Karla Cruz

1/30/21

Using the “ Independent Project Data” set file supplied above, perform an analysis in StatCrunch for the following using the variable(s) of your choice:

Think carefully about what kind of variables to choose for the given tasks. A short descriptive statement should accompany each of the above including a description of the variables used and any meaning that may be attached to the results. Write up the project in a WORD document for submission.

1. Frequency distribution of a variable and bar graph of the same variable

I chose the variable of Mental Health score, because it is similar to heart rate score used in previous examples for class .The meaning of the results shows that out of 950 patients in the data sample 91.16% of them were between a 50-60 Mental Health score, 56.74% were between 40-50, 30.11% between 30-40, 8.3% between 20-30, .84% between 10-20, and the range with the least participants was between 0-10 with only .10% participants in that range group. The bar graph shows a normal distribution with a negative skew for these scores.

**Frequency table results for MentalHlth:**
Count = 950

| **MentalHlth** | **Frequency** | **Percent of Total** | **Cumulative Percent of Total** |
| --- | --- | --- | --- |
| 0 to 10 | 1 | 0.10526316 | 0.10526316 |
| 10 to 20 | 7 | 0.73684211 | 0.84210526 |
| 20 to 30 | 71 | 7.4736842 | 8.3157895 |
| 30 to 40 | 207 | 21.789474 | 30.105263 |
| 40 to 50 | 253 | 26.631579 | 56.736842 |
| 50 to 60 | 327 | 34.421053 | 91.157895 |
| 60 to 70 | 84 | 8.8421053 | 100 |



1. Descriptives of a continuous variable: mean, median, mode, skewness, kurtosis, standard deviation and graph of that variable

I chose BMI as a continuous variable. When looking at the results for BMI including the histogram we can see the shape is unimodal with one peak, fairly symmetrical, with a positive skew and a leptokurtotic kurtosis.

**Summary statistics:**

| **Column** | **n** | **Mean** | **Median** | **Std. dev.** | **Skewness** | **Kurtosis** | **Mode** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| BMI | 948 | 29.167131 | 27.97 | 7.3633338 | 0.9961317 | 1.1320945 | 24.05 |



1. Cross tabulation of two variables with the appropriate statistical test

The two values I chose for cross tabulation were marital status and poverty because they are values I have calculated for the class before and felt comfortable doing again.

I tested the following hypothesis:

H0: There is no relationship between poverty status and marital status

H1: There is a relationship between poverty status and marital status

**Contingency table results:**

|  | Marital Status  |  |
| --- | --- | --- |
| Poverty Status  | Divorced | Married | Never married | Total |
| Below poverty% Within Poverty Status % Within Marital Status % of total  | 208(28.38%)(76.19%)(21.89%) | 53(7.23%)(63.86%)(5.58%) | 472(64.39%)(79.46%)(49.68%) | 733(100%)(77.16%)(77.16%) |
| Above poverty% Within Poverty Status % Within Marital Status % of total  | 65(29.95%)(23.81%)(6.84%) | 30(13.82%)(36.14%)(3.16%) | 122(56.22%)(20.54%)(12.84%) | 217(100%)(22.84%)(22.84%) |
|  Total% Within Poverty Status % Within Marital Status % of total  | 273(28.74%)(100%)(28.74%) | 83(8.74%)(100%)(8.74%) | 594(62.53%)(100%)(62.53%) | 950(100%)(100%)(100%) |

The results show that the chi-square was statistically significant because the p-value of .0059 is less than the alpha value of .05 and the critical value of 10.266 is greater than the table value of 5.99. Therefore we can reject the null hypothesis and accept the alternative hypothesis and conclude there is a relationship between poverty status and marital status.

**Chi-Square test:**

| **Statistic** | **DF** | **Value** | **P-value** |
| --- | --- | --- | --- |
| Chi-square | 2 | 10.266594 | 0.0059 |

1. Comparison of two groups (single variable) on a single continuous variable with the appropriate statistical test

I chose age and BMI within married participants. I used a two-sample t-test because the independent variable is nominal and the dependent variable approached interval measurement.

I tested the following hypothesis:

H0: There is no difference or relationship between age and weight within married participants.

H1: There is a difference or relationship between age and weight within married participants.

**Two sample T hypothesis test:**

μ1 : Mean of Married age
μ2 : Mean of Married bmi
μ1 - μ2 : Difference between two means
H0 : μ1 - μ2 = 0
HA : μ1 - μ2 ≠ 0
(with pooled variances)

**Hypothesis test results:**

| **Difference** | **Sample Diff.** | **Std. Err.** | **DF** | **T-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| μ1 - μ2 | 7.9910843 | 1.1188071 | 164 | 7.1425041 | <0.0001 |

The results show a p-value of .0001, which is less than the alpha value of .05, and the t-stat value of 7.14 is greater than the table value of 1.96 showing the results are statistically significant. Since the calculated T value was greater than the table t-value the null hypothesis is rejected and we accept the alternative hypothesis that there is a difference between age and weight within married participants in the study.

1. Comparison of the effect of three or more groups (single variable) on a single continuous variable with the appropriate statistical test

For this I chose to use an ANOVA test since the question requires the comparison of three or more groups and ANOVA compares the mean of two or more groups. The variables I chose were marital status since it had three groups (married, single, and divorced) and weight since the dependent variable had to be an interval or ratio.

H0: There is no difference in mean between weight and marital status

H1: There is a difference in mean between weight and marital status

**Analysis of Variance results:**
Data stored in separate columns.

**Column statistics**

| **Column** | **n** | **Mean** | **Std. Dev.** | **Std. Error** |
| --- | --- | --- | --- | --- |
| Divorced Weight | 273 | 177.48718 | 49.885759 | 3.0192235 |
| Married Weight | 83 | 173 | 46.340469 | 5.0865272 |
| Never married Weight | 594 | 171.44444 | 45.422073 | 1.8636902 |

**ANOVA table**

| **Source** | **DF** | **SS** | **MS** | **F-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| Columns | 2 | 6838.7587 | 3419.3794 | 1.5594709 | 0.2108 |
| Error | 947 | 2076442.9 | 2192.6535 |  |  |
| Total | 949 | 2083281.6 |  |  |  |

The results show a p-value of .2108, which is greater than .05, and the F-statistic value of 1.5594 is less than the table statistic of 2.99. Indicating this result is not a statistically significant.

1. Scatterplot and correlation between the two continuous variables with the appropriate statistical test

I selected BMI and age as two continuous variables to test if there was a correlation between age and BMI score.

H0: There is no relation between BMI score and Age

H1: There is a relation between BMI score and Age



**Simple linear regression results:**

Dependent Variable: bmi
Independent Variable: age
bmi = 27.302015 + 0.050900468 age
Sample size: 948
R (correlation coefficient) = 0.043783014
R-sq = 0.0019169524
Estimate of error standard deviation: 7.3601599

**Parameter estimates:**

| **Parameter** | **Estimate** | **Std. Err.** | **Alternative** | **DF** | **T-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 27.302015 | 1.4041841 | ≠ 0 | 946 | 19.44333 | <0.0001 |
| Slope | 0.050900468 | 0.037761904 | ≠ 0 | 946 | 1.3479317 | 0.178 |

**Analysis of variance table for regression model:**

| **Source** | **DF** | **SS** | **MS** | **F-stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| Model | 1 | 98.4261 | 98.4261 | 1.8169199 | 0.178 |
| Error | 946 | 51246.668 | 54.171954 |  |  |
| Total | 947 | 51345.094 |  |  |  |

The results show the following:

Coefficient r = 0.043783014`

Coefficient of determination=0.0019169524

Indicating a very slight and weak positive relationship between BMI score and age.

The coefficient of determination means there is 0.1% of the variation in BMI Score can be explained by participant age.

The value of .178 is greater than .05 and the Coefficient r-valueof .043783014 is less than the table value of making the correlation not statistically significant. We accept the null hypothesis that there is no relationship between BMI score and Age.