

Homework – 1

Assignment – 1A

Total Points – 50

Use R to Solve Question 7

Question 1

[5]

For the variables listed below [(a) – (j)] select the correct measurement scales from (i) to (v)-

- (i) Nominal
- (ii) Ordinal
- (iii) Binary
- (iv) Interval
- (v) Ratio

- (a) Gender
- (b) Temperature in Kelvin
- (c) Hair color of the students in the class
- (d) Ethnicity of the population in a county
- (e) Age in years/months of the students in the class
- (f) Education Diplomas earned by a student [Bachelors Diploma, MBA etc.]
- (g) European Countries
- (h) Alcoholic or non-alcoholic
- (i) Marital status of a person
- (j) Weight of students in the class

Question 2

[5]

Clinical Trial on New Drug for Weight loss

A random sample of 12 people in the clinical trial for new drug were asked to log their weights loss after 1 week in the program. Following is the weight loss (lbs) reported by the people in the clinical trial:

6.6, 5.5, 8.3, 7.2, 1.6, 18.6, 12.5, 22.0, 8.7, 3.4, 5.3, 2.2

Compute the central tendency theorem and analyze the data with respect to the weight loss.

Question 3

[5]

A sample of patients in the clinical trial of a new drug for weight loss produced the bell-shaped histogram. The mean and the standard deviation for the weight loss is 18 and 1.5. Explain how you understand this data with respect to severe weight loss or moderate weight loss.

Question 4

Define the following terms:

[5]

Sample Space, Independent event, Retrospective study, Prospective study, Random Variable

Question 5

[5]

Complete the below a Contingency Table for a hospital scenario of patients given drug 1 and drug 2 for treating Disease 1 and Disease 2. Based on the table answer the following questions:

	Drug 1	Drug 2	Total
Disease 1	13	59	
Disease 2	15	8	
Total			

- (i) What is probability of patients given Drug1
- (ii) What is probability of patients diagnosed with Disease 2
- (iii) What is the probability of patients with Disease 1 are treated with Drug 2
- (iv) What is the probability of the patients treated with Disease 2 are treated with Drug 1

Question 6

[5]

A study of emergency department services with respect to sexual assault. Using the below contingency table answer the following questions:

Forensic Nurse on ED staff	Rape crisis center linkage		
	No	Yes	Total
No	8	12	20
Yes	4	29	33
Total	12	41	53

- (i) Find the probability that ED will have a forensic nurse?
- (ii) Find the probability that ED will not have a forensic nurse?
- (iii) The probability that the ED will have a relationship with the rape center.
- (iv) The probability that the ED will not have a relationship with the rape center.
- (v) Does having a forensic nurse trained to assist victims of sexual violence on staff in ED affect the probability that a hospital will have a relationship with a rape crisis center?

Question 7

[20]

Data Exploration (no imputation needed)

You are given NationalDisease_Data. Use R to answer the below questions.

- (i) How many variables are present in the given dataset?
- (ii) Use histogram to understand the change in the age profile across the years. Is there a change in age profile before year 2000, 2000-2010 and after 2010? Explain in detail your findings.
- (iii) Analyze the variable sex and age for the years 1995, 2000, 2005, 2010, 2015 with respect to "Prediabetes", "obesity", "coronary heart disease", "stroke", and "Other heart disease" and "Diagnosed with Diabetes". (You can plot all variables in one table for different years or plot each variable individually) Explain your findings.
- (iv) For the variables sex, race and age analyze the trend for hospitalization across the following diseases for the years 1980, 1990, 2000, 2010, 2015, 2016: "Hospitalization for Hyperglycemic Crisis", "Hospitalization for hypoglycemia", "Hospitalization for Major CVD", "Hospitalization for Ischemic Heart Disease", "Hospitalization for Heart Failure", and "Hospitalization for Stroke". Explain your findings.
- (v) Analyze the diseases (Current Smoking, Obesity, A1c, Triglycerides, Systolic Blood Pressure, Diastolic Blood Pressure) for Hospitalization for Major CVD and Hospitalization for Stroke. (You can use other variables also to understand the trend) Explain your findings.

(Packages: ggplot, seaborn, stats, hist())

Note: There won't be a single answer for all the trends. It depends on how you identified the subset and chose the features. You can choose packages or work without using a package. We will look at the approach/effort to grade.