MA212 - LINEAR ALGEBRA AND STATISTICAL ANALYSIS

TUTORIAL - III

U19CS012

1.) Find the least square straight line for the following data:
(X, Y) : (1,6), (2,4), (3,3), (4,5), (5,4) & (6,2).
Estimate Y at X = 4 & X at Y = 4. Also find standard error estimate (se.)

1.) Step 1: for each (x,y) point calculate x^2 & xy

x	y	x^2	xy
1	6	1	6
2	4	4	8
3	3	9	9
4	5	16	20
5	4	25	20
6	2	36	12
$\Sigma x = 21$	$\Sigma y = 24$	$\Sigma x^2 = 91$	$\Sigma xy = 75$

Σ

Step 2: Sum all x, y, x^2 and xy to get $\Sigma x, \Sigma y, \Sigma x^2$ and Σxy

Step 3: Calculate Slope m:

$$m = \frac{N \Sigma(xy) - \Sigma x \Sigma y}{N \Sigma(x^2) - (\Sigma x)^2}$$

N = 6 = No. of points

$$= \frac{(6 \times 75) - (21 \times 24)}{6 \times 91 - (21)^2}$$

$$= \frac{-54}{105} = \boxed{-0.51428}$$

Step 4: Calculate intercept b:

$$b = \frac{\Sigma y - m \Sigma x}{N} = \frac{24 - (-0.51428)\,21}{6} = \boxed{5.79998}$$

ANS: Step 5: $y = mx + b$: $\boxed{y = (-0.51428)x + 5.79998}$ Least square straight line

U19CS012

$y = (-0.51428)x + 5.79998$

(A) Estimate $y = ?$ at $x = 4$

$y = (-0.51428)(4) + 5.79998$

$\boxed{y = 3.74286}$

(B) Estimate $x = ?$ at $y = 4$

$x = \frac{y - 5.79998}{(-0.51428)} = \frac{4 - 5.79998}{-0.51428} = \boxed{3.5}$

X	Y	$\bar{Y} = -0.5142X + 5.799$	$(error)^2$
1	6	5.2857	0.510224
2	4	4.77142	0.595089
3	3	4.25714	1.5804 ~~0.595089~~
4	5	3.74286	1.5804
5	4	3.22858	0.595089
6	2	2.7143	0.510224

$\Sigma e^2 = 5.37143$

(C) Standard error estimate (Se) $= \sqrt{\frac{\Sigma(y-\bar{y})^2}{N}} = \sqrt{\frac{5.37143}{6}} = \boxed{0.9461}$

2.> Estimate the blood pressure of women of age 45 from the following data which shows the age X and B.P 'Y' of 12 women. Are the two variables X and B.P 'Y' correlated? Find the correlation coefficient r.

Age (X)	56	42	72	36	63	47	55	49	38	42	68	60
Blood Pressure (Y)	147	125	160	118	149	128	150	145	115	140	152	155

U19CS012

(A) Correlation?

If we plot the points on cartesian plane, it will have (+ve) (positive) correlation as we can see if person's age increases, its blood pressure also increases.

(B)

Correlation Coefficient (r) = $\dfrac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2}\sqrt{\Sigma (y-\bar{y})^2}}$

X	Y	$X-\bar{X}$	$(X-\bar{X})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
56	147	3.67	13.46	6.67	44.48	24.47
42	125	-10.33	106.70	-15.33	235.00	158.35
72	160	19.67	386.90	19.67	386.90	386.90
36	118	-16.33	266.66	-22.33	498.62	364.64
63	149	10.67	113.84	8.67	75.16	92.50
47	128	-5.33	28.40	-12.33	152.02	65.71
55	150	2.67	7.12	9.67	93.50	25.81
49	145	-3.33	11.08	4.67	21.80	-15.55
38	115	-14.33	205.34	-25.33	641.60	362.97
42	140	-10.33	106.70	-0.33	0.10	3.40
68	152	15.67	245.54	11.67	136.18	182.86
60	155	7.67	58.82	14.67	215.20	112.51
$\bar{X}$ = 52.33	$\bar{Y}$ = 140.33		1550.56		2500.56	Σ = 1764.57

$\bar{X} = \dfrac{\Sigma x}{n} = \dfrac{628}{12} = 52.33$ $\quad \bar{Y} = \dfrac{\Sigma y}{n} = \dfrac{1684}{12} = 140.33$

ANS: $r = \dfrac{1764.57}{\sqrt{1550.56}\sqrt{2500.56}} = \dfrac{1764.57}{39.37 \times 50.00} = \boxed{0.8964}$

VISION

U19CS012

3) The pH solution is measured eight times using the same instrument and data obtained as in table. Find Mean, SD and Variance

X	$X-\bar{X}$	$(X-\bar{X})^2$
7.15	-0.03	0.0009
7.20	0.02	0.0004
7.18	0	0
7.19	0.01	0.0001
7.21	0.03	0.0009
7.20	0.02	0.0004
7.16	-0.02	0.0004
7.18	0	0
$\Sigma x = 57.47$		$\Sigma = 0.0031$

(i) Mean $= \frac{\Sigma x}{n} = \frac{57.47}{8} = \boxed{7.18375} \approx 7.18$

(ii) SD $= \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}} = \sqrt{\frac{0.0031}{8}}$

$\boxed{= 0.01968}$

(iii) Variance $= (SD)^2$

$\boxed{= 0.00038375}$

4) Let X be the height of a randomly chosen individual from population. In order to estimate the mean and variance of X, we observe a random sample $x_1, x_2, x_3 \ldots x_7$. Find Mean, SD and Variance

X	$X-\bar{X}$	$(X-\bar{X})^2$
166.8	-2	4
171.4	2.6	6.76
169.1	0.3	0.09
178.5	9.7	94.09
168	-0.8	0.64
157.9	-10.9	118.81
170.1	1.3	1.69
$\Sigma x = 1181.8$		$\Sigma = 226.08$

(i) Mean $= \bar{X} = \frac{1181.8}{7} = \boxed{168.8}$

(ii) Variance $= S^2 = \frac{\Sigma(x-\bar{x})^2}{(n-1)} = \frac{226.08}{7-1} = \boxed{37.7}$

(iii) Standard Deviation $= S = \sqrt{V} = \sqrt{37.7}$

$= \boxed{6.14}$

VISION

U19CS012

5.) Let $x_1, x_2, \dots x_n$ be a random sample from a geometric $(\hat{\theta})$ distribution, $\hat{\theta}$ is unknown. Find the maximum likelihood $\hat{\theta}$ based on random sample.

5.) if $x_i \sim$ Geometric (θ), then

$$P_{x_i}(x_i, \theta) = (1-\theta)^{x-1}\, \theta$$

Thus likelihood function is given by,

$$L(x_1, x_2, x_3, \dots x_n; \theta) = P_{x_1, x_2, \dots x_n}(x_1, x_2, \dots x_n; \theta)$$

$$= P_{x_1}(x_1; \theta)\, P_{x_2}(x_2; \theta) \dots P_{x_n}(x_n; \theta)$$

$$= (1-\theta)\left[\sum_{i=1}^{n}(x_i - n)\right] \theta^n$$

Then the log likelihood function is given by

$$\ln L(x_1, x_2, \dots x_n; \theta) = \left(\sum_{i=1}^{n}(x_i - n)\right) \ln(1-\theta) + n \ln(\theta)$$

$$\frac{d \ln L(x_1, x_2, \dots x_n; \theta)}{d\theta} = \left(\sum_{i=1}^{n} x_i - n\right) \cdot \frac{(-1)}{1-\theta} + \frac{n}{\theta}$$

By setting the derivative to zero, we can check that the maximum value of 'θ' is given by

$$\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}$$

$\therefore$ MLE can be written as $\boxed{\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}}$

U19CS012

6.) Let $x_1, x_2, \ldots x_n$ be a random sample from a uniform $(0, \hat{\theta})$ distribution, where $\hat{\theta}$ is unknown. Find the maximum likelihood of $\hat{\theta}$ based on this random sample.

6.) If $X_i \sim$ Uniform $(0, \theta)$ then

$$f_X(x) = \begin{cases} 1/\theta, & 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$$

The Likelihood function is given by

$$L(x_1, x_2, \ldots x_n; \theta) = f_{X_1 X_2 \cdots X_n}(x_1, x_2, x_3, \ldots x_n; \theta)$$

$$= f_{X_1}(x_1; \theta)\, f_{X_2}(x_2; \theta)\, f_{X_3}(x_3; \theta) \cdots f_{X_n}(x_n; \theta)$$

$$= \begin{cases} 1/\theta^n & 0 \le x_1, x_2, \ldots x_n \le \theta \\ 0 & \text{Otherwise} \end{cases}$$

Note that $1/\theta^n$ is a decreasing function of θ.

Thus to maximise it, we need to chose the smallest possible value for θ for $i = 1, 2, \ldots n$ we need to have $\theta \ge x_i$. Thus the smallest possible value for θ is

$$\hat{\theta}_{ML} = \max(X_1, X_2, \ldots X_n)$$

Therefore, MLE can be written as

$$\boxed{\hat{\theta}_{ML} = \max(X_1, X_2, \ldots, X_n)}$$

VISION

U19CS012

7.)

Sample 1

$n_1 = 40$

$\bar{x}_1 = 647$ hrs

$S_1 = 31$ hrs

Sample 2

$n_2 = 45$

$\bar{x}_2 = 742$ hrs

$S_2 = 29$ hrs

i) 95% confidence level

$\bar{x}_1 - \bar{x}_2 = 647 - 742 = -95$ $\quad \alpha = 1 - \frac{95}{100} = 0.05$

$\alpha/2 = 0.025$ $\quad \left[Z_{\alpha/2} = 1.96 \right]$

Interval : $(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

$= -95 \pm 1.96 \sqrt{\frac{(31)^2}{40} + \frac{(29)^2}{45}}$

$= -95 \pm 1.96 \sqrt{24.025 + 18.688}$

$= -95 \pm 1.96 \times 6.535$

$= -95 \pm 12.8086$

Interval $\Rightarrow$ $\boxed{(-107.8086, -82.1914)}$

ii) 99% confidence level

$\bar{x}_1 - \bar{x}_2 = 647 - 742 = -95$ $\quad \alpha = 1 - \frac{99}{100} = 0.01$

$\alpha/2 = 0.005$ $\quad Z_{\alpha/2} = 2.57$

Interval $= (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}$

$= -95 \pm 2.57 \sqrt{\frac{(31)^2}{40} + \frac{(29)^2}{45}}$

$= -95 \pm 2.57 \sqrt{(24.075 + 18.688)}$

$= -95 \pm 2.57 \times 6.535$

$= -95 \pm 16.794$

Interval $\Rightarrow$ $\boxed{(-111.794, -78.205)}$

U19CS012

8.) We need to find a 95% (= 100 (1 - α) %) confidence interval for μ_D.

$95\% = 100(1-\alpha)\%$

$\alpha = 1 - \frac{95}{100} = 0.05$ $\boxed{\frac{\alpha}{2} = 0.025}$

Acc to theorem,

If $\bar{d}$ (mean) and s_d (standard dev) for normally distributed differences of n random pairs of measurement) $100(1-\alpha)\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\boxed{\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}}$$

From data,

$d_1 = 38 - 45 = -7$

$d_2 = 23 - 25 = -2$

$d_3 = 35 - 31 = 4$

$d_4 = 41 - 38 = 3$

$d_5 = 44 - 50 = -6$

$d_6 = 29 - 33 = -4$

$d_7 = 37 - 36 = 1$

$d_8 = 31 - 40 = -9$

$d_9 = 38 - 43 = -5$

Hence, $\bar{d} = \frac{1}{n}\Sigma d_i$

$= \frac{1}{9}(-7 + (-2) + 4 + 3 + (-6) + (-4) + 1 + (-9) + (-5))$

$= \left(\frac{-25}{9}\right) = \boxed{-2.7778}$

The standard deviation is s_d is

$$s_d = \sqrt{\frac{1}{(n-1)}\Sigma(d_i - \bar{d})^2}$$

$$= \sqrt{\frac{1}{(9-1)}((-7 + 2.7778)^2 + (-2 + 2.7778)^2 + \dots (-5 + 2.7778)^2}$$

$= \sqrt{20.94444}$

$= \boxed{4.5765}$

To find $t_{0.025}$, using Table of critical values of t-distribution, t value which is leaving an area of 0.025 to right with $v = n - 1 = 9 - 1 = 8$ degrees of freedom,

$$\boxed{t_{0.025} = 2.306}$$

$$\bar{d} - t_{\alpha/2} \frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

$$-2.7778 - 2.306 \left(\frac{4.5765}{\sqrt{9}}\right) < \mu_D < -2.7778 + 2.306 \left(\frac{4.5765}{\sqrt{9}}\right)$$

$$-2.7778 - 3.5178 < \mu_D < -2.7778 + 3.5178$$

$$\boxed{-6.29 < \mu_D < 0.74}$$

So, the required 95% confidence interval is $\boxed{-6.29 < \mu_D < 0.74}$.

9.) A random sample of 10 chocolate energy bars of a certain brand has an average 230 calories with standard deviation of 15 calories, then construct 99% confidence interval for σ^2.

9) $n_1 = 10$, $\bar{x} = 230$ calories, SD = 15 calories

99% confidence interval for σ^2 (variance) = (?)

Given data: $\boxed{\chi^2_{\alpha/2} = 21.666}$ $\boxed{\chi^2_{1-\alpha/2} = 2.088}$ at 9 d.o.f

$\alpha = 1 - 99/100 = 0.01$ $\quad \alpha/2 = 0.005$

Interval:
$$\frac{(n-1)s^2}{\chi^2_{(\alpha/2)}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

$$\frac{(10-1)(225)}{21.666} \leq \sigma^2 \leq \frac{(10-1)(225)}{2.088}$$

ANSWER: $\boxed{93.464 \leq \sigma^2 \leq 969.33}$

x

SUBMITTED BY:

U19CS012

BHAGYA VINOD RANA

CSE