# ***Assignment Name - Statistics Advance***

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

|  |  |
| --- | --- |
| Column A | Column B |
| 25  35  21  67  98  27  64 | 52  10  5  98  52  36  69 |

**Answer: -**

**Cov (A, B)=**

**Here,**

**= 48.143**

**Similarly, = = 46**

**Cov(A, B)=**

**=**

**=**

**=471.857**

**These two variables have positive covariance.**

**We know, correlation is the ration of covariance and standard deviation. So ,**

**Cor(A,B) =**

**Here,**

**Sd(A) =**

**=**

**=**

**= = 30.194**

**Sd(B) =**

**=**

**=**

**= 712.122 = 26.686**

**Cor(A,B) = = = = 0.585**

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

**Answer: -** *Multi collinearity results in unstable parameter estimates which makes it very difficult to assess the effect of independent variables on dependent variables. Generally Multi Collinearity occurs when the variables are highly correlated to each other.*

*The ways to deal with Multicollinearity are:*

* *Remove highly correlated variables from the model.*
* [*Partial Least Squares Regression (PLS)*](http://blog.minitab.com/blog/statistics-and-quality-data-analysis/giving-thanks-for-the-regression-menu-v2)*or*[*Principal Components Analysis*](http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/multivariate/principal-components-and-factor-analysis/what-is-pca/)*, these two regression methods cut the number of variables to a smaller set of uncorrelated components.*

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

**Answer: -** *We consider correlation threshold as 0.7.The correlation between each of our independent variable, should not be 0.7. If two independent variables are having a correlation of 0.7 we can ignore one.   
 We need to check Tolerance and VIF values as well. If Tolerance (Tolerance is the amount of variability in one independent variable that is no explained by the other independent variables.) value is less than 0.10 and VIF (variance inflation factor) above 10 then we can say there is collinearity between two of our variables.*

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

**Answer: -** *There are two types of variables used in ANOVA.   
 1. Independent Variable*

*2. Dependent Variable*

***Independent Variable:***

*Statistical researchers can manipulate Independent variable to make grouping of observations.   
In One-way ANOVA, we consider only one Independent variables. There are two types of independent variables:*

* + - 1. *Attribute.*
      2. *Active*

*If the independent variable is an* ***active variable*** *then we manipulate the values of the variable to study its effect on another variable.*

*An* ***attribute independent variable*** *is a variable where we do not alter the variable during the manipulation.*

***Dependent Variable:***

*The dependent variable is defined as the variable that is, or is presumed to be, the result of manipulating the independent variable.*

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

**Answer: -** *Chi –Square test is to identify the significant relationship between two categorical variables. Null hypothesis denotes that there is no association between those two variables and Alternative hypothesis denotes the association between two variables.   
 Basically for any random sample the critical value of Chi Square statistic is approximately 0.05. If the observed chi-square test statistic is greater than the critical value, the null hypothesis can be rejected.*