**Part 1 (50 points total) –** *You will be able to complete this part of the assignment, once you have reviewed the materials in Module 1.*

**Part 1a (10 points total)**

Using the information in Module 1 and the data provided in your spreadsheet, please indicate the type of data that were collected for each variable in this table.

|  |  |
| --- | --- |
| Variable Name | Type of Data |
| COVID-19 Status  *(1pt)* |  |
| Age  *(1pt)* |  |
| Biological Sex  *(1pt)* |  |
| BMI  *(1pt)* |  |
| Hospitalization  *(1pt)* |  |
| Days in Hospital  *(1pt)* |  |
| Kidney Disease  *(1pt)* |  |
| Cardiovascular Disease  *(1pt)* |  |
| Healthcare Worker  *(1pt)* |  |

What is the overall sample size for this study?  *(1pt)*

**Part 1b (10 points total)**

**Among all participants in your dataset,** please calculate the following summary statistics for Age. Note, these statistics should only be calculated for individuals with data for a given variable. Please report your responses rounding to 1 DECIMAL PLACE.

|  |  |  |
| --- | --- | --- |
| Variable Name | Age | Point Value |
| Mean |  | *1* |
| Standard Deviation |  | *1* |
| Median |  | *1* |
| Range |  | *1* |
| Interquartile Range |  | *1* |

**Among all participants in your dataset,** please calculate the following summary statistics for Days in the Hospital. Note, these statistics should only be calculated for individuals with data for a given variable. Please report your responses rounding to 1 DECIMAL PLACE.

|  |  |  |
| --- | --- | --- |
| Variable Name | Days in Hospital | Point Value |
| Mean |  | *1* |
| Standard Deviation |  | *1* |
| Median |  | *1* |
| Range |  | *1* |
| Interquartile Range |  | *1* |

**Part 1c (10 points total)**

**Among participants that are COVID-19 POSITIVE**, please calculate the following summary statistics for Age. Note, these statistics should only be calculated for individuals with data for a given variable. Please report your responses rounding to 1 DECIMAL PLACE.

|  |  |  |
| --- | --- | --- |
| Variable Name | Age | Point Value |
| Mean |  | *1* |
| Standard Deviation |  | *1* |
| Median |  | *1* |
| Range |  | *1* |
| Interquartile Range |  | *1* |

**Among participants that are COVID-19 POSITIVE**, please calculate the following summary statistics for BMI. Note, these statistics should only be calculated for individuals with data for a given variable. Please report your responses rounding to 1 DECIMAL PLACE.

|  |  |  |
| --- | --- | --- |
| Variable Name | BMI | Point Value |
| Mean |  | *1* |
| Standard Deviation |  | *1* |
| Median |  | *1* |
| Range |  | *1* |
| Interquartile Range |  | *1* |

**Part 1d (10 points total)**

**Among all participants in your dataset,** please calculate the proportion of “Yes” or “Positive” responses for the variables in the table below. Please REPORT your response AS A % AND ROUND TO nearest whole number.

|  |  |  |
| --- | --- | --- |
| Variable Name | Proportion “Yes” or “Positive” (report as %) | Point Value |
| COVID-19 status |  | *1* |
| Hospitalization |  | *1* |
| Kidney Disease |  | *1* |
| Cardiovascular Disease |  | *1* |
| Healthcare Worker |  | *1* |

Using the information in your dataset and what you know about contingency tables, please fill in the contingency table below. Remember, the numbers in the contingency table should be whole numbers. *(5pts total)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | COVID-19 Positive | COVID-19 Negative | Total |
| Male | *(1pt)* | *(1pt)* |  |
| Female | *(1pt)* | *(1pt)* |  |
| Total |  |  | *(1pt)* |

**Part 1e (10 points total)**

**Draw two boxplots to show the distribution of age based on COVID-19 status. Make sure you draw and label your axis. (***Hint: There should be two boxplots for age on the axis below (one to show ages by COVID-19 POSITIVE participants, one to show ages by COVID-19 NEGATIVE participants. Be sure to include and label ALL major components of the boxplot: min, max, Q1, Q3, median.)*

*(5pts each, if both correct)*

**Part 2 (50 points total) –** *You will be able to complete this part of the assignment, once you have reviewed the materials in Module 2.*

**Part 2a (10 points total)**

Please use the Normal Distribution Probability Estimation Methods to complete the following questions. Please report your responses rounded to one decimal place.

|  |  |  |
| --- | --- | --- |
|  | Lower Value | Upper Value |
| Based on your dataset, there is a 95.4% probability that patients with COVID-19 will fall between which two values for BMI? | *(2pts)* | *(2pts)* |
| Based on your dataset, there is a 95.4% probability that patients with COVID-19 will fall between which two values for AGE? | *(2pts)* | *(2pts)* |

|  |  |  |
| --- | --- | --- |
|  | Report as a probability with 3 decimal places (*not %*) | Point Value |
| What is the probability that the AGE of a patient with COVID-19 is between 21 and 30 years. |  | *1* |
| What is the probability that the BMI of a patient with COVID-19 is 29.0 or greater? |  | *1* |

**Part 2b (10 points total)**

Please use the Binomial Distribution Probability Estimation Methods to answer the questions below.

1. Among COVID-19 positive patients in your dataset, what proportion were hospitalized? Report as a proportion (not a %) with two decimal places. *(2pts)*

|  |  |
| --- | --- |
|  | Report as a probability with 4 decimal places (*not %*) |
| The probability that exactly 18 of 25 COVID-19 positive patients in the community would be hospitalized. | *(2pts)* |

Next to the BMI column, insert a new variable (column) in your dataset and name it “Obesity Status”. Use the information in the BMI column to add data to the new “Obesity Status” variable for everyone in your dataset. Here’s how to do this: If BMI for a given participant is ≥30, then type in “yes” for that participant to indicate that they are obese. If BMI is <30, then type in “no” for that participant to indicate that they are not obese. For more information on how to create a new variable/column, see: <https://support.microsoft.com/en-us/office/insert-or-delete-rows-and-columns-6f40e6e4-85af-45e0-b39d-65dd504a3246#:~:text=To%20insert%20a%20single%20new,C%2C%20and%20click%20Insert%20Columns>.

1. Among COVID-19 positive patients in your dataset, please calculate the proportion of obese participants. Report as a proportion (not a %) with two decimals places. *(4pts)*

|  |  |
| --- | --- |
|  | Report as a probability with 4 decimal places (*not %*) |
| The probability that exactly 10 of 25 COVID-19 positive patients in the community are obese. | *(2pts)* |

**Part 2c (10 points total)**

*Please use the Poisson Distribution Probability Estimation Methods to answer the questions below:*

1. Among COVID-19 positive patients that were hospitalized, what is the average length of hospital stay? (In other words, what was the mean for hospital days among hospitalized COVID-19 positive patients?) Please report your response rounded to two decimal places. *(4pts)*

|  |  |
| --- | --- |
|  | Report as a probability with 4 decimal places (*not %*) |
| What is the probability that a hospitalized COVID-19 patient would spend 7 or fewer days in the hospital? | *(3pts)* |

|  |  |
| --- | --- |
|  | Report as a probability with 4 decimal places (*not %*) |
| What is the probability that a hospitalized COVID-19 patient would spend between 7 and 14 days in the hospital? | *(3pts)* |

**Part 2d (10 points total)**

Suppose the data in your file were collected from a case-control study to examine the association between previously diagnosed cardiovascular disease (independent variable) and hospitalization for COVID-19 (dependent variable). In other words, your research question is: *Is cardiovascular disease a risk factor for being hospitalized for COVID-19?*

To conduct this study, COVID-19 positive patients were enrolled based on whether they were hospitalized for COVID-19 (y/n) and then asked about previous diagnosis of cardiovascular disease (y/n). ***(NOTE: this will only include participants that were COVID-19 positive)***

1. **Using the data in your Excel file, please fill in the cells in this contingency table:**

|  |  |  |
| --- | --- | --- |
|  | Dependent Variable (yes):  Hospitalization (y) | Dependent Variable (no):  Hospitalization (n) |
| Independent Variable (yes):  Cardiovascular (y) | *(1pt)* | *(1pt)* |
| Independent Variable (no):  Cardiovascular (n) | *(1pt)* | *(1pt)* |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | 95% Confidence Interval  Lower Limit | 95% Confidence Interval  Upper Limit |
| Prevalence of Cardiovascular Disease  *(report with/round to 2 decimal places)* | *(1pt)* | *(1pt)* | *(1pt)* |
| Odds Ratio *(report with/round to 2 decimal places)* (Cardiovascular Disease -> COVID-19 Hospitalization) | *(1pt)* | *(1pt)* | *(1pt)* |

**Part 2e (10 points total)**

Suppose the data in your file were collected from a conduct a cohort study to determine if having COVID-19 (independent variable) increases the risk of kidney disease (dependent variable). In other words, your research question is: Are COVID-19 positive participants more likely to develop kidney disease than COVID-19 patients?

To answer this question, we will enroll participants at a medical clinic based on their COVID-19 status (y/n) and follow them for 6 months to see if kidney disease develops (y/n). ***(NOTE: this will include ALL participants in your dataset)***

1. **Fill in the cells based on your dataset:** *(1pt each)*

|  |  |  |
| --- | --- | --- |
|  | Dependent Variable (yes):  Kidney Disease y | Dependent Variable (no):  Kidney Disease n |
| Independent Variable (yes):  COVID-19 positive |  |  |
| Independent Variable (no):  COVID-19 negative |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | 95% Confidence Interval  Lower Limit | 95% Confidence Interval  Upper Limit |
| Cumulative Incidence of Kidney Disease among Study Sample  *(per 100 people) (round to 2 decimal places)* | *(1pt)* | NA | NA |
| Relative Risk *(report with/round to 2 decimal places)*  (COVID-19 -> Kidney Disease) | *(1pt)* | *(2pt)* | *(2pt)* |

**Part 3 (50 points total) –** *You will be able to complete this part of the assignment, once you have reviewed the materials in Module 3.*

**Part 3a (25 points)**

Let’s assume that Age is normally distributed among patients with and without COVID-19 in your dataset. Use the appropriate hypothesis test to determine if there is a difference in mean age between patients with COVID-19 and patients without COVID-19. Complete the following questions:

|  |  |
| --- | --- |
| What is the null hypothesis statement?  *(remember to be specific and use appropriate notation) (2pt)* |  |
| What is the alternative hypothesis statement?  *(remember to be specific and use appropriate notation)*  *(2pt)* |  |
| Which hypothesis test should you use and WHY?  *(3pt)* |  |

|  |  |  |
| --- | --- | --- |
|  | Group 1:  COVID-19 Negative | Group 2:  COVID-19 Positive |
| Sample Size | *(1pt)* | *(1pt)* |
| Mean | *(2pt)* | *(2pt)* |
| Variance | *(2pt)* | *(2pt)* |

|  |  |
| --- | --- |
|  | Results |
| Test Statistic  *(2pt)* |  |
| Degrees of Freedom  *(2pt)* |  |
| P value range *(2pt)*  *(use df=50 for table in book)* |  |
| Overall Conclusion  *(Hint: State your findings relative to the null hypothesis and explain WHY.)*  *(2pt)* |  |

**Part 3b (25 points)**

Let’s compare the proportion of patients with and without COVID-19 that are healthcare workers. Use the appropriate hypothesis test to determine if there is a difference in the proportion of healthcare workers between participants that are COVID-19 positive vs. COVID-19 negative. Complete the following tasks:

|  |  |
| --- | --- |
| What is the null hypothesis statement?  *(remember to be specific and use appropriate notation)*  *(2pt)* |  |
| What is the alternative hypothesis statement?  *(remember to be specific and use appropriate notation)*  *(2pt)* |  |
| Which hypothesis test should you use and WHY?  *(3pt)* |  |

|  |  |  |
| --- | --- | --- |
|  | Group 1:  COVID-19 Negative | Group 2:  COVID-19 Positive |
| Sample Size | *(1pt)* | *(1pt)* |
| Count of Individuals that ARE Healthcare Workers | **(a)**  *(1pt)* | **(b)**  *(1pt)* |
| Count of Individuals that ARE NOT Healthcare Workers | **(c)**  *(1pt)* | **(d)**  *(1pt)* |
| Proportion of Patients that are Healthcare Workers | *(2pt)* | *(2pt)* |

|  |  |
| --- | --- |
|  | Results |
| Test Statistic  *(2pt)* |  |
| Degrees of Freedom  *(2pt)* |  |
| P value range  *(2pt)* |  |
| Overall Conclusion  *(Hint: State your findings relative to the null hypothesis and explain WHY.) (2pt)* |  |

**Part 4 (50 points total) –** *You will be able to complete this part of the assignment, once you have reviewed the materials in Module 4.*

**Part 4a (24 points total)**

Let’s practice calculating confidence intervals for **mean BMI** among patients in your dataset.

Please provide your response rounded to 1 decimal place.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| GROUP | Mean BMI  (yrs) | 90% CI  Lower limit | 90% CI  Upper limit | 95% CI  Lower limit | 95% CI  Upper limit | 99% CI  Lower  limit | 99% CI  Upper limit |
| Overall  *(everyone)*  *(use df=50 for table in book)* |  |  |  |  |  |  |  |
| COVID-19 Positive |  |  |  |  |  |  |  |
| COVID-19 Negative |  |  |  |  |  |  |  |

**Part 4c (26 points total)**

Suppose we wanted to conduct a follow-up study to determine if there is a difference in the proportion of patients with cardiovascular disease between individuals that are COVID-19 positive and COVID-19 negative. Use the data in your dataset to help determine the number of individuals we would need to include in this follow-up study with 95% confidence and 80% power.

|  |  |  |
| --- | --- | --- |
|  | Response | Point value |
| Proportion of COVID-19 negative participants with cardiovascular disease *(use this for p0)* |  | *2* |
| Proportion of COVID-19 positive participants with cardiovascular disease *(use this for p1)* |  | *2* |
| zα/2 |  | *2* |
| zβ |  | *2* |
| Sample size needed for follow-up study |  | *3* |

There is mounting evidence suggesting obesity is a risk factor for severe COVID-19 illness. Suppose we wanted to conduct a follow-up study to determine if there is a difference in the mean BMI of patients with and without COVID-19 infection. Use the data in your dataset to help determine the number of individuals we would need to include in this follow-up study with 95% confidence and 80% power.

|  |  |  |
| --- | --- | --- |
|  | Response | Point value |
| Mean BMI among COVID-19 negative participants *(use this for µ1)* |  | *2* |
| BMI variance among COVID-19 negative participants *(use this for )* |  | *2* |
| Mean BMI among COVID-19 positive participants *(use this for µ2)* |  | *2* |
| BMI variance among COVID-19 positive participants *(use this for )* |  | *2* |
| zα/2 |  | *2* |
| zβ |  | *2* |
| Sample size needed for follow-up study |  | *3* |