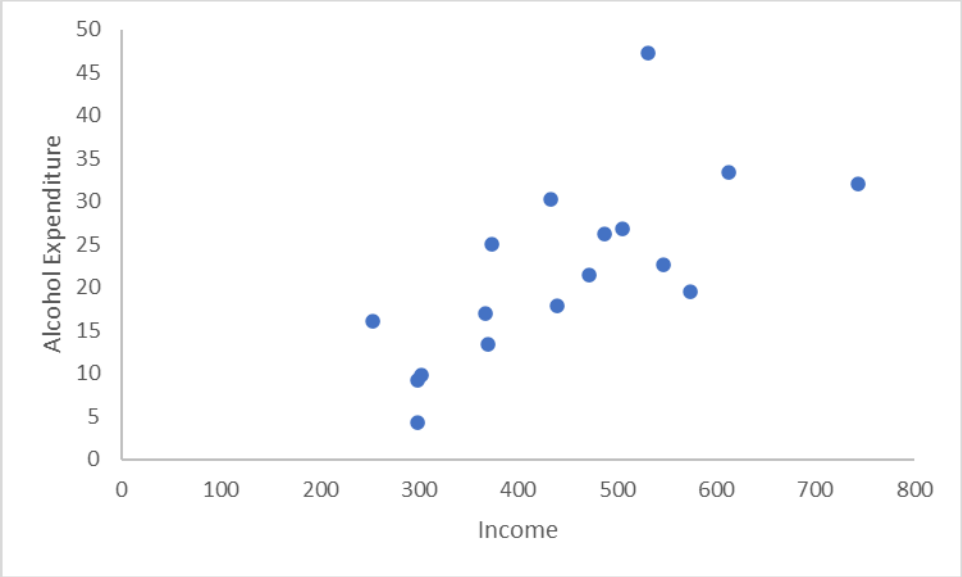
Solution:

(A)

The scatter plot can be drawn as –



From the above scatter plot, we can see an increasing trend which indicates that there is a positive relationship between Alcohol expenditure and Income.

(B)

Since there is an increasing trend, so, we shall expect the line of best fit to **slope up**.

(C)

$$Correlation\ coefficient = \frac{\sum XY - \frac{(\sum X) * (\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

$$= \frac{181911.25 - \frac{7610 * 371.99}{17}}{\sqrt{\left(3680748 - \frac{(7610)^2}{17}\right) * \left(9918.4547 - \frac{(371.99)^2}{17}\right)}} = 0.6969861$$

$$\boldsymbol{Correlation\ coefficient = 0.6970}$$

(D)

To test the significance of the correlation, the null and alternative hypotheses can be defined as –

$$H_0: \rho = 0, \qquad H_A: \rho \neq 0$$

Test statistic can be calculated as –

$$Test\ statistic = t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.6969861 * \sqrt{17-2}}{\sqrt{1-(0.6969861)^2}} = 3.764431$$

$$Test\ statistic = t = 3.764431$$

$$df = n - 2 = 17 - 2 = 15$$

$$p - value = 0.0019$$

$$\alpha = 0.05$$

Since the p-value is less than the level of significance, so we can reject the null hypothesis.

Thus, we can conclude that the correlation coefficient is significant.

This imply that there is a significant linear relationship between Alcohol expenditure and Income.

(E)

Running multiple linear regression in excel, we get the following output –

SUMMARY

OUTPUT

Regression Statistics	
Multiple R	0.8620
R Square	0.7430
Adjusted R Square	0.6837
Standard Error	5.9297
Observations	17

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1321.5606	440.520213	12.528567	0.000391213
Residual	13	457.09641	35.1612621		
Total	16	1778.657			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-8.3122	5.5259	-1.5042	0.1564	-20.2501	3.6257
Income	0.0323	0.0132	2.4509	0.0292	0.0038	0.0608
Adult	7.7909	2.2093	3.5265	0.0037	3.0181	12.5637
Smoker	2.5940	2.9242	0.8871	0.3912	-3.7234	8.9114

From the above output, the regression equation can be written as –

$$\widehat{Alcohol\ Expenditure} = -8.3122 + 0.0323 * Income + 7.7909 * Adult + 2.5940 * Smoker$$

All the coefficients have expected signs because as income increases, expenditure on Alcohol also increases, which is logical.

If there are a greater number of adults in the family, the expenditure on Alcohol likely to increase, so coefficient of adult being positive also makes sense.

Smokers are likely to drink more alcohol than non-smoker, so positive sense for smoker also makes sense.

(F)

$$p - value\ associated\ with\ Income = 0.0292$$

$$\alpha = 0.05$$

Since the p-value is less than the level of significance, so, we can conclude that the Income variable is significant.

$$p - value\ associated\ with\ Adult = 0.0037$$

$$\alpha = 0.05$$

Since the p-value is less than the level of significance, so, we can conclude that the variable "Adult" is significant.

$$p - value\ associated\ with\ Smoker = 0.3912$$

$$\alpha = 0.05$$

Since the p-value is greater than the level of significance, so, we can conclude that the variable "Smoker" is not significant.

From the ANOVA table –

$$P - value\ associated\ with\ the\ F\ test\ statistic = 0.0004$$

$$\alpha = 0.05$$

Since the p-value is greater than the level of significance, so, we can conclude that the overall model is significant.

(G)

Since the variable "Smoker" is not significant