Advance Statistics

1)Calculate covariance and correlation between below two columns A and B

Ans: Calculation steps for Co-variance

Step 1: Calculate means for two columns

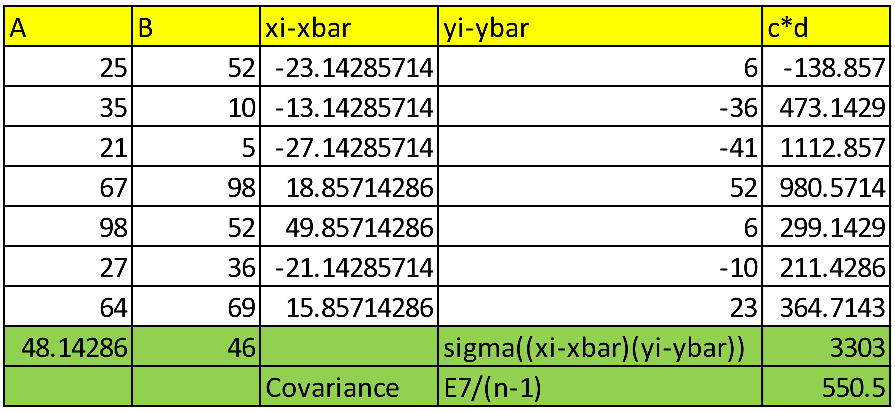
Step 2: Construct xi-xbar and yi-ybar

Step 3: Multiply both Xi-xbar and Yi-ybar columns

Step 4: Now add all the observations of step 3

Step 5: Divide answer of step 4 with (n-1) ‘6’

The above are the covariance steps The answer will be in step 6



A	B	xi-xbar	yi-ybar	c*d
25	52	-23.14285714	6	-138.857
35	10	-13.14285714	-36	473.1429
21	5	-27.14285714	-41	1112.857
67	98	18.85714286	52	980.5714
98	52	49.85714286	6	299.1429
27	36	-21.14285714	-10	211.4286
64	69	15.85714286	23	364.7143
48.14286	46		sigma((xi-xbar)(yi-ybar))	3303
		Covariance	E7/(n-1)	550.5

Calculation steps for Correlation

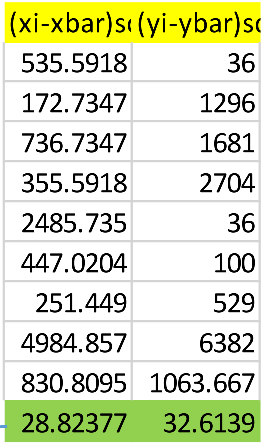
Step1: construct (xi-xbar) ^2 and (yi-ybar) ^2

Step2: Add the whole columns and divide it with n-1 ie ‘6’

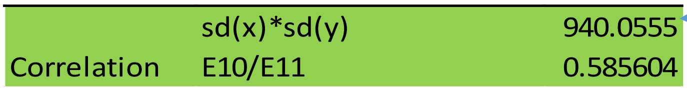
Step3: Now apply square root to the answer that arise in step 2

Step4: Multiply standard deviation of both columns

Step5: Divide the result of covariance with answer of step4



(xi-xbar)s	(yi-ybar)s
535.5918	36
172.7347	1296
736.7347	1681
355.5918	2704
2485.735	36
447.0204	100
251.449	529
4984.857	6382
830.8095	1063.667
28.82377	32.6139



	sd(x)*sd(y)	940.0555
Correlation	E10/E11	0.585604

2) What are the different ways to deal with multi collinearity?

Ans: There are two steps to deal with collinearity

1)If your main focus or objective is on y predictor then there is no problem in collinearity

2) Get rid of the redundant variables using a variable selection technique.

3)nearly combine the independent variables, such as adding them together.

4)Perform an analysis designed for highly correlated variables, such as principal components analysis or partial least squares regression.

3) What should be the correlation threshold value based on which we determine the highly collinear variables?

Ans: 0.8-1.0: strongly Correlated

0.6-0.8: Correlated

0.4-0.6: Moderate

0.2-0.4: weakly correlated

0: No correlation

Take the threshold to be larger than the usual 0.05

4) What are the two different types of variable we used in ANOVA?

Ans: The data consists of one variable as continuous and other variable as categorical then we use ANOVA test.

5) What are the null and alternate hypothesis in chi-square test?

Ans: chi-square test is used to check whether there is an association or not.

Ho (null-hypothesis): There is no association between variable

Ha (alternative hypothesis):There is an association between variables

If the observed chi-square test statistic is greater than the critical value, the null hypothesis can be rejected.