This assignment is based on a research article entitled “Mind-Set Matters: Exercise and the Placebo Effect,” which is hyperlinked and attached within Readings section of Module 8. [It is the third article within the Articles section]. This research article is also attached within Instructions section of the Final Analytic Project page on Assignments folder of Blackboard. The dataset associated with the aforementioned research article is modified and saved in Excel file format as “Mind-Set Matters,” which is attached within the Instructions section of the Final Analytic Project page on Assignments folder of Blackboard. Download and save this dataset in your computer and review various Sheets of this dataset. Then, review all previous assignments, Module Quizzes, and read the aforementioned research article carefully before preparing the final analytic project. Answer all of the following questions. [Note: The authors of the research article used *p*rep, an alternative of the *p*-value, but you use *p*-value approach in questions related with hypothesis testing].

1. How many rows (besides the row containing names of variables) and columns are there in the original dataset?
2. Is the original dataset an example of “big data”? If yes, why? If not, why?
3. In the original dataset, there are several missing values. What is your best guess in regard with these missing values?
4. Look at the data (including variable names in the first row) in the sheet named “modified dataset.” On the same Excel sheet “modified dataset,” show computation of Means and Standard Deviations for the variables “age,” “aex,” and “awt” for each of the two conditions (cond = Control and cond = Informed) using Descriptive Statistics tool of Excel’s Data Analysis ToolPack. Report these values below (on this Word file) in a Tabular form. What do these descriptive statistics mean in the context of variables? Write in your own words.
5. Suppose you are interested in testing the null hypothesis “Ho: There is no difference in age between participants in the Control group and Informed group” at 5% level of significance using a two-tailed t-test. Show necessary computation on the Excel sheet titled “question 5” using the tool “t-Test: Two-Sample Assuming Equal Variances” from Excel’s Data Analysis ToolPack. Below, state the correct alternative hypothesis and all steps that will exemplify the procedure of running an independent samples t-test with equal variance assumption. Use *p*-value approach to write conclusion about the null hypothesis. Refer to research article for help on writing your answer (specifically, check your answers with footnote 4 on page 167 of the research article).

Step 1. State the Null and Alternate Hypotheses - Null Hypothesis (H0): Control Group participants age is the same as the Informed Group participants age

Alternate Hypothesis (HA): Control Group participants age is not the same as the Informed Group participants age.

H0: 1 = 2

HA:

Step 2. Establish a Confidence, which is the estimate of the percentage of time you expect to get close to the estimate if you repeat the experiment or resample the population. The level of significance is .05. Which is written α = .05, represents the probability of making the wrong decision when the null hypothesis is true (i.e., the probability of making a Type 1 error).

The formula for calculating Confidence Level is 1 – α.

(1 - .05 = .95 or 95%), 95% is the most common Confidence Level.

*The EXCEL Data Analysis Toolpack will calculate the steps below.*

Step 3. Determine the N1 and N2 which is the number of data points in each sample group.

In the Control Group the N1 is 40 and in the Informal Group N2 is 43.

Step 4. Determine Degrees of Freedom

Degrees of freedom is a little different from the number of data points used in the final statistical calculation. It is the minimum number of independent values in a distribution that can vary without infringing on any constraints.

The Calculation for Degrees of Freedom is:

K = (N1 + N2) - 2

K = (40 +43) -2

K = 81

Step 5. Calculate the Mean

The mean is the mathematical average calculated by summing all values in each sample group, then divide by the number of values in the respective sample.

SUM (B2:B41) = 1696

Total Number of Values 40

, Control Group Mean = 42.4

SUM (B2:B41) = 1467

Total Number of Values 43

, Informed Group Mean = 34.12

Step 6. Calculate the Variance: The Variance tells how spread out data points are in a data set. The closer the variance is to zero, tells that the data points are clustered together.

The variance Sample Variance (S2):

S2 =

{\displaystyle s^{2}}……………

………………….

1. The variables ex1 and ex2 denote “Perceived amount of exercise” at Time 1 and Time 2, respectively. On Excel sheet named “question 6,” create separate histograms for ex1 and ex2 using Histogram tool of Excel’s Data Analysis ToolPack.
   1. Look at the graphs and answer this question on this Word file (below): Are these variables normally distributed, positively skewed, or negatively skewed?
   2. Suppose you are interested in testing the null hypothesis “Ho: There is no difference in perceived amount of exercise between participants in the Control group and Informed group at Time 2” at 5% level of significance using a two-tailed t-test. Show necessary computation on the Excel sheet titled “question 6” using the tool “t-Test: Two-Sample Assuming Equal Variances” from Excel’s Data Analysis ToolPack. Below, state the correct alternative hypothesis and all steps that will exemplify the procedure of running an independent samples t-test with equal variance assumption. Use *p*-value approach to write conclusion about the null hypothesis. Refer to research article for help on writing your answer. Check your means and standard deviations with the values shown in Table 1 of the research article.
2. This question is about Chi-square test of independence. Go to the Excel sheet named “question 7” and see Excel formulas used for computing observed frequencies in Table 1. Provided on this sheet is Table 2 of expected frequencies. Use Excel formula (as shown in one of the previous Analytic Projects) to compute each of the six expected frequencies under the null hypothesis “Ho: There is no association between outcome (weight loss = wtloss, no change = nochange, and weight gain = wtgain) and condition (Control and Informed).” You have to show formula for computing each expected frequency inside Table 2 of this sheet. Then, below Table 2 on this sheet, compute the value of Chi-square statistic, corresponding degrees of freedom, and *p*-value. On this Word file (below), state alternative hypothesis, write degrees of freedom, and using *p*-value approach write your conclusion about the null hypothesis.
3. This question is about computing Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of a logistic regression model (used to predict classification of subjects based on covariates), which was used to classify subjects to either Control or Informed condition. Consider this logistic regression model as a diagnostic test. The Excel sheet named “question 8” shows two variables “cond” and “cond\_pred,” where the latter represents predicted condition. Also shown in this sheet is a table with actual condition in the columns and predicted condition in the rows. For the rows, consider Control as + (test is positive) and Informed as – (test is negative). For the columns, consider Control as + and Informed as –. Compute the values of Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of this model. Show computations in the Excel sheet and report these values (below).
4. Write a Discussion section of all the results presented as answers to questions 1 through 8.
5. What additional tests would you suggest to enhance the findings of research article in the context of questions 7 and 8?