Civil Services (Main) Examination, 2013

STATISTICS

(Paper I)

Time Allowed : Three Hours **Maximum Marks : 250**

QUESTION PAPER SPECIFIC INSTRUCTIONS

(Please read each of the following instructions carefully before attempting questions)

There are EIGHT questions divided in two Sections.
Candidate has to attempt FIVE questions in all.
Question Nos. 1 and 5 are compulsory and out of the remaining, any THREE are to be attempted choosing at least ONE question from each Section.
The number of marks carried by a question/part is indicated against it.
Wherever any assumptions are made for answering a question, they must be clearly indicated.
Word limit in questions, wherever specified, should be adhered to.
Attempts of questions shall be counted in sequential order. Unless struck off, attempt of a question shall be counted even if attempted partly. Any page or portion of the page left blank in the Question-cum-Answer Booklet must be clearly struck off.

SECTION 'A'

1. (a) In the course of an experiment with a particular brand of DDT on flies, it is found that 80% are killed in the first application. Those which survive develop a resistance, so that the percentage of survivors killed in any later application is half that in the preceding application. Thus 40% of the survivors of the first application would succumb to the second, 20% of the survivors of the first two applications would succumb to the third, and so on. Find the probability that (i) a fly will survive four applications. (ii) it will survive four applications, given that it has survived the first one. **10**

(b) The joint distribution of two random variables X and Y has the pdf

$$f_{X,Y}(x,y)=\beta^{a+b}x^{a-1}(y-x)^{b-1}e^{-\beta y},\ 0<X<Y<\infty;\ =0 \text{ otherwise}$$

Derive the marginal distribution of X and the conditional distribution of Y given X. **10**

(c) Let X_1, X_2 be independent random variables following Poisson distribution with parameter λ. Show that X_1+X_2 is sufficient for λ. Verify if X_1+2X_2 is also sufficient for λ. **10**

(d) Let X be a negative Binomial random variable with parameters r and θ, r being known and θ unknown. Suppose the problem is to test $H_0:\theta=1/2$ against $H_1:\theta>1/2$. Find the UMP test for the problem and obtain its power function. **10**

(e) Use sign test statistic to find a non-trivial distribution free confidence interval for population median of a continuous distribution. **10**

2. (a) If $\varphi(t)$ denotes the characteristic function (ch.f.) of a random variable with distribution function $F(x)$, show that $e^{[\varphi(t)-1]}$ is a characteristic function. Hence find its corresponding distribution function. **20**

(b) For the Laplace distribution

$$f(x)=1/2\lambda e^{(-|x-\mu|)/\lambda},\ -\infty<x<\infty,\ \lambda>0,\ -\infty<\mu<\infty$$

find the moment generating function. **15**

(c) Let (X,Y) be jointly distributed with probability density function

$$f_{X,Y}(x,y,\theta)=\exp[-(x/\theta+\theta y)];\ X>0,\ Y>0;\ =0 \text{ otherwise}$$

Find Fisher information in a sample of n pairs. **15**

3. (a) Using Rao – Blackwell Theorem, derive the uniformly minimum variance unbiased estimator (UMVUE) of an estimable parametric function based on a complete sufficient statistic. Obtain the UMVUE of $P[X=r]$, $r=0,1,\ldots m$ where $X_1,X_2,\ldots X_n$ are independently identically distributed (i.i.d.) Bin(m,p). **20**

Optional Province · optionalprovince.com

(b) For one parameter exponential family obtain UMPU test for testing $H_0:\theta=0$ against $H_1:\theta\neq\theta_0$. Further prove that $E[\varphi(X)\,T(X)]=\alpha\,E[T(X)]$. **15**

(c) Let X be distributed as Bin(n,p) and $\pi(p)=1$ $(0<p<1)$ be the prior probability density of p. Then obtain the Bayes estimator (under squared error loss) and the Bayes risk. **15**

4. (a) Let N be the number of observations required by the sequential probability ratio test (SPRT) for testing $H_0:X\sim f(x,\theta_0)$ against $H_1:X\sim f(x,\theta_1)$, $\theta_1\neq\theta_0$. Stating appropriate assumptions, show that $P(N<\infty)=1$ at $\theta=\theta_0,\theta_1$. When $X\sim N(\theta,1)$, for testing $H_0:\theta=0$ against $H_1:\theta=1$, derive the expression of the percentage saving in sample sizes under H_0 based on SPRT over the fixed sample size based MP test. Take the strength of each procedure as (α,β). **20**

(b) State the strong law of large numbers. Decide whether the strong law of large numbers holds for the sequence of mutually independent random variables X_k with distribution

$$P\{X_k=\pm 2^k\}=2^{-(2k+1)},\ P\{X_k=0\}=1-2^{-2k}$$ **15**

(c) To compare two brands of tyres, the following milages (in '000' miles) were obtained for 8 tyres of each kind: Brand A: 32.1, 20.6, 17.8, 28.4, 19.6, 21.4, 19.9, 30.1 Brand B: 19.8, 27.6, 30.8, 27.8, 34.1, 18.7, 16.9, 17.9 Test the null hypothesis at $\alpha=0.05$ that the two samples come from the same population using Mann – Whitney test. [Table value: $U(8,8,0.025)=50$, $U(8,8,0.05)=48$] where $P[U>U(m,n,\alpha)]\leq\alpha$. **15**

SECTION 'B'

5. (a) Consider the following linear model:

$$y_1=\theta_1+\theta_2+\theta_3+\theta_4+\varepsilon_1,\ y_2=\theta_1+\theta_3-\theta_2-\theta_4+\varepsilon_2,\ y_3=\theta_1+\theta_2-\theta_3-\theta_4+\varepsilon_3,\ y_4=\theta_1+\theta_4-\theta_2-\theta_3+\varepsilon_4$$

where $\theta_1,\theta_2,\theta_3$ and θ_4 are parameters, $\varepsilon_j\sim(0,\sigma^2)$, $j=1\ldots4$. ε_j's are all independently distributed. Derive normal equations and obtain the BLUE of $\theta_1+\theta_2+\theta_3+\theta_4$. Also find the variance of the BLUE. **10**

(b) Consider a model $y_{ij}=\mu+\alpha_i+\varepsilon_{ij}$, $i=1,\ldots k$, $j=1,\ldots n$, ε_{ij}'s are i.i.d. $N(0,\sigma^2)$. Obtain least squares solutions of μ and α_i. When is a linear parametric function of α_i's (only) estimable? Derive a test statistic for testing $H_0:\sum_{i=1}^{n} C_i\alpha_i=0$. **10**

(c) Show that systematic sample is better than SRSWOR if $-1/(n-1)<\rho_w<-1/(N-1)$ where n and N are sample and population sizes respectively and ρ_w is the intraclass correlation between pairs of units that are in the same systematic sample. **10**

(d) What do you understand by a connected block design? The following is the incidence matrix of a block design with four treatments and five blocks:

$$[[1\ \ 1\ \ 1\ \ 0\ \ 0\,;1\ \ 0\ \ 0\ \ 1\ \ 1\,;0\ \ 1\ \ 0\ \ 1\ \ 0\,;0\ \ 0\ \ 1\ \ 0\ \ 1]]$$

Compute the C-matrix of the design. **10**

(e) Define multiple correlation coefficient $\rho_{1.23\ldots p}$ between X_1 and $X_2,X_3,\ldots X_p$ and show that $1-\rho_{1.2\ldots p}^2=(\sigma_{11}\sigma^{11})^{-1}$ where σ_{11} and σ^{11} are $(1,1)^{\text{th}}$ elements of the dispersion matrix Σ and Σ^{-1} respectively. **10**

6. (a) Let C be the C-matrix of a block design with parameters (v,b,r,k) and let Q be the vector of adjusted treatment totals. Show that, under the usual additive fixed effects linear model $E(Q)=C\boldsymbol{\tau}$ and $D(Q)=C\sigma^2$ where $\boldsymbol{\tau}$ is the vector of treatment effects and σ^2 is the common experimental error variance. Also show that $l'\boldsymbol{\tau}$ is estimable if and only if l' lies in the row space of C. **20**

(b) Let $X_1\ldots X_N$ $(N>p)$ be a random sample from $N_p(\mu,\Sigma)$, $\Sigma>0$ where $\mu=(\mu_1\ldots\mu_p)'$ and Σ are unknown. Show that for testing $H_0:\mu_1=\mu_2=\ldots=\mu_p$ against H_1: not all μ_i's are equal, the test statistic can be expressed as

$$T=N[\bar{X}'\,S'\,\bar{X}-(\bar{X}'\,S'\,\varepsilon)^2/\varepsilon'\,S'\,\varepsilon]$$ **15**

Optional Province · optionalprovince.com

where $\bar{X}$ and S are the sample mean vector and the unbiased estimator of Σ respectively and $\varepsilon=(1,1,\ldots 1)'$. What would be the non null distribution of T?

(c) Suppose it is required to estimate the average value of output of a group of 5000 factories in a region so that the sample estimate lies within 10% of the true value with confidence coefficient of 95%. The population coefficient of variation is known to be 60%. Determine the equation to calculate minimum sample size. **15**

7. (a) Show that for a BIBD with $v=nk$ for n, a positive integer, $b\geq v+r-1$. Show that for a symmetrical BIBD with v even, $r-\lambda$ is a perfect square. **20**

(b) Define Mahalanobis measure of distance squared between two populations with a common positive definite dispersion matrix and show that it is a non decreasing function of the number of characteristics. **15**

(c) A rural block in a district was divided into 3 strata. The following table gives the number of villages (N_h) and standard deviation (S_h) of area under wheat crop for different strata:

Stratum No.	No. of villages (N_h)	Standard deviation (S_h)
1	15	100
2	10	50
3	25	120

A sample of 20 villages is to be drawn. How many villages should be selected from each strata under (i) proportional allocation? (ii) Neyman allocation? **15**

8. (a) Let $X_{2p\times 1}=[[X_1 ; X_2]]$ be distributed as $N_{2p}(\mu,\Sigma)$ where $\mu=[[\mu_1 ; \mu_2]]$, $\Sigma=([\Sigma_1 \quad \Sigma_2 ; \Sigma_2 \quad \Sigma_1])$, X_1,X_2,μ_1 and μ_2 are p-vectors and Σ_1,Σ_2 are $p\times p$ matrices. Obtain the characteristic function of X. Derive the distribution of $[[X_1+X_2 ; X_1-X_2]]$. **20**

(b) Let $X=(X_1,X_2,X_3)'$ has the correlation matrix R given by

$$R=[[1 \quad 1/4\sqrt{2} \quad 0 ; 1/4\sqrt{2} \quad 1 \quad 1/4\sqrt{2} ; 0 \quad 1/4\sqrt{2} \quad 1]]$$

Obtain the first two principal components and the population variance explained by the first two principal components. **15**

(c) Define a ratio estimator. Show that it is biased. Suggest a suitable sampling procedure for which the classical ratio estimator of a population ratio is unbiased. For the suggested procedure, propose an unbiased estimator of the variance of the ratio estimator. **15**

Optional Province · optionalprovince.com