FALAK JAIN
2274350452

1. D
2. A
3. C
4. C
5. B
6. F
7. B
8. E
9. A
10. C
11. G
12. A
13. F
14. E
15. A
16. B
17. A
18. A
19. B
20. D
21. B
22. B
23. B
24. A
25. B

FALAK JAIN
22743504S2

26. Sign of Spearman's correlation : positive
Sign of Pearson's correlation : positive

27. $\log y_{new} = 3 + 90 \log x_{new}$ —— ①
$\log y_{old} = 3 + 90 \log x_{old}$ —— ②

Subtracting ② from ① :

$$\log\left(\frac{y_{new}}{y_{old}}\right) = 90 \log\left(\frac{x_{new}}{x_{old}}\right) \longrightarrow \{\text{for a 1\% change in } x : x_{new} = 1.01\, x_{old}$$

$$\Rightarrow \log\left(\frac{y_{new}}{y_{old}}\right) = 90 \times 0.01 = 0.9$$

$$\Rightarrow \frac{y_{new}}{y_{old}} = e^{0.9} = 2.46.$$

$$\Rightarrow y_{new} = 2.46\, y_{old}$$

Therefore, a 1% increase in x leads to a 146% increase in y.

28. (i) Backward Stepwise Selection : $1 + \frac{p(p+1)}{2} = 1 + \frac{20(21)}{2} = 211.$

(ii) Best subset selection : $2^p = 2^{20} = 1,048,576$

29. $AIC = \frac{1}{n\hat{\sigma}^2}(RSS + 2d\hat{\sigma}^2)$, $C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2)$

As we can see above, AIC has an additional term $\hat{\sigma}^2$ in the denominator which is a constant. The presence of a constant term will still lead to the same rankings whether we use AIC or C_p. Therefore, both metrics return the same models.

FALAK JAIN
2274350452

30. Cluster 1 : x_1 & x_3

$(0,0)$ & $(100,100)$

$\therefore$ centroid $= \frac{0+100}{2}, \frac{0+100}{2}$

$= (50,50)$ ①

Cluster 2 : x_2, x_4, x_5

$(1,1), (101,101), (123,123)$

$\therefore$ centroid $= \left(\frac{1+101+123}{3}\right), \left(\frac{1+101+123}{3}\right)$

$= (75,75)$ ②

31. Gini Index :

$$= \frac{5}{30}\left(1-\frac{5}{30}\right) + \frac{5}{30}\left(1-\frac{5}{30}\right) + \frac{10}{30}\left(1-\frac{10}{30}\right) + \frac{10}{30}\left(1-\frac{10}{30}\right)$$

$$= \frac{5}{30}\left(\frac{25}{30}\right) + \frac{5}{30}\left(\frac{25}{30}\right) + \frac{10}{30}\left(\frac{20}{30}\right) + \frac{10}{30}\left(\frac{20}{30}\right)$$

$$= \frac{125+125+200+200}{900} = \frac{13}{18} = 0.722$$