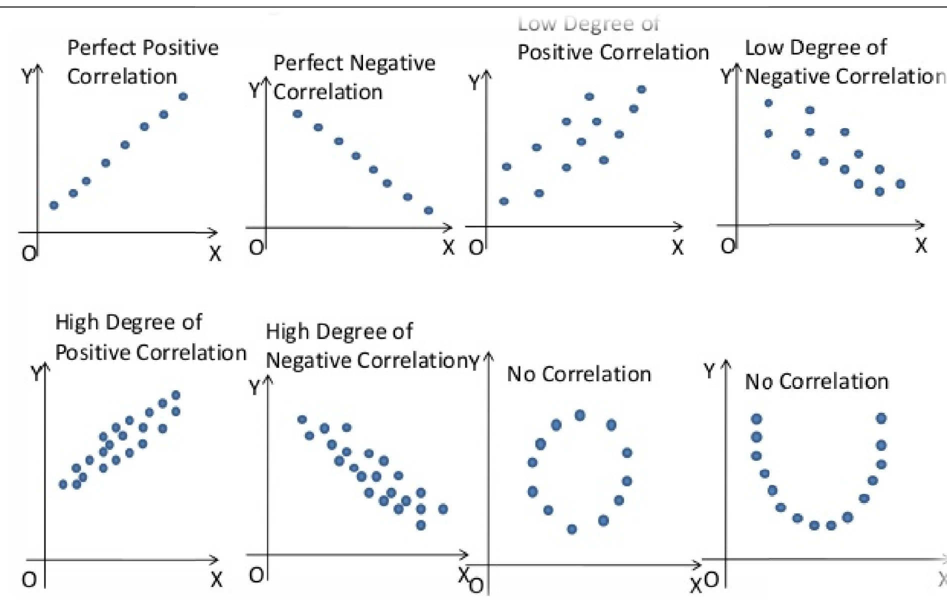
CORRELATION

<table>
<tr><td>Bi-Variate Data</td><td>When data are collected on two discrete variables simultaneously, they are known as Bi-Variate data</td></tr>
<tr><td>Bi-Variate Distribution</td><td>Distribution of Bi-Variate data is called as Bivariate Distribution</td></tr>
<tr><td>Bi-Variate Frequency Distribution</td><td>
<table>
<tr><td>Meaning</td><td>Frequency distribution involving two discrete variables.</td></tr>
<tr><td>Marginal Distribution</td><td>If we make a separate distribution from bi-variate frequency distribution where we take aggregate of only one variable at a time. Total no. of marginal distributions = 2</td></tr>
<tr><td>Conditional Distribution</td><td>If we make a separate distribution from bi-variate frequency distribution where we take one variable related one class interval of another variable. Total no. of conditional distributions = m + n (m = no. of rows, n = no. of columns)</td></tr>
</table>
</td></tr>
<tr><td>Correlation</td><td>While studying two variables at the same time, if it is found that the change in one variable leads to change in the other variable either directly or inversely, then the two variables are known to be associated or correlated.
<table>
<tr><td>Positive Correlation</td><td>If two variables move in the same direction</td></tr>
<tr><td>Negative Correlation</td><td>If two variables move in the opposite direction</td></tr>
<tr><td>No Correlation</td><td>If no change due to each other</td></tr>
</table>
</td></tr>
<tr><td>Measure of Correlation</td><td>A measurement or formula that represents the nature/ direction and/or magnitude of correlation.
<table>
<tr><th>Method</th><th>Helps in obtaining</th></tr>
<tr><td>Scatter Diagram</td><td>Only direction of correlation</td></tr>
<tr><td>Karl Pearson's Product moment correlation coefficient</td><td>Direction as well as strength of correlation. Best Method – Most accurate</td></tr>
<tr><td>Spearman's rank correlation co-efficient</td><td>Direction as well as strength of correlation. Useful for attributes.</td></tr>
<tr><td>Co-efficient of concurrent deviations</td><td>Direction as well as strength of correlation. Only preferred for direction and not magnitude. Quickest method.</td></tr>
</table>
</td></tr>
<tr><td>Scatter Diagram</td><td>

</td></tr>
</table>

Learn with CA. Pranav
Transforming students to professionals

<table>
<tr><td rowspan="5">Karl Pearson's
Product moment
correlation coefficient</td><td>Defined as</td><td>the ratio of covariance between the two variables to the product of the standard deviations of the two variables</td></tr>
<tr><td>Main Formula</td><td>$r_{xy} = \frac{Cov(x,y)}{\sigma_x.\sigma_y}$</td></tr>
<tr><td>Formula for Covariance</td><td>$Cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$ or $\frac{\Sigma xy}{n} - \bar{x}.\bar{y}$</td></tr>
<tr><td>Formula for Standard Deviation σ_x or σ_y</td><td>$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{\text{n}}}$ or $\sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2}$</td></tr>
<tr><td>Properties</td><td>→ It is a unit-free measurement
→ Value of r lies from -1 to +1 both inclusive
→ Change of origin or Scale
<table><tr><td>Change of Origin</td><td>No impact</td></tr><tr><td>Change of Scale</td><td>No impact of value, but if change of scale of both variables are of different sign then sign r will also change</td></tr></table></td></tr>
<tr><td rowspan="4">Spearman's
Rank
Correlation coefficient</td><td>Applied to</td><td>find the level of agreement (or disagreement) between two judges so far as assessing a qualitative characteristic is concerned</td></tr>
<tr><td>Main Formula</td><td>$r_R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$, here d means difference in ranks</td></tr>
<tr><td>Adjustment Value in case of Tie Rank</td><td>$\Sigma \frac{(t^3-t)}{12}$ here t is a tie length and we need to do summation of all ties</td></tr>
<tr><td>Formula in case of Tie length</td><td>$r_R = 1 - \frac{6(\Sigma d^2 + \text{value of adjustment})}{n(n^2-1)}$</td></tr>
<tr><td rowspan="3">Co-efficient of concurrent deviations</td><td>Use</td><td>A very simple and casual method of finding correlation when we are not serious about the magnitude of the two variables</td></tr>
<tr><td>Steps in this method</td><td>This method involves in attaching a positive sign for a x-value (except the first) if this value is more than the previous value and assigning a negative value if this value is less than the previous value. Applies to both variable and then these signs are compared. If signs match – pair is counted as concurrent deviation.</td></tr>
<tr><td>Formula</td><td>$r_c = \pm\sqrt{\pm\frac{2c-m}{m}}$
Here, m = total no. of deviations (it is one less than total no. of pairs under observation i.e m=n-1), c = no. of concurrent deviations, r_c also lies between -1 and 1 incl.</td></tr>
</table>

Learn with CA. Pranav
Transforming students to Professionals

REGRESSION

Regression Analysis	Estimation of one variable for a given value of another variable on the basis of an average mathematical relationship between the two variables	
Estimation of Y (when it is dependent on X)	Line	Regression line of Y on X
	Regression Coefficient	Regression Coefficient of Y on X denoted by $\boldsymbol{b_{yx}}$
	Form	$\boldsymbol{Y - \bar{Y} = b_{yx}\,(X - \bar{X})}$, $\bar{X}$ *and* $\bar{Y}$ *are means of X series and Y series*
Estimation of X (when it is dependent on Y)	Line	Regression line of X on Y
	Regression Coefficient	Regression Coefficient of X on Y denoted by $\boldsymbol{b_{xy}}$
	Form	$\boldsymbol{X - \bar{X} = b_{xy}\,(Y - \bar{Y})}$, $\bar{X}$ *and* $\bar{Y}$ *are means of X series and Y series*
Important Theory Points	When linear relationship exists between two variables (i.e. correlation is perfect, $r_{xy} = -1\ or + 1$)	The linear equation so arrived can be used both ways for Y on X and X on Y. It means regression lines are identical.
	When no linear relationship exist between two variables	In that case, we need to estimate the regression lines with the help of Method of Least Squares
	To derive regression line of y on x	The minimisation of vertical distances in the scatter diagram is to be done
	To derive regression line of x on y	The minimisation of horizontal distances in the scatter diagram is to be done
Regression Coefficient	Defined as the ratio of	$\frac{\text{Covariance between two variables}}{\text{Variance of Independent variable}}$
	Regression Coefficient of Y on X	$b_{yx} = r.\frac{\sigma_y}{\sigma_x}$ or $b_{yx} = \frac{Cov(x,y)}{{\sigma_x}^2}$
	Regression Coefficient of X on Y	$b_{xy} = r.\frac{\sigma_x}{\sigma_y}$ or $b_{xy} = \frac{Cov(x,y)}{{\sigma_y}^2}$
	$\boldsymbol{r}$ used here is Karl Pearson's Correlation Coefficient	
Properties of Regression lines and coefficient	Change of origin	The regression coefficients remain unchanged
	Change of scale	: If original pair is X, Y and modified pair is U, V where U = $\frac{X-m}{p}$ and V = $\frac{Y-n}{q}$, then $b_{vu} = b_{yx}\frac{q}{p},\ b_{uv} = b_{xy}\ \frac{p}{q}$
	Intersection of two regression lines	Two regression (if not identical) will intersect at the point $(\bar{x}, \bar{y})$ [means]
	Relation between correlation and regression coefficients	$r = \pm\sqrt{\pm b_{xy} \times b_{yx}}$ b_{xy}, b_{yx} *and* r all will have same sign

Learn with CA. Pranav
Transforming students to Professionals

<table>
<tr><td rowspan="3">Coefficient of Determination</td><td>Coefficient of Determination</td><td>r^2 (square of correlation coefficient)</td></tr>
<tr><td>Interpretation of value of r^2</td><td>It explains the percentage of variation in dependent variable due to variation in independent variable</td></tr>
<tr><td>Example: if $r_{xy} = 0.8, then\ r^2 = 0.64$</td><td>It means 64% of variation in $\boldsymbol{x}$ is due to variation in $\mathbf{y}$ and remaining 36% due to other factors. It shows the reliability of correlation coefficient.</td></tr>
<tr><td rowspan="4">Probable Error</td><td>Formula</td><td>Probable Error [P.E] = $^2/_3$ × Standard Error [S.E.]</td></tr>
<tr><td>Standard Error</td><td>$\frac{1 - r^2}{\sqrt{n}}$</td></tr>
<tr><td>Use</td><td>Probable Error is used to test the reliability of $\boldsymbol{r}$</td></tr>
<tr><td>Test</td><td><table><tr><td>If r is less than PE</td><td>The value of r is not significant. Not reliable</td></tr><tr><td>If r is greater than six times of PE</td><td>The value of r is significant and there is evidence of correlation</td></tr></table></td></tr>
</table>

Learn with CA. Pranav
Transforming students to Professionals