ST. LAWRENCE HIGH SCHOOL

FOR GOD AND COUNTRY

PRETEST EXAMINATION

Subject: Statistics **Class: XII** **F. M. 70**

Duration: 2 HRS. 30 MIN **Date: 04.08.2018**

PART A

1. Answer the following questions. 2X4

i. Show that normal distribution is symmetric about mean.

ii. Give an example each of (a) Increasing trend and (b) Decreasing trend.

OR

State two differences between seasonal variation and cyclical variation.

iii. Define scatter diagram.

iv. The coordinates of two points on a scatter diagram are (10,6) and (8.14).Find the correlation coefficient between x and y.

2. Answer the following questions. 3X8

i. A random variable X follows Binomial distribution with parameters n and p. Find mean deviation about mean. OR

Find the quartile deviation of rectangular (1,6).

ii. Obtain the expression of coefficient of determination from the regression linen y on x.

iii. Prove that correlation coefficient lies between – 1 & +1. OR

For any two random variables X & Y, show that V(X+Y) = V(X) + V(Y) +2COV(X,Y)

iv. Derive the mean deviation about median of Uniform distribution (discrete).

OR

The pmf f(x) of random variable x, assuming values 0,1,..,15 satisfies the relation

$f(x)=\frac{16-x}{x}\ \frac{1}{2}f(x-1)$. Determine f(x).

v. A perfect coin is tossed 3 times in succession. Given X = 1 if first toss gives head, X = 0 if first toss gives a tail, and Y = number of heads obtained in 3 tosses, construct the joint distribution of X and Y, and find the correlation coefficient between them.

vi. Write the important properties of normal distribution.

vii. Describe the least square method in time series analysis.

OR

Describe the method of moving average in times series analysis.

viii. For 5 pairs of values of x and y, the values of x + y are 24, 28, 30, 33, 35 and variances of x and y are 6 and 2 respectively. Calculate the correlation coefficient between x and y.

OR

Each of two persons tosses an unbiased coin 10 times. Find the probability that both of them get same number of heads.

3. Answer the following questions. 5X4

i. Derive the expression for Spearman's Rank correlation coBinefficient.

ii. Derive the expression of standard error of estimate of y from its linear regression on x. Hence determine the possible range of correlation coefficient . OR

Derive the regression line X on Y from a scattered diagram having n points.

iii. An unbiased coin is tossed 200 times. Find the probability that number of heads obtained is between 110 and 120 (inclusive both).

Given Area under standard normal curve between 0 and 2.89 and 1.30 are 0.498 and 0.403 respectively. OR

A random variable X follows poisson (m). Find the skewness of the distribution.

iv. Write a short note on components of time series analysis.

PART B

1. Select the correct answer from the following options. 1X10

i. If x~ R(a, b) , then E(X) is

a. (a+ b)/2 b. (a-b)/2 c.(b-a)/2 d.$(b-a)^2/2$

ii. If $4u = 2x + 7, 6v = 2y - 15$ and the regression coefficient of y on x is 3, then the regression coefficient of v on u is

a 2 b. 3 c 9/2 d none of these

iii. If $3y - 2x = 9$ is the regression line of variable y on variable x, correlation coefficient between x and y is 1/3, and variance of x is 4, then variance of y is

a 4 b. 9 c. 1 d. none of these

iv. First three moments (μ_1', μ_2', μ_3') are same for

a binomial distribution b Poisson distribution

c. exponential distribution d none of these

v. Mean and s.d. are same for

a binomial distribution b Poisson distribution

c uniform distribution d exponential distribution.

vii. The correlation coefficient between income and religion is

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

viii. If the trend equation fitted from a given set of data of production of a fertilizer factory is 13y = 8424 + 312t considering time unit 1 year, the monthly increase in production of fertilizer in thousand tones is

a. 54 b. 24 c.2 d. none of these

ix. Two regression lines coincide then r is equal to

a. ±1 b. +1 c. -1 d. none of these

x. The probability of a continuous random variable is denoted by

a. pdf b. cdf c. pmf d. none of these

1. Answer the following questions. 1X8

i Write down the pmf of a symmetric binomial distribution.

ii. What is pmf? OR Define correlation.

iii. Give an example of cyclical variation.

iv. 2x +3y = 5 and 3x +2y = 5 are two regression lines.Find the values of mean of x and mean of y.

v. What is a parameter of a probability distribution?

vi. X ~ Bin(n, p). Find probability of X> n.

vii. Show that the sum of two regression coefficient is greater than twice the correlation coefficient.

viii. Define time series data.

....

Sanjay Bhattacharya
9/8/18

ST. LAWRENCE HIGH SCHOOL

PRETEST EXAMINATION

FOR GOD AND COUNTRY **Subject: Statistics** **Class: XII** **F.M. 70**

~~M. 70~~

Duration: 2 HRS. 30 MIN **Date: 04.08.2018**

Part 1

Q1.i. $f(\mu - x) = f(\mu + x)$. Hence symmetric about μ.

ii. (a)Population growth in India. (b) Volume of glacier.

OR In case of seasonal variation time period is one year but in case of cyclical trend time period is more than one year.

iii. Bivariate observations are being plotted on a graph with two axes is known as scatter diagram.

OR Correlation coefficient between x and y = -1.

Q2.i. MD $\mu(x) = \sum_{x=k+1}^{n} (x - np)\, n_{c_x}\ p^x\ (1-p)^{n-x}$. Where k=[np]

$= \gamma_{(k+1)}$.

Where $\gamma_{(k+1)} = \sum_{x=k+1}^{n} |x - np|\, n_{c_x}\ p^x\ (1-p)^{n-x}$.

$= 2np(1-p)\ n_{c_k}\ p^k\ (1-p)^{n-1-k}$.

ii. $Q_3 = 4.25$ and $Q_1 = 2.25$. QD = (4.25 – 2.25)/2 = 1.

iii. Take $u_i = \frac{x_i - \bar{x}}{s_x}$ and $v_i = \frac{y_i - \bar{y}}{s_y}$. By C.S. inequality $-1 \le r \le 1$.

OR $V(X+Y) = \sum_{i=1}^{n}(x_i + y_i - \bar{x} - \bar{y})^2 = \sum_{i=1}^{n}((x_i - \bar{x}) + (y_i - \bar{y}))^2$

Expanding V(X+Y) = V(X) + V(Y) +2COV(X,Y).

iv. mean deviation about median = $\sum_{i=k+1}^{n}\left(a + ih - a - \frac{1}{2}(n-1)\right)\frac{1}{n}$.

Where k= $[a - \frac{1}{2}(n-1)]$.

v.

x\y	0	1	2	3
0	1/8	2/8	1/8	0
1	0	1/8	2/8	1/8

$r = \frac{E(XY) - E(X).E(Y)}{s_X s_Y} = 0.577$

vi. Normal distribution shows a bell shaped, symmetric, assymptotic, mesokurtic, thin-tail curve. Mean, median and mode lie at the same point $x = \mu$. The ratio between the quartile deviation and standard deviation is 76:100. Points of inflexion lie at $\mu \pm \sigma$. $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$. X be any random variable discrete or continuous what-so-ever $\frac{X - E(X)}{\sqrt{V(X)}}$ follows N(0,1). Using th transformation $Z = \frac{X-\mu}{\sigma}$, Z follows N(0,1).

vii. Write the mathematical model of linear, parabolic and exponential form. The scale of time has to be taken for both odd and even number of time periods. Using the normal equations the unknown constants to be found out and by replacing in the original equation the trend equation can be found. In this method the trend values of extreme classes can not be found and hence it can not be used for forecasting.

viii. V(X+Y) = V(X) + V(Y) +2COV(X,Y)

ie, r = 0.98

OR $10_{c_5} 2^{-10}$

3.i.

Spearman's rank correlation coefficient can be derived as the simple product-moment correlation coefficient taking u_i and v_i as variate values.

$$\bar{u} = \frac{1}{n}\sum_i u_i = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

and $$s_u^2 = \frac{1}{n}\sum_i (u_i - \bar{u})^2 = \frac{1}{n}\sum_i u_i^2 - \bar{u}^2$$

$$= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12},$$

Similarly, $\bar{v} = \frac{n+1}{2}$ and $s_v^2 = \frac{n^2-1}{12}$.

Again, $$\frac{1}{n}\sum_i d_i^2 = \frac{1}{n}\sum_i (u_i - v_i)^2 = \frac{1}{n}\sum_i \{(u_i - \bar{u}) - (v_i - \bar{v})\}^2, \text{ (since } \bar{u} = \bar{v})$$

$$= \frac{1}{n}\sum_i (u_i - \bar{u})^2 + \frac{1}{n}\sum_i (v_i - \bar{v})^2 - \frac{2}{n}\sum_i (u_i - \bar{u})(v_i - \bar{v})^2$$

$$= s_u^2 + s_v^2 - 2\,\text{Cov}(u,v) = \frac{n^2-1}{6} - 2\,\text{Cov}(u,v).$$

$$\therefore \text{Cov}(u, v) = \frac{n^2-1}{12} - \frac{1}{2n}\sum_i d_i^2.$$

$$\text{So, } r_{uv} = \frac{\text{Cov}(u,v)}{s_u s_v} = \frac{\frac{n^2-1}{12} - \frac{1}{2n}\sum_i d_i^2}{\frac{n^2-1}{12}} = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)} = r_R.$$

ii.

(4) $\text{Var}(e) = \frac{1}{n}\sum_i e_i^2, \quad (\because \bar{e} = 0)$

$$= \frac{1}{n}\sum_i (y_i - Y_i)^2 = \frac{1}{n}\sum_i \left\{(y_i - \bar{y}) - r\frac{s_y}{s_x}(x_i - \bar{x})\right\}^2$$

$$= \frac{1}{n}\sum_i (y_i - \bar{y})^2 - 2r\frac{s_y}{s_x}\cdot\frac{1}{n}\sum_i (y_i - \bar{y})(x_i - \bar{x}) + r^2\frac{s_y^2}{s_x^2}\cdot\frac{1}{n}\sum_i (x_i - \bar{x})^2$$

$$= s_y^2 - 2r\frac{s_y}{s_x}.rs_xs_y + r^2\frac{s_y^2}{s_x^2}.s_x^2 = s_y^2 - 2r^2s_y^2 + r^2s_y^2 = s_y^2(1 - r^2).$$

So, the standard deviation of e (which is called the *standard error of estimate* y from its linear regression on x) $= s_y\sqrt{1-r^2}$.

Since $\text{Var}(e) \geq 0$, we get

$s_y^2(1 - r^2) \geq 0$ or, $1 - r^2 \geq 0$ or, $r^2 \leq 1$.

$\therefore \quad -1 \leq r \leq 1$.

iii. X follows Bin(200 , ½)

E(x) =100 and V(x) =50.

$P(\frac{109.5-100}{\sqrt{50}} < Z < \frac{120.5-100}{\sqrt{50}}) = \Phi(2.9) - \Phi(1.3) = 0.95$.

iv. Write four components and give the definitions with one example each.

Part B

1.i.a ii.a iii. d iv. b and c v.b vii.d viii. ix. a x.b

2.i. $f(x)= n_{c_x} \left(\frac{1}{2}\right)^n$

ii. Probability mass function.

iii. Recession

iv. 1 and 1 respectively.

v. Which specifies the distribution

vi. 0

vii. Use am > gm taking the reg.coefficients as observations.

viii. Data which depend upon time.