PFS1243 – BUSINESS STATISTICS

GROUP ASSIGNMENT

**Problem Description:**

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide.  
Heart failure is a common event caused by CVDs and this dataset contains 12 features that can be used to predict mortality by heart failure.

Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies.

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidemia or already established disease) need early detection and management. Therefore, the report on the heart failure need to be done.

 You are part of the research team that conduct on this research. Then, your and team are needed to prepare a report on the Cardiovascular diseases (CVDs). You will use descriptive statistics and your knowledge of continuous distributions to complete this task. Data are given in csv format.

Data description:

#age – age

#anamea - Decrease of red blood cells or hemoglobin (boolean)

# creatinine\_phosphokinase - Level of the CPK enzyme in the blood (mcg/L)

# diabetes – 1(yes),0(No)

# ejection\_fraction - Percentage of blood leaving the heart at each contraction (percentage)

# high\_blood\_pressure -1(yes),0(No)

# platelets - Platelets in the blood (kiloplatelets/mL)

# serum\_creatinine - Level of serum creatinine in the blood (mg/dL)

# serum\_sodium - Level of serum sodium in the blood (mEq/L)

#sex- 1(Female),0(Male)

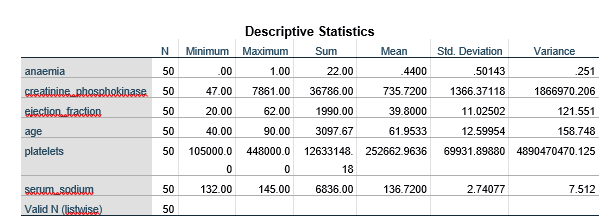
#smoking -1(yes),0(No)

#time - Follow-up period (days)

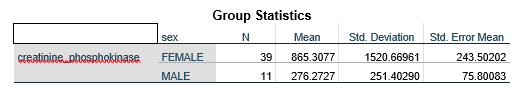
**Required:**

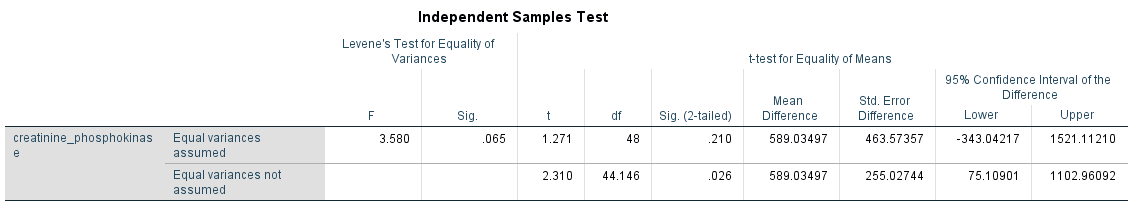
1. Randomly select 50 data from dataset, then used the data to estimate descriptive statistics for each of these factors Decrease of red blood cells or hemoglobin (boolean), Level of the CPK enzyme in the blood (mcg/L), Percentage of blood leaving the heart at each contraction (percentage), Platelets in the blood (kiloplatelets/mL), and Level of serum sodium in the blood (mEq/L). Compare all these factors. Be sure to comment on the central tendency, variability and shape of these two investments. Based on your findings, discuss on the mean and variability.

Assume that the data are normally distributed with a mean and standard deviation (as estimated in (a) rounded to the nearest integer). Answer the following questions:

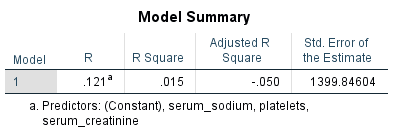


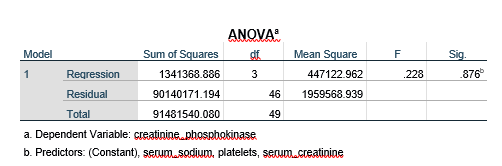
1. Find the probability that *Level of serum sodium in the blood (mEq/L)* will exceed 145*(mEq/L)*
2. Find the probability that *age*  exceed 70
3. Find the probability that *Percentage of blood leaving the heart at each contraction (percentage)* will be between 35% and 42%
4. Creatine phosphokinase (CPK) is an enzyme in the body. It is found mainly in the heart, brain, and skeletal muscle. Therefore, what factors does effect level of these enzyme. Test whether there have significant different between gender on level of the CPK enzyme in the blood (mcg/L). Use 0.05 level of significant and interpret your answer.





1. Test whether platelets in the blood, level of serum creatinine in the blood, level of serum sodium in the blood effect on level of the CPK enzyme in the blood. Use 0.05 level of significant and interpret your answer.





|  |
| --- |
|  |



BUSINESS STATISTICS ASSIGNMENT RUBRIC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
| **Category (Points)** | Below quality (1-3) | Collegiate quality (4) | Excellent quality (5) | Total  Possible | Total  Earned |
| **Descriptive Statistics**  **(clo 2)** | Descriptive statistics omitted or not given clearly | Adequate overview of descriptive statistics | Thorough, clear, concise overview of descriptive statistics | 10 |  |
| **Scatter plot (clo 2)** | Contains only 1 of the following items: title, labels, or consistent and appropriate intervals.  Dots on graph are not in proper order and cross over each other. | Lacking 2 of the following items: title, labels, or consistent and appropriate intervals.  Needs to work on connecting and plotting dots. | All graphs consists of title, labels, and consistent and appropriate intervals.  Followed directions and connected dots and are in correct order moving gradually to the right of the graph. | 10 |  |
| **Hypotheses (clo 2)** | Null and alternative hypotheses not stated or incorrect | Null and alternative hypotheses stated, but not in mathematical terms | Null and alternative hypotheses both stated correctly in mathematical terms. | 10 |  |
| **Regression analysis (clo 3)** | Inappropriate statistical tests conducted or statistical tests not conducted or explained correctly | Appropriate statistical tests are conducted correctly and explained adequately | Explanation of statistical tests is thorough, articulate, and precise | 10 |  |
| **Interpretation of Results (clo 3)** | P-values not computed correctly or not interpreted correctly | P-values computed correctly and interpreted correctly with respect to hypothesis | Interpretation of P-values includes clear discussion of significance level | 10 |  |

**TOTAL MARKS:**

**LECTURER:**

\*\*Please attach this marking scheme together with your report during submission.