SECOND YEAR HIGHER SECONDARY EXAMINATION MARCH 2018

SUBJECT: STATISTICS **CODE NO. 9020**

Qn No	Scoring key	Split score	Total score
1 a 1 b	(iv) 0.5 (ii) 20	1 1	2
2	A — B (i) No correlation — (d) r =0 (ii) $\frac{Cov(x,y)}{\sigma_x \times \sigma_y}$ — (c) Correlation coefficient (iii) $1-\frac{6\sum d^2}{n^3-n}$ — (a) Rank correlation coefficient (iv) Perfect correlation — (b) r = ± 1	½ × 4	2
3	$\sum X = 247, \sum Y = 263, \sum X^2 = 7345, \sum Y^2 = 7537, \sum XY = 7259, n = 10$ $r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2} \times \sqrt{n\sum Y^2 - (\sum Y)^2}}$ $= \frac{10 \times 7259 - 247 \times 263}{\sqrt{10 \times 7345 - 247^2} \times \sqrt{10 \times 7537 - 263^2}}$ $= \frac{7629}{\sqrt{12441 \times 6201}} = 0.87$ (or $r = \frac{Cov(x,y)}{\sigma_x \times \sigma_y}$) $r = \frac{76.29}{11.15 \times 7.87} = 0.87$	1 ½ ½	2
4	Since f(x) is a pdf, we have $\int_{-\infty}^{+\infty} f(x)dx = 1$ $ie, \int_0^2 kx\, dx = 1$ $\Rightarrow k\left[\frac{x^2}{2}\right]_0^2 = 1 \quad \Rightarrow k = \frac{1}{2}$	½ ½ 1	2
5	Simple AM price index = $\frac{\sum x}{n}$, where x is the price relative. Here $\sum x = 111.11 + 122.86 + 110.42 = 344.39$, n = 3 $\therefore$ Simple AM price index = $\frac{344.39}{3} = 114.797$	1 ½ ½	2
6	$b_{yx} = 0.23,\ \gamma = 0.45,\ \sigma_x = 10,\ \sigma_y = ?$ We have, $b_{yx} = \gamma \frac{\sigma_y}{\sigma_x}$ $\Rightarrow 0.23 = 0.45 \times \frac{\sigma_y}{10}$ $\Rightarrow \sigma_y = 5.11$	1 ½ ½	2
7	Let μ be the average mileage of that particular model automobile. Here, $H_0 : \mu = 23\ and\ H_1 : \mu \neq 23$ (or use $H_1 : \mu < 23$) Given that $\bar{x} = 21.8, s^2 = 7.84\ and\ n = 50$ Since n is large, the test statistic to be used is: $Z = \frac{(\bar{x} - \mu)}{s/\sqrt{n}} \sim N(0,1)$	½ ½	2

<table>
<tr>
<td></td>
<td>$$ie,\ Z=\frac{21.8-23}{\sqrt{7.84}/\sqrt{50}}=-3.03$$
For $\alpha=0.05$, the critical region is $|Z|\ge Z_{\alpha/2}$, $ie, |Z|\ge 1.96$ (or $|Z|>2.58$)
Here $|Z|=3.03>1.96$ (or $|Z|=3.03>2.58$)
$\therefore$ We reject H_0. The given data does not agree with the claim of the manufacturer.</td>
<td>½
½</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>We have the trend equation $y=18.04x+126.55$ with origin 2010. We have to shift the origin to 2015. The trend equation with shifted origin is:
$y=18.04(x+k)+126.55$, where k = 2015 – 2010 = 5
$ie,\ y=18.04(x+5)+126.55$
$or\ y=18.04x+216.75$</td>
<td>1
1</td>
<td>2</td>
</tr>
<tr>
<td>9.</td>
<td>Given n = 100, m = 10 and d = 170
$\bar{p}=\frac{d}{mn}=\frac{170}{10\times 100}=0.17$ and $\bar{q}=1-0.17=0.83$
The control limits of np chart are:
$CL=n\bar{p}=100\times 0.17=17$
$LCL=n\bar{p}-3\sqrt{n\bar{p}\bar{q}}=17-3\sqrt{14.11}=5.73$
$UCL=n\bar{p}+3\sqrt{n\bar{p}\bar{q}}=17+3\sqrt{14.11}=28.27$</td>
<td>½
½
½
½</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>$Mean=E(X)=\sum xp(x)=1\times 0.5+2\times 0.3+3\times 0.2=1.7$
$E(X^2)=\sum x^2p(x)=1^2\times 0.5+2^2\times 0.3+3^2\times 0.2=3.5$
$Variance=E(X^2)-[E(X)]^2$
$=3.5-1.7^2=0.61$</td>
<td>1
1
½
½</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>The population values are 13, 11, 15, 17 and 18.
The population mean = $\frac{13+11+15+17+18}{5}=\frac{74}{5}=14.8$
The possible number of SRSWORs of size 2 is $5C_2=10$.
<table>
<tr><th>Sample No.</th><th>Sample</th><th>Sample mean</th></tr>
<tr><td>1</td><td>13, 11</td><td>12</td></tr>
<tr><td>2</td><td>13, 15</td><td>14</td></tr>
<tr><td>3</td><td>13, 17</td><td>15</td></tr>
<tr><td>4</td><td>13, 18</td><td>15.5</td></tr>
<tr><td>5</td><td>11, 15</td><td>13</td></tr>
<tr><td>6</td><td>11, 17</td><td>14</td></tr>
<tr><td>7</td><td>11, 18</td><td>14.5</td></tr>
<tr><td>8</td><td>15, 17</td><td>16</td></tr>
<tr><td>9</td><td>15, 18</td><td>16.5</td></tr>
<tr><td>10</td><td>17, 18</td><td>17.5</td></tr>
<tr><td colspan="2">Total</td><td>148</td></tr>
</table>
Mean of sample means = $\frac{148}{10}=14.8$
Mean of sample means = population mean. $\therefore$ Sample mean is unbiased for population mean.</td>
<td>1
1
½
½</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>Let X denotes the weight of a particular kind of apple sold at a fruit market. Then X is normally distributed with $\mu=100$ and $\sigma=20$.</td>
<td></td>
<td></td>
</tr>
</table>

<table>
<tr>
<td></td>
<td>$\therefore Z = \frac{X-\mu}{\sigma} = \frac{X-100}{20}$ follows N(0, 1).
a) $P(70 < X < 110) = P\left(\frac{70-100}{20} < \frac{X-\mu}{\sigma} < \frac{110-100}{20}\right)$
$= P(-1.5 < Z < 0.5) = 0.6247$
b) $P(X > 110) = P(Z > 0.5) = 0.3085$
∴ No. of apples weigh greater than 110 gm = $1000 \times 0.3085 = 308$ (or 309)</td>
<td>½
½
½
1
½</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>
<table>
<tr><th>Commodity</th><th>p_0</th><th>q_0</th><th>p_1</th><th>q_1</th><th>p_0q_0</th><th>p_1q_0</th></tr>
<tr><td>A</td><td>12</td><td>100</td><td>15</td><td>120</td><td>1200</td><td>1500</td></tr>
<tr><td>B</td><td>6</td><td>210</td><td>7</td><td>240</td><td>1260</td><td>1470</td></tr>
<tr><td>C</td><td>10</td><td>110</td><td>13</td><td>150</td><td>1100</td><td>1430</td></tr>
<tr><td>Total</td><td></td><td></td><td></td><td></td><td>3560</td><td>4400</td></tr>
</table>
Laspyre's index number $= \frac{\sum p_1q_0}{\sum p_0q_0} \times 100$
$= \frac{4400}{3560} \times 100 = 123.6$</td>
<td>1½
1
½</td>
<td>3</td>
</tr>
<tr>
<td>14 a
14 b</td>
<td>(ii) Level of significance
Let X be the thickness of washers produced by the machine and let μ be the average thickness. We have to test $H_0 : \mu = 0.05$ against $H_1 : \mu \neq 0.05$.
Given $n = 10, \bar{x} = 0.053$ and $s = 0.003$.
The test statistic is $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$, follows $t_{(n-1)}$.
Here, $t = \frac{0.053-0.05}{0.003/\sqrt{9}} = 3$
For $\alpha = 0.01$ and degrees of freedom 9, the critical region is $|t| \geq 3.25$
Here |t| = 3 < 3.25. So we accept H_0 .The machine is working in proper order.</td>
<td>1
½
½
½
½</td>
<td>1
2</td>
</tr>
<tr>
<td>15</td>
<td>$\bar{R} = \frac{5+6+...+6}{10} = 5.8$, n = 5
The control limits for R – Chart are:
$CL = \bar{R} = 5.8$
$LCL = D_3\bar{R} = 0 \times 5.8 = 0$
$UCL = D_4\bar{R} = 2.115 \times 5.8 = 12.267$
All observed values are within the control limits. So the process is in control.</td>
<td>1
½
½
½
½</td>
<td>3</td>
</tr>
<tr>
<td>16 a
16 b</td>
<td>(iii) Seasonal variations
<table>
<tr><th>Month</th><th>No. of homes</th><th>3 yearly moving total</th><th>3 yearly moving average</th></tr>
<tr><td>June</td><td>6</td><td></td><td></td></tr>
<tr><td>July</td><td>7</td><td>22</td><td>7.33</td></tr>
<tr><td>August</td><td>9</td><td>24</td><td>8</td></tr>
<tr><td>September</td><td>8</td><td>26</td><td>8.67</td></tr>
<tr><td>October</td><td>9</td><td>27</td><td>9</td></tr>
<tr><td>November</td><td>10</td><td>31</td><td>10.33</td></tr>
<tr><td>December</td><td>12</td><td></td><td></td></tr>
</table>
</td>
<td>1
2</td>
<td>1
2</td>
</tr>
</table>

17 a	$y = x^3 + 7x^2 + 10x + 6$ $\frac{dy}{dx} = 3x^2 + 14x + 10$ $\frac{d^2y}{dx^2} = 6x + 14$	1 1	2
17 b	$\int_0^k x^2 dx = 9 \quad \Rightarrow \left[\frac{x^3}{3}\right]_0^k = 9$ ie, $\frac{k^3}{3} - 0 = 9 \Rightarrow k = 3$	1 1	2
18 a	Let X be the number of candidates selected out of n = 1000 candidates appeared. The probability of selection , $p = 0.2\% = 0.002$ is very small, we can use the Poisson distribution with $\lambda = np = 1000 \times 0.002 = 2$ We have the pmf, $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \ x = 0,1,2,...$ $ie, \ f(x) = \frac{e^{-2}2^x}{x!}, \ x = 0,1,2,...$ P(3 persons selected) = $P(X = 3)$ $= \frac{e^{-2}2^3}{3!} = \frac{0.1353 \times 8}{6} = 0.1804$ (Attempt with Binomial Distribution also consider)	½ ½ ½ ½	2
18 b	Let X be the number of workers suffering from the occupational decease. Then X follows Binomial distribution with n = 6 , p = 0.2 and q = 0.8. The pmf is: $f(x) = {}^nC_x p^x q^{n-x}, \ x = 0,1,...,n$ $ie, \ f(x) = {}^6C_x (0.2)^x (0.8)^{6-x}, \ x = 0,1,...,6$ P(4 will suffer from the decease) = P(X=4) $= {}^6C_4 \times (0.2)^4 \times (0.8)^2 = 15 \times 0.0016 \times 0.64 = 0.01536$	½ ½ ½ ½	2
19 a	(iv) 12	1	1
19 b	**Any 3 of the relations given below. Each carries 1 score.** 1. The square of a Standard Normal variable is a Chi – square variable with 1 degrees of freedom. 2. The sum of squares of n independent Standard Normal variables is Chi – square variable with n degrees of freedom. 3. If X is a Standard Normal variable and Y is a Chi – square variable with degrees of freedom, n. Then $t = \frac{X}{\sqrt{Y/n}}$ is a t variable with degrees of freedom n. 4. If X_1 is a Chi – square variable with degrees of freedom n_1 and X_2 is another independent Chi – square variable with degrees of freedom n_2. Then $F = \frac{X_1/n_1}{X_2/n_2}$ is a F variable with degrees of freedom (n_1, n_2) . 5. The square of a t variable with degrees of freedom n is a F variable with degrees of freedom (1, n).	3 X 1	3

20 a (iii) θ — 1 — 1

20 b

$$t_1 = \frac{x_1+x_2+x_3+x_4}{4} \text{ is unbiased.}$$

$$V(t_1) = V\left(\frac{x_1+x_2+x_3+x_4}{4}\right) = \frac{\sigma^2+\sigma^2+\sigma^2+\sigma^2}{16} = \frac{\sigma^2}{4} = 0.25\sigma^2$$ 1

$$t_2 = \frac{2x_1+x_2+x_3+x_4}{5} \text{ is unbiased.}$$ 3

$$V(t_2) = V\left(\frac{2x_1+x_2+x_3+x_4}{5}\right) = \frac{4\sigma^2+\sigma^2+\sigma^2+\sigma^2}{25} = \frac{7\sigma^2}{25} = 0.28\sigma^2$$ 1

Here, $V(t_1) < V(t_2)$. So t_1 is efficient than t_2 . 1

21 Let *x* denotes the rank of number of years of smoking and *y* denotes the rank of lung damage grade.

x	y	$d = x - y$	d^2
10	10	0	0
9	7	2	4
8	6	2	4
7	9	-2	4
6	5	1	1
5	8	-3	9
4	4	0	0
3	2	1	1
2	3	-1	1
1	1	0	0
		Total	$\sum d^2 = 24$

2 — 4

Rank correlation coefficient, $\rho = 1 - \frac{6\sum d^2}{n^3 - n}$ 1

$$= 1 - \frac{6 \times 24}{10^3 - 10} = 1 - \frac{144}{990} = 0.8545$$ 1

(or attempt using Karl Pearsons formula, give 2 scores)

22 a Correlation coefficient. — 1 — 1

22 b The two regression lines are:

$3x + 2y = 26$ ---------(1) and $6x + y = 31$----------(2)

(i) Let (1) denotes the regression line of Y on X and (2) denotes the regression line of X on Y.

Equation (1) becomes, $2y = -3x + 26$ or $y = \frac{-3}{2}x + 13 \quad \Rightarrow b_{yx} = \frac{-3}{2}$ ½

Equation (2) becomes, $6x = -y + 31$ or $x = \frac{-1}{6}y + \frac{31}{6} \quad \Rightarrow b_{xy} = \frac{-1}{6}$ ½

4

Now, $b_{yx} \times b_{xy} = \frac{-3}{2} \times \frac{-1}{6} = \frac{1}{4} < 1$. Hence our assumption is correct

$$\therefore r = \pm\sqrt{b_{yx} \times b_{xy}} = \pm\sqrt{\frac{1}{4}} = \frac{-1}{2} = -0.5$$ 1

(ii) Let $\bar{x}$ and $\bar{y}$ be the averages. The equations (1) and (2) become:

$3\bar{x} + 2\bar{y} = 26$ ----------(3) 1

$6\bar{x} + \bar{y} = 31$ ----------(4)

$(3) - 2\times(4) \Rightarrow -9\bar{x} = -36 \Rightarrow \bar{x} = 4$, average price = 4 ½

$(4) - 2\times(3) \Rightarrow -3\bar{y} = -21 \Rightarrow \bar{y} = 7$, average demand =7 ½

<table>
<tr><td>23 a</td><td>(iv) Equality of more than two means</td><td>1</td><td>1</td></tr>
<tr><td>23 b</td><td>

H_0 : The yields are equal and H_1 : The yields are not equal.

<table>
<tr><th>A</th><th>B</th><th>C</th></tr>
<tr><td>20</td><td>18</td><td>25</td></tr>
<tr><td>21</td><td>20</td><td>28</td></tr>
<tr><td>23</td><td>17</td><td>22</td></tr>
<tr><td>$T_1 = 64$</td><td>$T_2 = 55$</td><td>$T_3 = 75$</td></tr>
</table>

Grand Total, $G = 64 + 55 + 75 = 194$

Correction factor, $CF = \frac{G^2}{n} = \frac{194^2}{9} = 4181.78$

Total Sum of Squares, TSS = Sum of squares of all observations – CF

$$= 20^2 + 21^2 + ... + 22^2 - 4181.78 = 94.22$$

Between Sum of Squares, SSB = $\sum \frac{T_i^2}{n_i} - CF$

$$= \frac{64^2}{3} + \frac{55^2}{3} + \frac{75^2}{3} - 4181.78 = 66.89$$

ANOVA Table

<table>
<tr><th>Source</th><th>df</th><th>SS</th><th>MSS</th><th>F</th><th>$F_{0.5}$</th></tr>
<tr><td>Between
Within</td><td>2
6</td><td>66.89
27.33</td><td>33.445
4.555</td><td>7.34</td><td>5.14</td></tr>
<tr><td>Total</td><td>8</td><td>94.22</td><td></td><td></td><td></td></tr>
</table>

Here $F > F_{0.5}$. So we reject H_0 . That is the yields are not equal for the three varieties of seeds.

</td><td>½
½
½
½
1 ½
½</td><td>4</td></tr>
<tr><td>24</td><td>

The hypotheses are:

H_0 : Drinking habit and the favour in local sale of liquor are independent.

H_1 : Drinking habit and the favour in local sale of liquor are not independent.

We are given a 2X2 contingency table. The contingency table along with the expected frequencies is:

<table>
<tr><th rowspan="2">Question 1</th><th colspan="2">Question 2</th><th rowspan="2">Total</th></tr>
<tr><th>Yes</th><th>No</th></tr>
<tr><td>Yes</td><td>56 (58)</td><td>31 (29)</td><td>87</td></tr>
<tr><td>No</td><td>18 (16)</td><td>6 (8)</td><td>24</td></tr>
<tr><td>Total</td><td>74</td><td>37</td><td>111</td></tr>
</table>

The test statistics is $\chi^2 = \sum \frac{(O-E)^2}{E}$ follows Chi – square distribution with degrees of freedom $(R-1)(C-1) = (2-1)(2-1) = 1$.

The value of test statistic is:

$$\chi^2 = \frac{(56-58)^2}{58} + \frac{(31-29)^2}{29} + \frac{(18-16)^2}{16} + \frac{(6-8)^2}{8} = 0.96$$

The critical region with df=1 is, $\chi^2 > 3.84$.

Here $\chi^2 = 0.96 < 3.84$. So we accept H_0. Hence the drinking habit and the favour in local sale of liquor are independent.

</td><td>1
1 ½
1
1
½</td><td>5</td></tr>
</table>