Sanjoy Bhattacharya
19/11/18

FOR GOD AND COUNTRY

ST. LAWRENCE HIGH SCHOOL

SELECTION TEST SOLUTION

Subject:Statistics **Class: XII** **F. M. 70**
Duration: 3 HRS. 15 MIN **Date: 17.11.2018**

PART : A

Q1. Answer the following questions. **2X4**

i. When does a discrete random variable follow Normal distribution?

Ans. For large sample $\frac{X-E(X)}{V(X)} \sim N(0,1)$

ii. What is power of test? OR Write a short note on percentile scaling.

Ans. Probability of rejecting false null hypothesis = 1 – P(type II error).

OR The distribution trait must be Rectangular.To determine the scale value corresponding to a raw score, the percentile position to be considered.

iii. Find V(X) in case of uniform distribution.

Ans. $E(X) = \frac{1}{n}\sum_{i=0}^{n-1}(a+ih) = \frac{1}{n}(na + h\sum_{i=0}^{n-1} i) = a + \frac{1}{2}h(n-1)$

$V(X) = E(X-E(X))^2 = \frac{1}{n}\sum_{i=0}^{n-1}((a+ih) - (a+\frac{1}{2}h(n-1)))^2 = \frac{h^2}{12}(n^2-1)$

iv. Write the components of time series data. OR

Write the normal equations in parabolic trend fitting.

Ans. Components: secular trend(T), seasonal variation(S), cyclical variation(C), irregular fluctuation(I).

OR. For parabolic trend fitting of $y = a + bt + ct^2$, the normal equations are

$\sum y = an + b\sum t + c\sum t^2$

$$\sum yt = a\sum t + b\sum t^2 + c\sum t^3$$

$$\sum yt^2 = a\sum t^2 + b\sum t^3 + c\sum t^4$$

Q2. Answer the following questions. **3X8**

i. Find and unbiased estimator of population variance where population units follow $N(\mu, \sigma^2)$

Ans. $E(s^2) = \frac{1}{n}\sum Ex_i^2 - E(m^2) = \frac{1}{n}\sum(v(x_i) + E(x_i)^2) - (v(m) + E(m))^2$

$= \frac{n-1}{n}\sigma^2$. So the unbiased estimator of population variance is $\frac{n-1}{n}s^2$. Where m = sample mean.

ii. Determine f(x), the p.m.f., from $f(x) = \frac{\lambda}{x}f(x-1), x = 1,2,3,..$, where f(x) is non-zero for non-negative integral values of the random variable x. Find also the probability that X is greater than zero.

Ans. $f(x) = \frac{\lambda^x}{x!}f(0)$, ie, Since total probability is 1, $f(0) = (e^\lambda - 1)^{-1}$. So $f(x) = \frac{\lambda^x}{x!}.(e^\lambda - 1)$

iii. Derive mean deviation about mean of a random variable X ≈ Poisson(m).

OR Find the expectation of number of throws required to get the third '6' in repeated throw of an unbiased die.

Ans. EI X- λI = $2\sum_{i=k+1}^{\infty}(x_i - \lambda)f(x) = 2(\gamma_x - \gamma_{x+1})where\ \gamma_x = e^{-\lambda}\frac{\lambda^x}{(x-1)!}$

So mean deviation is $2\gamma_{k+1}$

OR. Let X be the no of throws required to get third 6.

$f(x) = (\frac{5}{6})^{x-1}\frac{1}{6}$. So $E(X) = \sum_{x=1}^{\infty} x.(\frac{5}{6})^{x-1}\frac{1}{6}. = 6$

500 and standard deviation Rs 50. Find the minimum wage of 10 highest paid workers. Given $\int_0^x \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 0.4$ for x= 1.286. OR

Derive the points of inflection in normal distribution.

Ans. X: wage of workers. a: minimum wage of 10 highest paid workers.

$P(X \geq a)=0.1$, ie, $\Phi(\frac{a-500}{\sqrt{50}}) = 0.9, ie, a = 500 + \sqrt{50}.1.286 = 509.09 = 509.$

OR. $\frac{d^2}{dx^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = 0$, ie, $x = \mu - \sigma$ and $\mu + \sigma$

v. Write a short note on moving average method used in time series. OR

Write a short note on Bias in sample survey

Ans. In moving average meyhod we can not find the secular trend values for first n and last n years for 2n or (2n+1) yearly moving average and for this reason we can not forecast in this method which can be done using curve fitting method. In case of inclusion or rectification of some values moving average is better.

OR. A. Procedural bias : Response bias, Interviewer bias, Non response bias, Observational bias.

B. Sampling bias: Bias due to defective sampling method
Bias due to substitution
Bias due to faulty demarcation of units
Bias due to wrong selection of statistic.

vi. Write advantages of sampling over complete enumeration.

Ans. Greater speed, less cost, more scope and coverage, measurement of error and minimisation of it and greater accecibility

vii Derive the expression of standard error of estimate of y from the regression

equation y on x.

Ans. v(e) $= \frac{1}{n}\sum(y_i - Y_i)^2 = {s_y}^2 - 2r^2{s_y}^2 + r^2{s_y}^2 = {s_y}^2(1-r^2)$. Square root of which gives standard error of estimate of y from the regression

equation y on x.

viii. Write the control limits of mean chart for both the cases when standards are given and not given. OR

Write the control limits of range chart for both the cases when standards are given and not given.

Ans. MEAN CHART

Case1: standard given σ= σ'

LCL= μ' - $\frac{3}{\sqrt{n}}\sigma'$ = μ' - Aσ', CL= = μ', UCL= μ' + $\frac{3}{\sqrt{n}}\sigma'$ = μ' + Aσ'

Case1: standard not given. Replace μ' by $\bar{\bar{x}}$ and σ by $\frac{\bar{s}}{c_2}$ or $\frac{\bar{R}}{d_2}$ to get the limits

LCL = $\bar{\bar{x}} - A_1\bar{s}$, CL =$\bar{\bar{x}}$, UCL = $\bar{\bar{x}} + A_1\bar{s}$

LCL = $\bar{\bar{x}} - A_2\bar{R}$, CL =$\bar{\bar{x}}$, UCL = $\bar{\bar{x}} + A_2\bar{R}$

OR. Range chart:

standard given $\sigma = \sigma'$

LCL= $d_2\sigma' - 3D\sigma' = D\sigma'_1$, CL== $d_2\sigma'$, UCL == $d_2\sigma' + 3D\sigma' = D_2\sigma'$

Standard not given

Replace σ by $\frac{\bar{R}}{d_2}$

LCL= $\bar{R} - 3D\frac{\bar{R}}{d_2} = D_3\bar{R}$, CL==$\bar{R}$, UCL =$\bar{R} + 3D\frac{\bar{R}}{d_2} = D_4\bar{R}$

Q3. Answer the following questions. **5X4**

i. Give the theory behind control chart technique.

Ans. The following points are to be mentioned.

a. Selection of rational subgroup.
b. Calculation of proper statistic.
c. $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

Chebyshev's inequality $P(\mu - 3\sigma < T < \mu + 3\sigma) > 8/9$.

ii. Derive the expression of standard error of sample mean in SRSWOR. OR

Derive the expression of standard error of sample proportion in SRSWOR.

Ans. Here x_i *and* x_j *are not independent*. $E(x_i) = \mu$, $v(x_i) = \sigma$

and cov $(x_i, x_j) = \frac{\sigma^2}{N-1}$. $V(\bar{x}) = \sqrt{\frac{N-n}{N-1}} \frac{\sigma}{\sqrt{n}}$.

OR. Here replace *by* $\sqrt{p(1-p)}$.

iii. Derive the expression for coefficient of determination from regression equation y on x for n bivariate observations. OR

Derive the expression of regression equation of y on x.

Ans. $V(Y) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{y} + r\frac{s_y}{s_x}(x_i - \bar{x}) - \bar{y})^2 = r^2 s_y^2$

$IrI = \frac{v(Y)}{v(y)}$ is coefficient of determination.

OR. In a scatter diagram let ith plotted point be(x_i, y_i) and the predicted best fitted line be Y=a+bX . $e_i = (y_i - Y_i) = (y_i - a - bx_i)$

Sum of square of errors $E = (y_i - a - bx_i)^2$

Minimizing E by differentiating partially w.r.t.a and b we get the normal equations

$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$ and $\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2$

Solving we get, $\hat{b} = \frac{cov(x,y)}{V(x)} = b_{xy}$. Hence the regression equation is

$y - \bar{y} = b_{xy}(x - \bar{x})$.

iv. Let λ be the average number of misprints per page of a book. To test the hypothesis $H_0 : \lambda = 0.5$ against the alternative hypothesis $H_a : \lambda = 1$, if a page chosen at random from the book contains more than one misprint, then the hypothesis $\lambda = 0.5$ is rejected. Find the probabilities of type I error and type II error, and the power of the test. (Given $e^{-\frac{1}{2}} = 0.61$).

Ans. P(type I error) = P(x>1I λ=0.5) =1- (1+ 0.5) $e^{-0.5}$= 0.09

P(type II error) = P(x≤1I λ=1)= e^{-1}(1+1)= 2/e= 0.735

Power of test = 1- 0.735= 0.265

PART B

Q1. Select the correct alternatives. **1x10**

i. The correlation coefficient between x and y in $2x + 3y = 7$ is

a. 1 **b. -1** c. 0 d. none of these

ii. The unbiased estimator of population mean is

a. sample mean b. sample median c. sample median d. none of these

iii. The maximum probability of type I error is

a. level of significance b. power of test c. Type II error d. none of these

iv. Observational bias is

a. sampling bias **b. non sampling bias** c. response bias d. none of these

v. A random variable follows N(25 , 9). Then P(X = 7) is

a.0 b. 1 c. 0.5 d. none of these

vi. Which of the following control chart is used for attribute%

a. mean chart b. range chart **c. fraction defective** d. none of these

vii. X follows Poisson(4), then C.V. is equal to

a. 25% **b.50%** c.75% d. none of these

viii. The correlation coefficient between religion and income is

a. negative b. positive c. zero **d. none of these**

ix. Pdf is the

a. probability **b. Area** c. cumulative probability d. none of these

x. The unbiased estimator of P when X follows Bin(n, P) is

a. $\frac{X}{n}$ **b.** $\frac{\bar{x}}{n}$ c. nX d. none of these

Q2.. Answer the following questions. **1X8**

i. Show that unbiasednessof the estimator $\hat{\theta}$ of the parameter θ does not necessarily imply that

$\sqrt{\hat{\theta}}$ will be an unbiased estimator of $\sqrt{\theta}$,

Ans. $v(\sqrt{\hat{\theta}}) > 0$,ie, $E(\hat{\theta})^2 > (E(\sqrt{\theta}))^2$

ii. What do you mean by composite hypothesis?

Ans. The hypothesis which does not completely specify the parameter.

iii. What are probability and non-probability sampling?

Ans. when only some population units have probability 1 and other have probability zero is called non-probability sampling and when all the population units have some probability then that is known as probability sampling.

iv. What is meant by assignable causes in statistical quality control?

Ans. The manufactured products are to be divided into several rational subgroups where each subgroup cosists of homogeneous items and the variation occurs between two subgroups is due to assignable cause.

v. What is fpc ?

Ans. Finite population correction. $\sqrt{\frac{N-n}{N-1}}$.

vi. Write the moment estimator of p in Bin (n,p).

Ans. $\frac{\bar{x}}{n}$

vii. Define rational subgroup.

Ans. The manufactured products are to be divided into several subgroups such that s each subgroup cosists of homogeneous items and if any variation due to chance cause and the variation occurs between two subgroups is due to assignable cause.

viii. If a Binomial distribution has mode X = 3 and 4, then find C.V. of the distribution.

Ans. $\sqrt{1-p}100\%$ where p is probability of success.