IIHMR DELHI | INTERNATIONAL INSTITUTE OF HEALTH MANAGEMENT RESEARCH

Post Graduate Diploma in Management (Hospital & Health Management)
PGDM – 2023-25 Batch

1st Year – 1st Term Examination

Course & Code	:	**Biostatistics-CC 604**	**Reg. No.**	:	
Term & Batch	:	**I, 2023-25**	**Date**	:	**December 16, 2023**
Duration	:	**3 Hrs.**	**Max. Marks**	:	**70**

Instructions:

- Budget your time as per the marks given for each question and write your answer accordingly.
- Don't write anything on the Question Paper except writing your Registration No.
- Mobile Phones are not allowed even for computations.

Part A- 15 Questions (Objective type or very short answer type questions: carrying 1 mark each)

1. What is the components of Descriptive Statistics?
a. To visualise data only.
b. To calculate summary measures only.
c. To calculate summary measures and to visualize data.
d. To visualise data, calculate summary measures and to do statistical hypothesis testing.

2. Which statistical test is used to assess the association between two categorical variables?
a. Pearson's correlation coefficient
b. Chi-square test
c. ANOVA
d. Mann-Whitney U test

3. What is true for a null hypothesis to be rejected if p represents the p-value and α is the level of significance?
a. $p < \alpha$
b. $p = \alpha$
c. $p > \alpha$
d. $p \sim \alpha$

4. What is true in context of standard error?
a. Standard deviation of the different samples of a same population
b. Standard deviation of different observations in a sample
c. Standard deviation of different samples from different populations
d. Standard deviation of different observations from different samples.

5. What is not correct in context of test statistic?
a. It captures the deviation of sample estimate from the population.
b. It captures the standard error of the sample taken.
c. It provides the basis for rejection or not rejection of null hypothesis.
d. It provides the level of significance.

6. Which statistical test is appropriate for comparing three or more group means?
a. t-test
b. ANOVA
c. Chi-square test
d. Wilcoxon signed-rank test

7. What is the correct formula for calculating 95% Confidence interval for sample mean ($\overline{X}$) if S.D. and S.E. are standard deviation and standard error of the sample respectively?
a. ($\overline{X}$ - 1.96 × S.D., $\overline{X}$ + 1.96 × S.D.)
b. ($\overline{X}$ + 1.96 × S.D., $\overline{X}$ - 1.96 × S.D.)
c. ($\overline{X}$ - 1.96 × S.E., $\overline{X}$ - 1.96 × S.E.)
d. ($\overline{X}$ + 1.96 × S.E., $\overline{X}$ - 1.96 × S.E.)

8. Which of the following plot is used to detect association between two variables?
a. Box plot
b. Bar plot
c. Histogram
d. Scatter plot

9. In a normal distribution, what percentage of data falls within two standard deviations of the mean?
a. 68%
b. 75%
c. 95%
d. 99%

10. Which statistical test is used to determine if there is a significant difference between the means of two related groups?
a. t-test
b. ANOVA
c. Chi-square test
d. Paired t-test

6. Which of the following is not a measure of dispersion?
a. Range
b. Standard deviation
c. Interquartile range (IQR)
d. Mode

11. Which of the following are measures for shape of data?
a. Skewness and Kurtosis
b. Inter Quartile Range (IQR).
c. Variance and Standard Deviation.
d. Mean and Mode.

12. A researcher is interested in studying the relationship between smoking and lung cancer. What types of variables will be used to do the study smoking status and lung cancer.
a. Both quantitative
b. Both Categorical or Qualitative
c. Smoking Categorical and Lung Cancer Quantitative
d. Smoking Quantitative and Lung Cancer Categorical

13. What are the lower and upper whisker limits in a Box-plot?
a. Q3- IQR*1.5 and Q1+IQR*1.5
b. Q1 +IQR*1.5 and Q3-IQR*1.5
c. Q1- IQR*1.5 and Q3+IQR*1.5
d. Q3+ IQR*1.5 and Q3- IQR*1.5

14. What does a confidence interval represent?
a. The range of values that are likely to include the population parameter.
b. The precise value of the population parameter.
c. The variability of the sample data.
d. The probability of making a Type I error.

15. Which of the following plot is used to detect association between two variables?
a. Box plot
b. Bar plot
c. Histogram
d. Scatter plot

Part B- 5 Questions (Short answer questions carrying 5 marks each)

Q 16. What do you understand by standard error and how it differs from standard deviation?

OR

Calculate standard error if Standard Deviation is 0.3 and sample size is 30? What will be the null and alternative hypothesis if one wants to test whether the average haemoglobin of the sample is 11.5 and that of population is 12?

Q.17. Explain 95% confidence Interval for mean?

OR

What will be 95% confidence interval for mean if standard deviation is 1, sample size is 10 and sum of all the observations is 30? Assume normal distribution for the sample.

Q 18. Explain the differences between qualitative and quantitative variables through arbitrary health data-set?

Q 19. Explain different types of data visualisation techniques?

OR

Identify which one of the following variables in the arbitrary data set is Discrete, Continuous and Categorial/Qualitative variable? Write R-syntax to calculate mean, median, IQR and for obtaining boxplot for the Anxiety Scores?

Exam_Marks	Anxiety_Scores	Gender
40	86.298	Male
65	88.716	Female
80	70.178	Male
80	61.312	Male
40	89.522	Male
70	60.506	Female
20	81.462	Female
55	75.82	Female
50	69.372	Female
40	82.268	Female

Q 20. What are outliers? How to detect them? Explain by giving examples?

OR

Calculate IQR and outliers for the following data-set. Write R-syntax to obtain both box-plot and five-number summary statistics for following data-set.
10,20,30,40,50,60,70,100,170

Part C- 3 Questions (Long answer questions carrying 10 marks each)

Q 21. Explain differences between descriptive statistics and inferential statistics? Explain different components of descriptive statistics and inferential statistics?

OR

What is the role of statistics? Explain why we need statistics? Draw the hierarchy of Research Methodology and explain where does statistics fit into it?

Q 22. Find below the dataset for 14 patients collected by Divyanshu. Write appropriate methodology/steps including null and alternative hypothesis, test statistic, reject accept criteria for testing the following situations.

(i) t-test for testing average blood pressure and average insulin levels among patients with no cardiovascular disease (CVD) and with CVD.

(ii) Chi-square test for testing association between low BMI and high BMI patients with CVD and no CVD patients. You can set cutoff for low and high BMIs as per your knowledge.

Pregnancies	**Glucose**	**Blood Pressure**	**Skin Thickness**	**Insulin**	**BMI**	**Diabetes Pedigree Function**	**Age**	**CVD**
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	24	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	27	0
1	189	60	23	846	30.1	0.398	29	1

Q 23. How do you approach for calculating sample size for any survey? Define various types sampling techniques and sample size calculations?