INDIAN STATISTICAL INSTITUTE

Statistical Methods IV

Semestral Examination

B II, Semester II, 2024-25

Date: 02.05.2025 Time: 3 hours

Answer as many as you can. The maximum you can score is 50.

1. Suppose that a random sample of size 10 is available from a trivariate normal distribution, with mean vector $\mu = (\mu_1, \mu_2, \mu_3)^T$, and covariance matrix $\Sigma_{3\times 3}$. The sample mean and the sample covariance matrix are

$$\bar{X} = (1.23, 1.96, 0.84), S = \begin{pmatrix} 0.2994 & 0.1774 & 0.1278 \\ 0.1774 & 0.1075 & 0.0747 \\ 0.1278 & 0.0747 & 0.0693 \end{pmatrix}.$$

Using the Hotelling's T^2 statistic, test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$, against the alternative that all the μ_i's are not equal (at level $\alpha = 0.05$). [8]

2. (a) Explain how, and under what conditions, the delete-1 jackknife is successful in reducing the bias of the estimator of the target parameter.

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution F. Suppose that the median functional is defined as $T(F) = F^{-1}(1/2)$. Then

$$T(F_n) = F_n^{-1}(1/2) = X_{([n+1]/2)}.$$

where F_n is the empirical distribution function based on $(X_1, X_2, \ldots, X_n)$. Let $(X_1^*, X_2^*, \ldots, X_n^*)$ be a bootstrap resample drawn from the original sample, and let F_n^* be the empirical distribution based on this resample, and $T(F_n^*)$ is the m-th order statistic of the bootstrap resample, where $m = ([n+1]/2)$. Show that, conditional on the original sample

$$P(T(F_n^*) = X_{(k)}) = \sum_{j=0}^{m-1} \left\{ \binom{n}{j} \left(\frac{k-1}{n}\right)^j \left(1 - \frac{k-1}{n}\right)^{n-j} - \binom{n}{j} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \right\},$$

for $k = 1, 2, \ldots, n$, where $X_{(k)}$ is the k-th order statistic of the original sample.

[5+5=10]

3. A diagnostic test for gout is based on serum uric acid level. Using appropriate units the level is approximately distributed as $N_1(5, 10)$ among healthy adults and $N_1(8.5, 1)$ among adults with gout. If 1% of the population has gout, how should a patient with level x be classified so that the total probability of misclassification is minimized? (You need to derive the rule, not just write down the formula). Also calculate the posterior probabilities when $x = 7$, and indicate the corresponding classification. [8]

4. Suppose a random variable is modeled by the $N(\mu, \sigma^2)$ family, where $\sigma^2 = 2$ is known. An i.i.d. random sample drawn from this distribution yields the values 2, 6, 4, 3 and 5. We are interested in testing the hypothesis $H_0 : \mu = 3$. Perform the likelihood ratio test, the Wald test and the score test for this hypothesis based on these data. Give your conclusions and state whether the tests (or some of the tests) are equivalent in this case. [9]

5. (a) Consider the data represented in the following 2×2 table, obtained through a retrospective study design. What is the estimated odds ratio, and how do you interpret it? Use these data to set up a 95% confidence interval for the odds ratio.

	Cancer	Control	
Smokers	83	72	155
Nonsmokers	3	14	17
	86	86	172

(b) For the Donner party data discussed in class, the fitted logistic regression model is

$$\text{logit}(\hat{\pi}) = 1.63 - 0.078\ age + 1.60\ fem.$$

Here π is the survival probability, the age variable has its usual meaning, and the variable fem is a binary variable which is indicator of the female sex (takes value 1 for females and 0 for males). Find the odds ratio of survival for females over males of the same age, and interpret it. [(4+2)+(3+1)=10]

6. Under appropriate assumptions, find the seasonally adjusted values for the following time series using an appropriate method. (That is, find the values of the series after the seasonal component has been removed).

Year	Quarter	Series
1	I	28.4
	II	33.8
	III	37.7
	IV	32.6
2	I	32.7
	II	37.0
	III	41.3
	IV	36.5
3	I	36.0
	II	41.3
	III	45.1
	IV	40.9
4	I	40.6
	II	33.2
	III	37.4
	IV	32.2

[10]