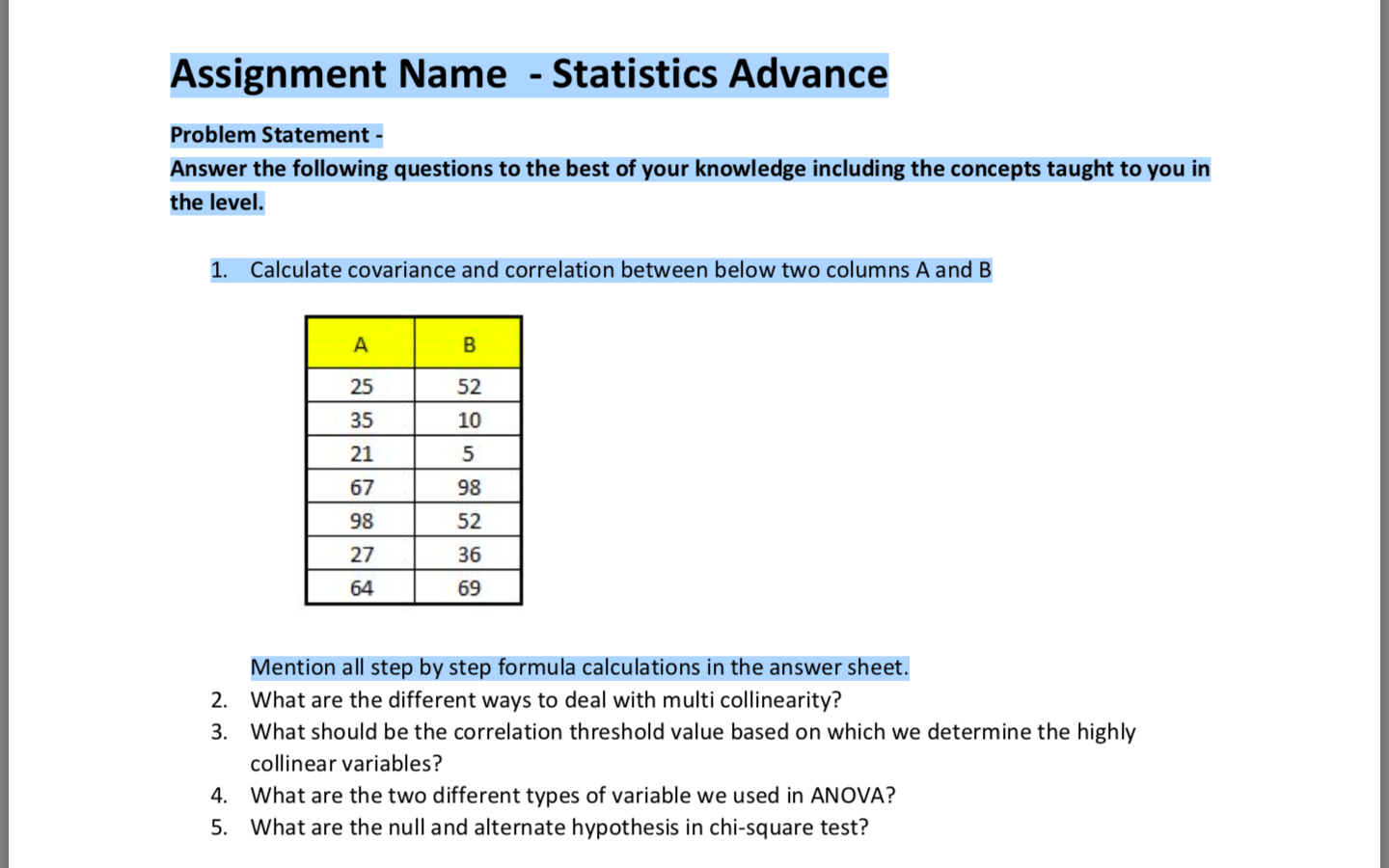
***Assignment Name - Statistics Advance***

***Problem Statement -  
Answer the following questions to the best of your knowledge including the concepts taught to you in the level.***

1. ***Calculate covariance and correlation between below two columns A and B***

***Mention all step by step formula calculations in the answer sheet.***

******

***solution:***

*Given Data Sets are -*

*A = {25,35,21,67,98,27,64}*

*B={52,10,5,98,52,36,69}*

*NA = 7 and NB= 7*

*Sum of all the values of A : = 337*

*Sum of all the values of B : = 322*

*Mean of A = ( Σ xi ) / N = 337/7 = 48.14285714 ~* ***48.14***

*Mean of B = ( Σ xi ) / N = 322/7 =* ***46***

*A – MeanA = { -23.14,-13.14,-27.14,18.86,49.86,-21.14,18.86,49.86,-21.14,15.86}*

*B – MeanB = {6,-36,-41,52,6,-10,23}*

*VarianceA = ( Σ (xi – mean)2 ) / N*

*=(535.4596+172.6596+736.5796+355.6996+2486.0196++446.8996+251.5396)/7*

|  |
| --- |
|  |
|  |
|  |

*=4984.8572/7 = 712.1224571 ~* ***712.12***

*VarianceB =( Σ (xi – mean)2 ) /N = (36+1296+1681+2704+36+100+529)/7*

*=6382/7 = 911.7142857 ~* ***911.71***

*SdA = SQRT (VarianceA) = SQRT (712.12) =26.68562267 ~* ***26.686***

*SdB = SQRT (VarianceB) = SQRT (911.71) = 30.1946069 ~* ***30.195***

***Covariance Cov ( A , B ) =***

*= { -23.14, -13.14, -27.14, 18.86,49.86, - 21.14, 15.86}*

*= { 6, -36, -41, 52, 6, -10, 23 }*

***=*** *{ -138.84, 473.04, 1112.74, 980.72, 299.16, 211.4, 364.78}*

*sum = 3303*

*Cov ( A , B ) = 3303 / 7 = 471.857*

***covariance = 471.857 ~ 471.86***

***Correlation ( A , B ) = Cov ( A , B ) / SdA \* SdB***

*= 471.857 / 26.686\*30.195*

*= 471.857/805.7618863*

***= 0.585603701 ~ 0.59***

***Summary of calculations –***

*1.Covariance of A and B is* ***471.86***

*2.Corelation between A and B is* ***0.59***

*3.The observation shows that A and B variables are postively corelated with each other.*

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

***Solution :***

*There are 2 major methods to deal with Collinearity :*

1. *To remove the redundancy , by eleminating the independent variables that are corelated to each other with the help of some selection techniques.*
2. *to ignore it, if our focus is mainly on the target variable , and if there is an independent variable correlated with the target variable then we can ignore the independent variable and focus on the target variable.*

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

***Solution :***

*if a correlation value is nearly or equal to 1 something which is above 0.74 then we can say that they are highly correlated with each other.*

*Suppose there are two values with high correlation value which is more than 0.74 . Then they can come under either of the categories*

*• X1 and X2 may be two different ways of measuring the same thing.*

*• X1 and X2 are so strongly confounded that their predictive contributions will be very difficult to separate.*

*Based on the above analysis we can determine that the two variable are highly collinear variables.*

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

***Solution :***

1. *Numerical variable – for example mean of the categorial variables*
2. *Categorial Variable – for example the two gender male , female*

*we calculate the mean of male population and female population. then perform the statistical analyisis to obtain conclusions.*

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

***Solution :***

*The Null Hyposthesis ( H0) = The two variables are independent to each other*

*Alternative Hypothes ( H1)  = The two variables are dependent to each other*

*ex:*

*There are two workdone and health.*

*Work done = { completed , incomplete }*

*Health = { Well , not Well }*

*H0  = The Workdone status is independent of health*

*H1= The work done status depends on health*