Indian Statistical Institute, Bangalore
B. Math II, First Semester, 2020-21
Back Paper Examination, Statistics I
Maximum Score 100 **Duration: 3.5 Hours**

Students are allowed to consult the book Statistics by McClave and Sincich.
Values from the normal distribution qnorm(0.9)=1.281552, qnorm(0.995)=2.575829, qnorm(0.95)=1.644854
Instructor: Rituparna Sen rsen@isibang.ac.in ritupar.sen@gmail.com 9176620249

1. (10+2+3) Let $X_1, X_2 \cdots X_n$ be a random sample from the Pareto distribution with pdf given by

$$f(x \mid \theta) = c\theta^c x^{-(c+1)}, \quad \text{if} \quad x > \theta, \tag{1}$$

where $\theta \in \mathbb{R}^+$ and $c > 2$. Both θ and c are unknown.

   (a) Obtain the method of moments (MoM) estimator for θ.

   (b) Is this the only MoM estimator, or is it possible to obtain other estimators based on the method of moments?

   (c) Is the estimator consistent? Justify your answer.

2. (5+5+5) Let $U \sim Unif(0,1)$ and $W = \tan(\pi(U - 1/2))$.

   (a) Find the distribution of W.

   (b) Use this to describe how to generate a random variable from the Cauchy distribution with parameters 10 and 5.

   (c) Find the mean, median and first quartile of the distribution in part (b).

3. (5+5) Answer any two of the following questions related to the class presentations.

   (a) How to test if data comes from N(0,1) when frequency table is given?

   (b) Describe the Friedman test for ordinal data.

   (c) What is the model for quantile regression? What is the quantity that is minimized to estimate the parameters?

4. (5+5+5) Let $X_1, X_2, \cdots, X_n$ be a random sample from the population with distribution function F. Let $F_n(x)$ be the empirical distribution function, that is

$$F_n(x) = \frac{\#\{X_i \leq x\}}{n}$$

Let $c < d$ be two given real numbers. Define $V = F_n(c)$ and $W = F_n(d)$.

   (a) Find the expectation and variance of V.

   (b) Find the distribution of V

   (c) Find the covariance between V and W.

5. (2+5+5+5+3) A logger knows the average time for his cutting machine to cut a tree is 9.8 minutes. A new machine on the market claims to cut the trees in less than 9.8 minutes. A random sample of 25 test runs on the new machine yielded a mean of 8.5 minutes with a standard deviation of 1.5. Using $\alpha = 0.01$, perform the appropriate test of hypothesis to determine if the new machine cuts faster than the logger's machine. In particular, answer the following questions.

   (a) State the null and alternative hypotheses.

   (b) State the test statistic and find its distribution under the null hypothesis.

   (c) State the assumptions required.

   (d) Compute the value of the test statistic. What will be the R command for finding the p-value?

   (e) Suppose the p-value is 0.0001. Is the null hypothesis rejected? What is the conclusion that the owner can draw regarding the defective rate of computers?

6. (7+3) In a study of the relationship between birth order and college success, an investigator found that 140 in a sample of 200 college graduates were firstborn or only children. In a sample of 120 non-graduates of comparable age and socioeconomic background, the number of firstborn or only children was 66.

   (a) Estimate the difference between the proportions of firstborn or only children in the two populations from which these samples were drawn. Use a 95% confidence interval.

   (b) Can we conclude that the proportions are different in the two populations?

7. (4+4) The iris data consists of 4 characters (sepal length, sepal width, petal length, petal width) measured on 50 flowers from each of 3 species (setosa, versicolor, virginica). We run the following command in R.

```
summary(aov(formula = Sepal.Width ~ Species, data = iris))
```

   (a) Complete the table of output.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Species		11.35			<2e-16
Residuals			0.115		

   (b) Carry out the ANOVA test using the above output stating the null and alternative hypotheses, assumptions and conclusions.

8. (7) Explain what the following R code and output is doing. The data is on hair and eye color of 592 individuals. State the model, hypotheses, data, assumptions, test statistic, its distribution and conclusion.

```
> data
       Eye
Hair    Brown Blue Hazel Green
  Black    68   20    15     5
  Brown   119   84    54    29
  Red      26   17    14    14
  Blond     7   94    10    16
> chisq.test(data)

Pearson's Chi-squared test

data:  data
X-squared = 138.29, df = 9, p-value < 2.2e-16
```