table**

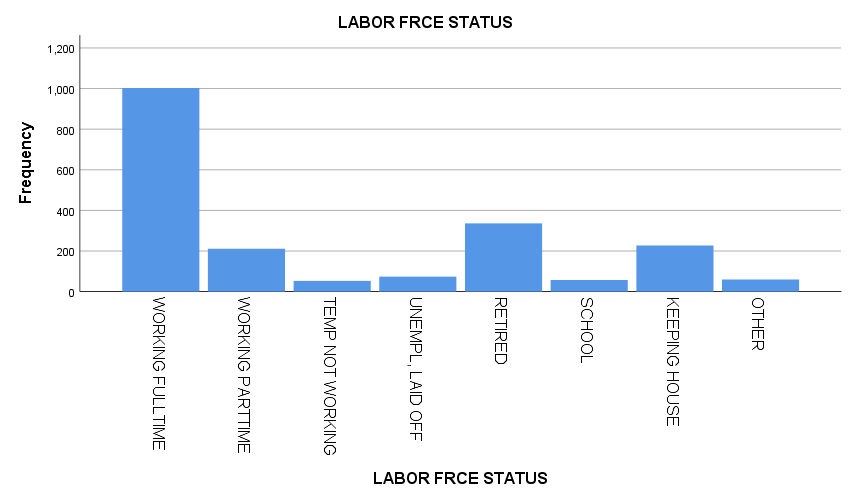
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **RESPONDENTS SEX** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | MALE | 929 | 45.9 | 45.9 | 45.9 |
| FEMALE | 1094 | 54.1 | 54.1 | 100.0 |
| Total | 2023 | 100.0 | 100.0 |  |

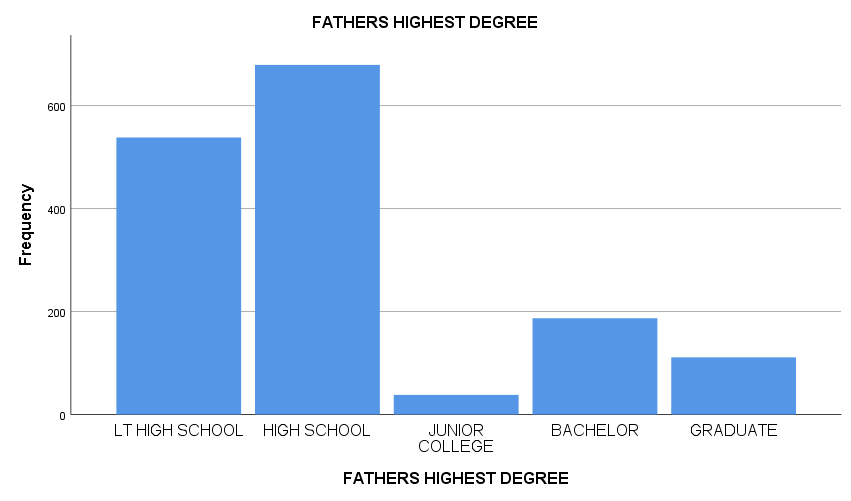
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **LABOR FRCE STATUS** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | WORKING FULLTIME | 1003 | 49.6 | 49.6 | 49.6 |
| WORKING PARTTIME | 211 | 10.4 | 10.4 | 60.1 |
| TEMP NOT WORKING | 53 | 2.6 | 2.6 | 62.7 |
| UNEMPL, LAID OFF | 74 | 3.7 | 3.7 | 66.4 |
| RETIRED | 336 | 16.6 | 16.6 | 83.0 |
| SCHOOL | 57 | 2.8 | 2.8 | 85.8 |
| KEEPING HOUSE | 227 | 11.2 | 11.2 | 97.0 |
| OTHER | 60 | 3.0 | 3.0 | 100.0 |
| Total | 2021 | 99.9 | 100.0 |  |
| Missing | NA | 2 | .1 |  |  |
| Total | | 2023 | 100.0 |  |  |

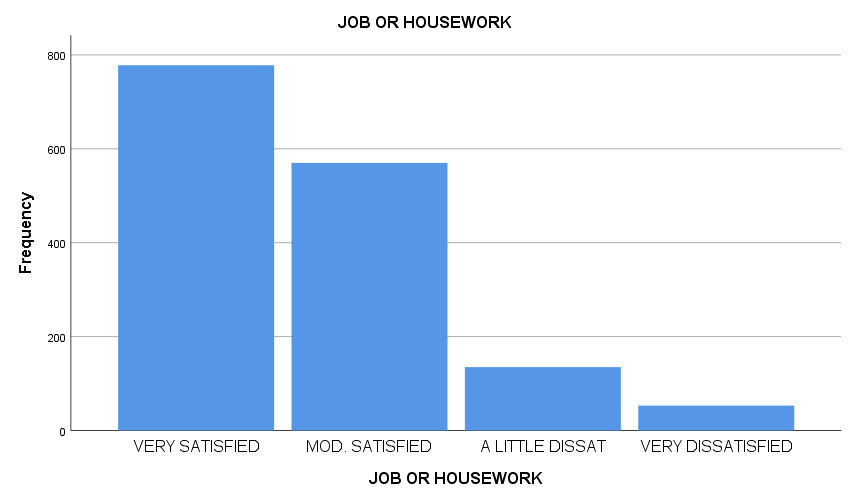
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **FATHERS HIGHEST DEGREE** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | LT HIGH SCHOOL | 538 | 26.6 | 34.6 | 34.6 |
| HIGH SCHOOL | 679 | 33.6 | 43.7 | 78.4 |
| JUNIOR COLLEGE | 38 | 1.9 | 2.4 | 80.8 |
| BACHELOR | 187 | 9.2 | 12.0 | 92.9 |
| GRADUATE | 111 | 5.5 | 7.1 | 100.0 |
| Total | 1553 | 76.8 | 100.0 |  |
| Missing | IAP | 354 | 17.5 |  |  |
| DK | 109 | 5.4 |  |  |
| NA | 7 | .3 |  |  |
| Total | 470 | 23.2 |  |  |
| Total | | 2023 | 100.0 |  |  |

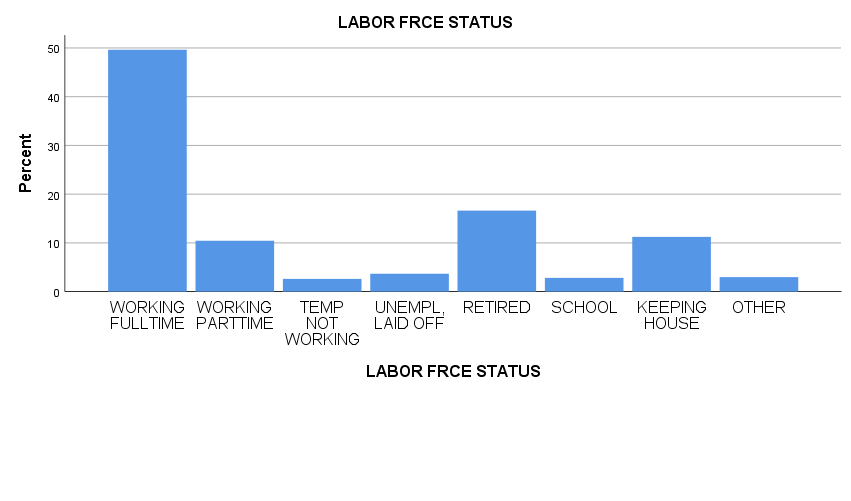
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **JOB OR HOUSEWORK** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | VERY SATISFIED | 778 | 38.5 | 50.7 | 50.7 |
| MOD. SATISFIED | 570 | 28.2 | 37.1 | 87.8 |
| A LITTLE DISSAT | 135 | 6.7 | 8.8 | 96.5 |
| VERY DISSATISFIED | 53 | 2.6 | 3.5 | 100.0 |
| Total | 1536 | 75.9 | 100.0 |  |
| Missing | IAP | 453 | 22.4 |  |  |
| DK | 19 | .9 |  |  |
| NA | 15 | .7 |  |  |
| Total | 487 | 24.1 |  |  |
| Total | | 2023 | 100.0 |  |  |

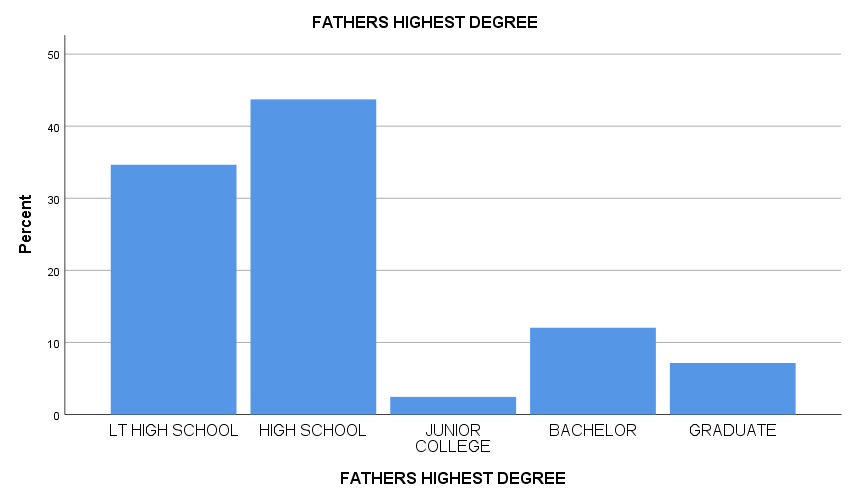
Bar Charts:

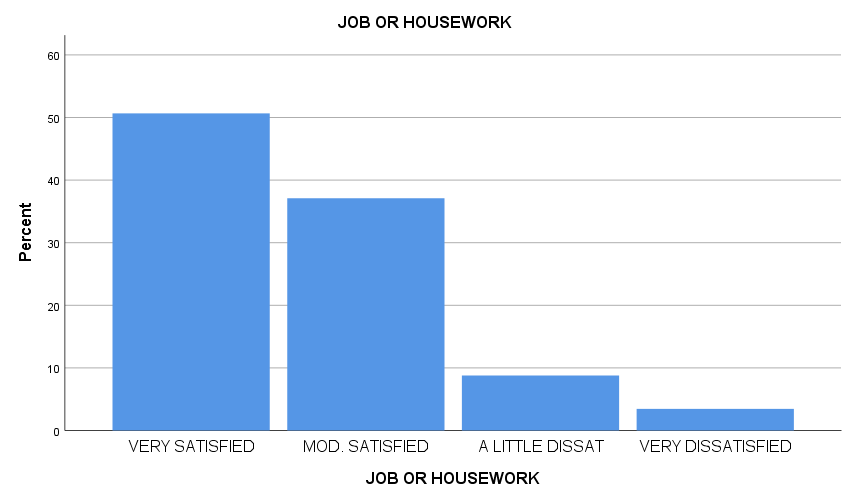












2. For which of these variables is it appropriate to use the median? What conclusions can you draw about the distributions of these variables?

It is appropriate to use the median for the parent education variable. The reason why this is okay to do is because this variable is actually measured with numeric values.

3. What percent of respondents have a bachelor’s degree, or higher?

14.7%

What percent of respondents are working?

60%

4. How might you combine some of the categories of wrkstat to insure that there are a sufficient number of respondents in each category?

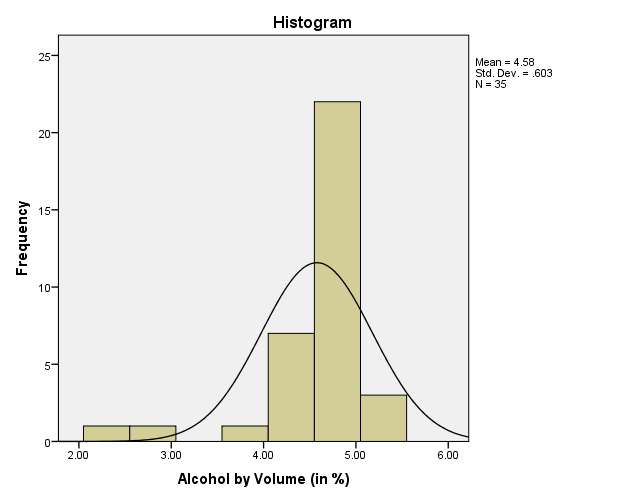
The way to combine some of the categories of wrkstat is to do employed, unemployed, temporally out of work, and school.

**4.18 Learning Activity**

1. Run **Frequencies** on the variable *alcohol*, requesting the summary statistics median and mean, plus a histogram with a superimposed normal curve. Suppress the display of the frequency table.

[DataSet1] C:\Users\ctitus\Downloads\Drinks (3).sav

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| Alcohol by Volume (in %) | | |
| N | Valid | 35 |
| Missing | 0 |
| Mean | | 4.5771 |
| Median | | 4.7000 |



1. What is the value of value of *alcohol* that splits the distribution in half?

The value of the splits the distribution in half looks to be 4.7.

Is the median the same as the mean?

No. the median and the mean are not the same.

Which value is lower?

The mean value is lower and it is 4.5771 which turns out to be lower than the median.

What does that tell you about the shape of the distribution of *alcohol*?

This tells me that the data in question has a negative skew and the histogram shows that above.

1. Does the histogram verify your description of the distribution of *alcohol*?

Yes, the reason is because the distribution of the data actually shows that most of the data in question is clustered on the left. Thus meaning that it is indeed supported.

How does it differ from a normal distribution?

The way that it differs from a normal distribution is that the tail is actually longer on the left side. Whereas in a normal distribution from my understanding is that the tails are more equal on either side of the peak.

1. Run **Descriptives** to obtain default statistics for *price* and *calories*.

DESCRIPTIVES VARIABLES=calories price

/STATISTICS=MEAN STDDEV MIN MAX.

**Descriptives**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| Calories per 12 Fluid Ounces | 35 | 68 | 175 | 139.77 | 24.447 |
| Price per 6-pack | 35 | 1.59 | 7.19 | 3.0274 | 1.12343 |
| Valid N (listwise) | 35 |  |  |  |  |

On which variable is there more dispersion?

The most dispersion is 12 Fluid ounces because of the range of 107 when it is compared to the 6-pack.

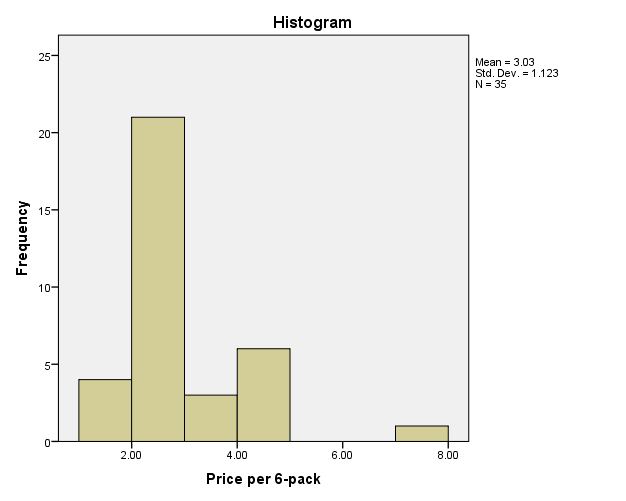
Is it even realistic to compare these two variables since they are on different scales?

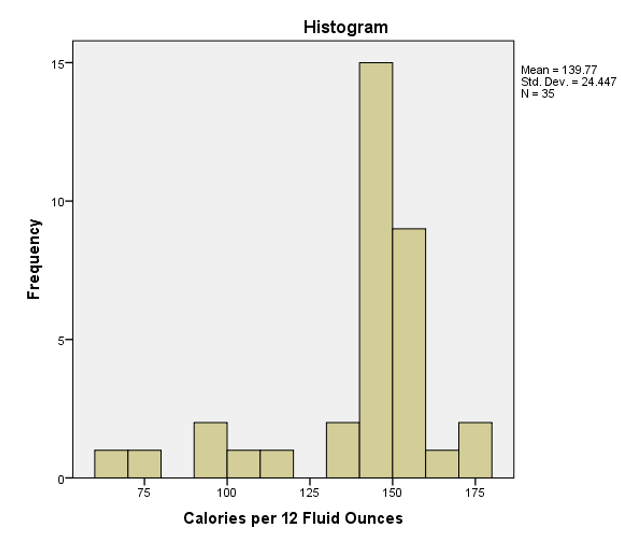
I don’t believe that it is realistic to compare these 2 because it really does not help with the information.

5. Continuing your analysis of *price* and *calories*, run the **Explore** procedure for these two variables. Request a histogram in addition to the defaults.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Case Processing Summary** | | | | | | |
|  | Cases | | | | | |
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| Price per 6-pack | 35 | 100.0% | 0 | 0.0% | 35 | 100.0% |
| Calories per 12 Fluid Ounces | 35 | 100.0% | 0 | 0.0% | 35 | 100.0% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Descriptives** | | | | |
|  | | | Statistic | Std. Error |
| Price per 6-pack | Mean | | 3.0274 | .18989 |
| 95% Confidence Interval for Mean | Lower Bound | 2.6415 |  |
| Upper Bound | 3.4133 |  |
| 5% Trimmed Mean | | 2.9317 |  |
| Median | | 2.6500 |  |
| Variance | | 1.262 |  |
| Std. Deviation | | 1.12343 |  |
| Minimum | | 1.59 |  |
| Maximum | | 7.19 |  |
| Range | | 5.60 |  |
| Interquartile Range | | .86 |  |
| Skewness | | 1.781 | .398 |
| Kurtosis | | 4.338 | .778 |
| Calories per 12 Fluid Ounces | Mean | | 139.77 | 4.132 |
| 95% Confidence Interval for Mean | Lower Bound | 131.37 |  |
| Upper Bound | 148.17 |  |
| 5% Trimmed Mean | | 141.83 |  |
| Median | | 147.00 |  |
| Variance | | 597.652 |  |
| Std. Deviation | | 24.447 |  |
| Minimum | | 68 |  |
| Maximum | | 175 |  |
| Range | | 107 |  |
| Interquartile Range | | 12 |  |
| Skewness | | -1.687 | .398 |
| Kurtosis | | 2.539 | .778 |





1. Does the standard error of each variable help you better determine which variable has more dispersion?

No because the standard error does not do anything in regards to determining which variable has more dispersion. The reason is because the standard of error is looking at the possibility of error within the data. Dispersion on the other hand looks at the measure of the range that can be found within the data set.