Stat 328

Part-1

Statistical Analysis using Excel

There are many statistical analysis methods can be achieved using Excel

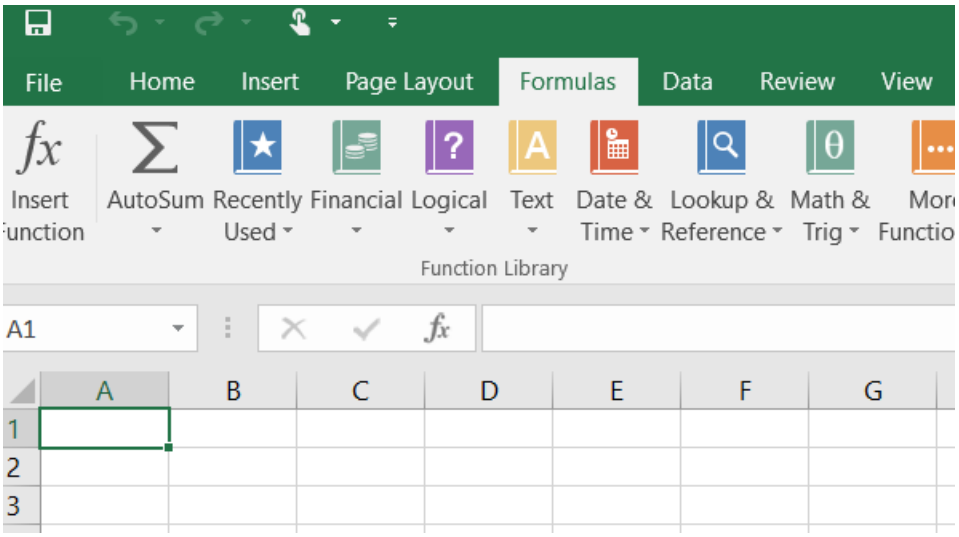
1-1 Functions

Three different type of function can be used in statistics; they are:

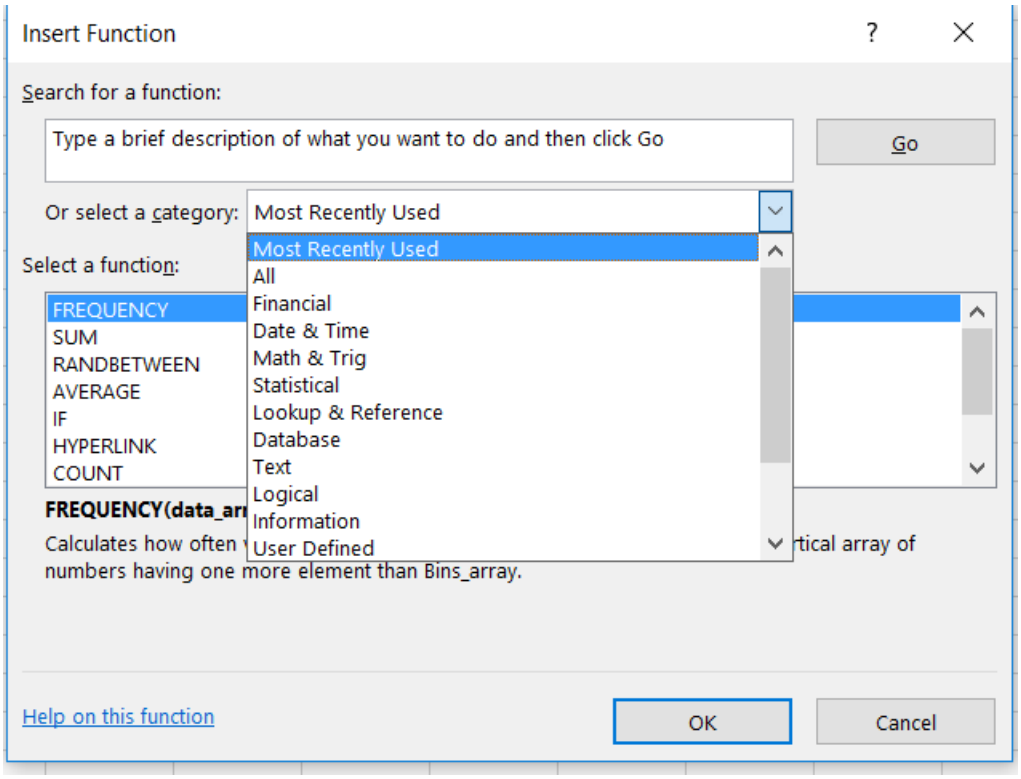
(a) Statistical Functions
(b) Mathematical Functions
(c) Logical Functions

These function can be obtained from excel as:

Click the **Formulas** tab, then select the **Insert Function** command


File
Home
Insert
Page Layout
Formulas
Data
Review
View
Insert Function
AutoSum
Recently Used
Financial
Logical
Text
Date & Time
Lookup & Reference
Math & Trig
Function Library
A1
A
B
C
D
E
F
G
1
2
3

Then, we get


Insert Function
Search for a function:
Type a brief description of what you want to do and then click Go
Go
Or select a category:
Most Recently Used
Select a function:
FREQUENCY
SUM
RANDBETWEEN
AVERAGE
IF
HYPERLINK
COUNT
Most Recently Used
All
Financial
Date & Time
Math & Trig
Statistical
Lookup & Reference
Database
Text
Logical
Information
User Defined
numbers having one more element than Bins_array.
Help on this function
OK
Cancel

So, if we select the statistical to get a list of possible statistical functions as follows:

Count & Frequency	
COUNT	Returns the number of numerical values in a supplied set of cells or values
COUNTA	Returns the number of non-blanks in a supplied set of cells or values
COUNTBLANK	Returns the number of blank cells in a supplied range
COUNTIF	Returns the number of cells (of a supplied range), that satisfy a given criteria
COUNTIFS	Returns the number of cells (of a supplied range), that satisfy a set of given criteria *(New in Excel 2007)*
FREQUENCY	Returns an array showing the number of values from a supplied array, which fall into specified ranges

Finding the Largest & Smallest Values	
MAX	Returns the largest value from a list of supplied numbers
MAXA	Returns the largest value from a list of supplied values, counting text and the logical value FALSE as the value 0 and counting the logical value TRUE as the value 1
MAXIFS	Returns the largest value from a subset of values in a list that are specified according to one or more criteria. *(New in Excel 2016)*
MIN	Returns the smallest value from a list of supplied numbers
MINA	Returns the smallest value from a list of supplied values, counting text and the logical value FALSE as the value 0 and counting the logical value TRUE as the value 1
MINIFS	Returns the smallest value from a subset of values in a list that are specified according to one or more criteria. *(New in Excel 2016)*
LARGE	Returns the Kth LARGEST value from a list of supplied numbers, for a given value K
SMALL	Returns the Kth SMALLEST value from a list of supplied numbers, for a given value K

Percentiles, Quartiles & Rank

PERCENTILE	Returns the K'th percentile of values in a supplied range, where K is in the range 0 - 1 (inclusive) *(Replaced by Percentile.Inc function in Excel 2010)*
PERCENTILE.INC	Returns the K'th percentile of values in a supplied range, where K is in the range 0 - 1 (inclusive) *(New in Excel 2010 - replaces the Percentile function)*
PERCENTILE.EXC	Returns the K'th percentile of values in a supplied range, where K is in the range 0 - 1 (exclusive) *(New in Excel 2010)*
QUARTILE	Returns the specified quartile of a set of supplied numbers, based on percentile value 0 - 1 (inclusive) *(Replaced by Quartile.Inc function in Excel 2010)*
QUARTILE.INC	Returns the specified quartile of a set of supplied numbers, based on percentile value 0 - 1 (inclusive) *(New in Excel 2010 - replaces the Quartile function)*
QUARTILE.EXC	Returns the specified quartile of a set of supplied numbers, based on percentile value 0 - 1 (exclusive) *(New in Excel 2010)*
RANK	Returns the statistical rank of a given value, within a supplied array of values *(Replaced by Rank.Eq function in Excel 2010)*
RANK.EQ	Returns the Mode (the most frequently occurring value) of a list of supplied numbers (if more than one value has same rank, the top rank of that set is returned) *(New in Excel 2010 - replaces the Rank function)*
RANK.AVG	Returns the statistical rank of a given value, within a supplied array of values (if more than one value has same rank, the average rank is returned) *(New in Excel 2010)*
PERCENTRANK	Returns the rank of a value in a data set, as a percentage (0 - 1 inclusive) *(Replaced by Percentrank.Inc function in Excel 2010)*
PERCENTRANK.INC	Returns the rank of a value in a data set, as a percentage (0 - 1 inclusive) *(New in Excel 2010 - replaces the Percentrank function)*
PERCENTRANK.EXC	Returns the rank of a value in a data set, as a percentage (0 - 1 exclusive) *(New in Excel 2010)*

Averages

AVERAGE	Returns the Average of a list of supplied numbers
AVERAGEA	Returns the Average of a list of supplied numbers, counting text and the logical value FALSE as the value 0 and counting the logical value TRUE as the value 1
AVERAGEIF	Calculates the Average of the cells in a supplied range, that satisfy a given criteria *(New in Excel 2007)*
AVERAGEIFS	Calculates the Average of the cells in a supplied range, that satisfy multiple criteria *(New in Excel 2007)*

MEDIAN	Returns the Median (the middle value) of a list of supplied numbers
MODE	Returns the Mode (the most frequently occurring value) of a list of supplied numbers *(Replaced by Mode.Sngl function in Excel 2010)*
MODE.SNGL	Returns the Mode (the most frequently occurring value) of a list of supplied numbers *(New in Excel 2010 - replaces the Mode function)*
MODE.MULT	Returns a vertical array of the most frequently occurring values in an array or range of data *(New in Excel 2010)*
GEOMEAN	Returns the geometric mean of a set of supplied numbers
HARMEAN	Returns the harmonic mean of a set of supplied numbers
TRIMMEAN	Returns the mean of the interior of a supplied set of values

Deviation & Variance

AVEDEV	Returns the average of the absolute deviations of data points from their mean
DEVSQ	Returns the sum of the squares of the deviations of a set of data points from their sample mean
STDEV	Returns the standard deviation of a supplied set of values (which represent a sample of a population) *(Replaced by Stdev.S function in Excel 2010)*
STDEV.S	Returns the standard deviation of a supplied set of values (which represent a sample of a population) *(New in Excel 2010 - replaces the Stdev function)*
STDEVA	Returns the standard deviation of a supplied set of values (which represent a sample of a population), counting text and the logical value FALSE as the value 0 and counting the logical value TRUE as the value 1
STDEVP	Returns the standard deviation of a supplied set of values (which represent an entire population) *(Replaced by Stdev.P function in Excel 2010)*
STDEV.P	Returns the standard deviation of a supplied set of values (which represent an entire population) *(New in Excel 2010 - replaces the Stdevp function)*
STDEVPA	Returns the standard deviation of a supplied set of values (which represent an entire population), counting text and the logical value FALSE as the value 0 and counting the logical value TRUE as the value 1
VAR	Returns the variance of a supplied set of values (which represent a sample of a population) *(Replaced by Var.S function in Excel 2010)*

VAR.S	Returns the variance of a supplied set of values (which represent a sample of a population) *(New in Excel 2010 - replaces the Var function)*
VARA	Returns the variance of a supplied set of values (which represent a sample of a population), counting text and the logical value FALSE as the value 0 and counting the logical value TRUE as the value 1
VARP	Returns the variance of a supplied set of values (which represent an entire population) *(Replaced by Var.P function in Excel 2010)*
VAR.P	Returns the variance of a supplied set of values (which represent an entire population) *(New in Excel 2010 - replaces the Varp function)*
VARPA	Returns the variance of a supplied set of values (which represent an entire population), counting text and the logical value FALSE as the value 0 and counting the logical value TRUE as the value 1
COVAR	Returns population covariance (i.e. the average of the products of deviations for each pair within two supplied data sets) *(Replaced by Covariance.P function in Excel 2010)*
COVARIANCE.P	Returns population covariance (i.e. the average of the products of deviations for each pair within two supplied data sets) *(New in Excel 2010 - replaces the Covar function)*
COVARIANCE.S	Returns sample covariance (i.e. the average of the products of deviations for each pair within two supplied data sets) *(New in Excel 2010)*

Confidence Intervals	
CONFIDENCE	Returns the confidence interval for a population mean, using a normal distribution *(Replaced by Confidence.Norm function in Excel 2010)*
CONFIDENCE.NORM	Returns the confidence interval for a population mean, using a normal distribution *(New in Excel 2010 - replaces the Confidence function)*
CONFIDENCE.T	Returns the confidence interval for a population mean, using a Student's t distribution *(New in Excel 2010)*

Trend Line Functions	
FORECAST	Predicts a future point on a linear trend line fitted to a supplied set of x- and y- values
INTERCEPT	Calculates the best fit regression line, through a supplied series of x- and y-values and returns the value at which this line intercepts the y-axis
LINEST	Returns statistical information describing the trend of the line of best fit, through a supplied series of x- and y- values

SLOPE	Returns the slope of the linear regression line through a supplied series of x- and y- values
TREND	Calculates the trend line through a given set of y-values and returns additional y-values for a supplied set of new x-values
GROWTH	Returns numbers in a exponential growth trend, based on a set of supplied x- and y- values
LOGEST	Returns the parameters of an exponential trend for a supplied set of x- and y- values
STEYX	Returns the standard error of the predicted y-value for each x in the regression line for a set of supplied x- and y- values

PERMUT	Returns the number of permutations for a given number of objects
PERMUTATIONA	Returns the number of permutations for a given number of objects (with repetitions) that can be selected from the total objects *(New in Excel 2013)*

Distribution & Tests of Probability

BETADIST	Returns the cumulative beta probability density function *(Replaced by Beta.Dist function in Excel 2010)*
BETA.DIST	Returns the cumulative beta distribution function or the beta probability density function *(New in Excel 2010 - replaces the Betadist function)*
BETAINV	Returns the inverse of the cumulative beta probability density function *(Replaced by Beta.Inv function in Excel 2010)*
BETA.INV	Returns the inverse of the cumulative beta probability density function *(New in Excel 2010 - replaces the Betainv function)*
BINOMDIST	Returns the individual term binomial distribution probability *(Replaced by Binom.Dist function in Excel 2010)*
BINOM.DIST	Returns the individual term binomial distribution probability *(New in Excel 2010 - replaces the Binomdist function)*

BINOM.DIST.RANGE	Returns the probability of a trial result using a binomial distribution *(New in Excel 2013)*
NEGBINOMDIST	Returns the negative binomial distribution *(Replaced by Negbinom.Dist function in Excel 2010)*
NEGBINOM.DIST	Returns the negative binomial distribution *(New in Excel 2010 - replaces the Negbinomdist function)*
CRITBINOM	Returns the smallest value for which the cumulative binomial distribution is greater than or equal to a criterion value *(Replaced by Binom.Inv function in Excel 2010)*
BINOM.INV	Returns the smallest value for which the cumulative binomial distribution is greater than or equal to a criterion value *(New in Excel 2010 - replaces the Critbinom function)*
CHIDIST	Returns the right-tailed probability of the chi-squared distribution *(Replaced by Chisq.Dist.Rt function in Excel 2010)*
CHISQ.DIST.RT	Returns the right-tailed probability of the chi-squared distribution *(New in Excel 2010 - replaces the Chidist function)*
CHISQ.DIST	Returns the chi-squared distribution (probability density or cumulative distribution function) *(New in Excel 2010)*
CHIINV	Returns the inverse of the right-tailed probability of the chi-squared distribution *(Replaced by Chisq.Inv.Rt function in Excel 2010)*
CHISQ.INV.RT	Returns the inverse of the right-tailed probability of the chi-squared distribution *(New in Excel 2010 - replaces the Chiinv function)*
CHISQ.INV	Returns the inverse of the left-tailed probability of the chi-squared distribution *(New in Excel 2010)*
CHITEST	Returns the chi-squared statistical test for independence *(Replaced by Chisq.Test function in Excel 2010)*
CHISQ.TEST	Returns the chi-squared statistical test for independence *(New in Excel 2010 - replaces the Chitest function)*
CORREL	Returns the correlation coefficient between two sets of values
EXPONDIST	Returns the exponential distribution *(Replaced by Expon.Dist function in Excel 2010)*
EXPON.DIST	Returns the exponential distribution *(New in Excel 2010 - replaces the Expondist function)*
FDIST	Returns the right-tailed F probability distribution for two data sets *(Replaced by F.Dist.Rt function in Excel 2010)*

F.DIST.RT	Returns the right-tailed F probability distribution for two data sets *(New in Excel 2010 - replaces the Fdist function)*
F.DIST	Returns the F probability distribution (probability density or cumulative distribution function) *(New in Excel 2010)*
FINV	Returns the inverse of the right-tailed F probability distribution for two data sets *(Replaced by F.Inv.Rt function in Excel 2010)*
F.INV.RT	Returns the inverse of the right-tailed F probability distribution for two data sets *(New in Excel 2010 - replaces the Finv function)*
F.INV	Returns the inverse of the Cumulative F distribution *(New in Excel 2010)*
FISHER	Returns the Fisher transformation
FISHERINV	Returns the inverse of the Fisher transformation
FTEST	Returns the result of an F-Test for 2 supplied data sets *(Replaced by F.Test function in Excel 2010)*
F.TEST	Returns the result of an F-Test for 2 supplied data sets *(New in Excel 2010 - replaces the Ftest function)*
GAMMADIST	Returns the gamma distribution *(Replaced by Gamma.Dist function in Excel 2010)*
GAMMA.DIST	Returns the gamma distribution *(New in Excel 2010 - replaces the Gammadist function)*
GAMMAINV	Returns the inverse gamma cumulative distribution *(Replaced by Gamma.Inv function in Excel 2010)*
GAMMA.INV	Returns the inverse gamma cumulative distribution *(New in Excel 2010 - replaces the Gammainv function)*
GAMMA	Return the gamma function value for a supplied number *(New in Excel 2013)*
GAMMALN	Calculates the natural logarithm of the gamma function for a supplied value
GAMMALN.PRECISE	Returns the natural logarithm of the gamma function for a supplied value *(New in Excel 2010)*
GAUSS	Calculates the probability that a member of a standard normal population will fall between the mean and z standard deviations from the mean *(New in Excel 2013)*
HYPGEOMDIST	Returns the hypergeometric distribution *(Replaced by Hypgeom.Dist function in Excel 2010)*
HYPGEOM.DIST	Returns the hypergeometric distribution *(New in Excel 2010 - replaces the Hypgeomdist function)*

KURT	Returns the kurtosis of a data set
LOGNORMDIST	Returns the cumulative log-normal distribution *(Replaced by Lognorm.Dist function in Excel 2010)*
LOGNORM.DIST	Returns the log-normal probability density function or the cumulative log- normal distribution *(New in Excel 2010 - replaces the Lognormdist function)*
LOGINV	Returns the inverse of the lognormal distribution *(Replaced by Lognorm.Inv function in Excel 2010)*
LOGNORM.INV	Returns the inverse of the lognormal distribution *(New in Excel 2010 - replaces the Loginv function)*
NORMDIST	Returns the normal cumulative distribution *(Replaced by Norm.Dist function in Excel 2010)*
NORM.DIST	Returns the normal cumulative distribution *(New in Excel 2010 - replaces the Normdist function)*
NORMINV	Returns the inverse of the normal cumulative distribution *(Replaced by Norm.Inv function in Excel 2010)*
NORM.INV	Returns the inverse of the normal cumulative distribution *(New in Excel 2010 - replaces the Norminv function)*
NORMSDIST	Returns the standard normal cumulative distribution *(Replaced by Norm.S.Dist function in Excel 2010)*
NORM.S.DIST	Returns the standard normal cumulative distribution *(New in Excel 2010 - replaces the Normsdist function)*
NORMSINV	Returns the inverse of the standard normal cumulative distribution *(Replaced by Norm.S.Inv function in Excel 2010)*
NORM.S.INV	Returns the inverse of the standard normal cumulative distribution *(New in Excel 2010 - replaces the Normsinv function)*
PEARSON	Returns the Pearson product moment correlation coefficient
RSQ	Returns the square of the Pearson product moment correlation coefficient
PHI	Returns the value of the density function for a standard normal distribution, for a supplied number *(New in Excel 2013)*
POISSON	Returns the Poisson distribution *(Replaced by Poisson.Dist function in Excel 2010)*
POISSON.DIST	Returns the Poisson distribution *(New in Excel 2010 - replaces the Poisson function)*
PROB	Returns the probablity that values in a supplied range are within given limits

SKEW	Returns the skewness of a distribution
SKEW.P	Returns the skewness of a distribution based on a population *(New in Excel 2013)*
STANDARDIZE	Returns a normalized value
TDIST	Returns the Student's T-distribution *(Replaced by T.Dist.2t & T.Dist.Rt functions in Excel 2010)*
T.DIST.2T	Returns the two-tailed Student's T-distribution *(New in Excel 2010 - replaces the Tdist function)*
T.DIST.RT	Returns the right-tailed Student's T-distribution *(New in Excel 2010 - replaces the Tdist function)*
T.DIST	Returns the Student's T-distribution (probability density or cumulative distribution function) *(New in Excel 2010)*
TINV	Returns the two-tailed inverse of the Student's T-distribution *(Replaced by T.Inv.2t function in Excel 2010)*
T.INV.2T	Returns the two-tailed inverse of the Student's T-distribution *(New in Excel 2010 - replaces the Tinv function)*
T.INV	Returns the left-tailed inverse of the Student's T-distribution *(New in Excel 2010)*
TTEST	Returns the probability associated with a Student's T-Test *(Replaced by T.Test function in Excel 2010)*
T.TEST	Returns the probability associated with a Student's T-Test *(New in Excel 2010 - replaces the Ttest function)*
WEIBULL	Returns the Weibull distribution *(Replaced by Weibull.Dist function in Excel 2010)*
WEIBULL.DIST	Returns the Weibull distribution *(New in Excel 2010 - replaces the Weibull function)*
ZTEST	Returns the one-tailed probability value of a z-test *(Replaced by Z.Test function in Excel 2010)*
Z.TEST	Returns the one-tailed probability value of a z-test *(New in Excel 2010 - replaces the Ztest function)*

Similarly, if we select the Math& Trig We get a list of possible Mathematical functions as follows:

Basic Numeric Information	
ABS	Returns the absolute value (i.e. the modulus) of a supplied number
SIGN	Returns the sign (+1, -1 or 0) of a supplied number
GCD	Returns the Greatest Common Divisor of two or more supplied numbers
LCM	Returns the Least Common Multiple of two or more supplied numbers

Basic Mathematical Operations	
SUM	Returns the sum of a supplied list of numbers
PRODUCT	Returns the product of a supplied list of numbers
POWER	Returns the result of a given number raised to a supplied power
SQRT	Returns the positive square root of a given number
QUOTIENT	Returns the integer portion of a division between two supplied numbers
MOD	Returns the remainder from a division between two supplied numbers
AGGREGATE	Performs a specified calculation (e.g. the sum, product, average, etc.) for a list or database, with the option to ignore hidden rows and error values *(New in Excel 2010)*
SUBTOTAL	Performs a specified calculation (e.g. the sum, product, average, etc.) for a supplied set of values

Rounding Functions	
CEILING	Rounds a number away from zero (i.e. rounds a positive number up and a negative number down), to a multiple of significance
CEILING.PRECISE	Rounds a number up, regardless of the sign of the number, to a multiple of significance *(New in Excel 2010)*

ISO.CEILING	Rounds a number up, regardless of the sign of the number, to a multiple of significance. *(New in Excel 2010)*
CEILING.MATH	Rounds a number up to the nearest integer or to the nearest multiple of significance *(New in Excel 2013)*
EVEN	Rounds a number away from zero (i.e. rounds a positive number up and a negative number down), to the next even number
FLOOR	Rounds a number towards zero, (i.e. rounds a positive number down and a negative number up), to a multiple of significance
FLOOR.PRECISE	Rounds a number down, regardless of the sign of the number, to a multiple of significance *(New in Excel 2010)*
FLOOR.MATH	Rounds a number down, to the nearest integer or to the nearest multiple of significance *(New in Excel 2013)*
INT	Rounds a number down to the next integer
MROUND	Rounds a number up or down, to the nearest multiple of significance
ODD	Rounds a number away from zero (i.e. rounds a positive number up and a negative number down), to the next odd number
ROUND	Rounds a number up or down, to a given number of digits
ROUNDDOWN	Rounds a number towards zero, (i.e. rounds a positive number down and a negative number up), to a given number of digits
ROUNDUP	Rounds a number away from zero (i.e. rounds a positive number up and a negative number down), to a given number of digits
TRUNC	Truncates a number towards zero (i.e. rounds a positive number down and a negative number up), to the next integer.

Matrix Functions	
MDETERM	Returns the matrix determinant of a supplied array
MINVERSE	Returns the matrix inverse of a supplied array
MMULT	Returns the matrix product of two supplied arrays
MUNIT	Returns the unit matrix for a specified dimension *(New in Excel 2013)*

Random Numbers	
RAND	Returns a random number between 0 and 1

RANDBETWEEN	Returns a random number between two given integers

Conditional Sums	
SUMIF	Adds the cells in a supplied range, that satisfy a given criteria
SUMIFS	Adds the cells in a supplied range, that satisfy multiple criteria *(New in Excel 2007)*

Advanced Mathematical Operations	
SUMPRODUCT	Returns the sum of the products of corresponding values in two or more supplied arrays
SUMSQ	Returns the sum of the squares of a supplied list of numbers
SUMX2MY2	Returns the sum of the difference of squares of corresponding values in two supplied arrays
SUMX2PY2	Returns the sum of the sum of squares of corresponding values in two supplied arrays
SUMXMY2	Returns the sum of squares of differences of corresponding values in two supplied arrays
SERIESSUM	Returns the sum of a power series

Trigonometry Functions	
PI	Returns the constant value of pi
SQRTPI	Returns the square root of a supplied number multiplied by pi
DEGREES	Converts Radians to Degrees
RADIANS	Converts Degrees to Radians
COS	Returns the Cosine of a given angle
ACOS	Returns the Arccosine of a number
COSH	Returns the hyperbolic cosine of a number
ACOSH	Returns the inverse hyperbolic cosine of a number
SEC	Returns the secant of an angle *(New in Excel 2013)*
SECH	Returns the hyperbolic secant of an angle *(New in Excel 2013)*
SIN	Returns the Sine of a given angle
ASIN	Returns the Arcsine of a number

SINH	Returns the Hyperbolic Sine of a number
ASINH	Returns the Inverse Hyperbolic Sine of a number
CSC	Returns the cosecant of an angle *(New in Excel 2013)*
CSCH	Returns the hyperbolic cosecant of an angle *(New in Excel 2013)*
TAN	Returns the Tangent of a given angle
ATAN	Returns the Arctangent of a given number
ATAN2	Returns the Arctangent of a given pair of x and y coordinates
TANH	Returns the Hyperbolic Tangent of a given number
ATANH	Returns the Inverse Hyperbolic Tangent of a given number
COT	Returns the cotangent of an angle *(New in Excel 2013)*
COTH	Returns the hyperbolic cotangent of an angle *(New in Excel 2013)*
ACOT	Returns the arccotangent of a number *(New in Excel 2013)*
ACOTH	Returns the hyperbolic arccotangent of a number *(New in Excel 2013)*

Exponents & Logarithms

EXP	Returns *e* raised to a given power
LN	Returns the natural logarithm of a given number
LOG	Returns the logarithm of a given number, to a specified base
LOG10	Returns the base 10 logarithm of a given number

Factorials

FACT	Returns the Factorial of a given number
FACTDOUBLE	Returns the Double Factorial of a given number
MULTINOMIAL	Returns the Multinomial of a given set of numbers

Miscellaneous

BASE	Converts a number into a text representation, with the supplied base *(New in Excel 2013)*
DECIMAL	Converts a text representation of a number in a specified base into a decimal number *(New in Excel 2013)*
COMBIN	Returns the number of combinations (without repititions) for a given number of objects
COMBINA	Returns the number of combinations (with repetitions) for a given number of items *(New in Excel 2013)*
ARABIC	Converts a Roman numeral to an Arabic numeral *(New in Excel 2013)*

ROMAN	Returns a text string depicting the roman numeral for a given number

Also, if we select the Logical, we get a list of possible logical functions as follows:

Function	Description
AND function	Returns TRUE if all of its arguments are TRUE
FALSE function	Returns the logical value FALSE
IF function	Specifies a logical test to perform
IFERROR function	Returns a value you specify if a formula evaluates to an error; otherwise, returns the result of the formula
IFNA function	Returns the value you specify if the expression resolves to #N/A, otherwise returns the result of the expression
IFS function	Checks whether one or more conditions are met and returns a value that corresponds to the first TRUE condition.
NOT function	Reverses the logic of its argument
OR function	Returns TRUE if any argument is TRUE
SWITCH function	Evaluates an expression against a list of values and returns the result corresponding to the first matching value. If there is no match, an optional default value may be returned.
TRUE function	Returns the logical value TRUE
XOR function	Returns a logical exclusive OR of all arguments

Examples

1- Math & Tri functions

No	Function	syntax	Description	Examples
1	ABS	ABS(number)	Valuate the absolute value of a number	Abs(6) = 6 Abs(-5) = 5
2	Combin	Combin(n;x)	Calculate the binomial coefficients $\binom{n}{x}$	Combin(5;3)=10
3				
4				
5				
6				

2-Statistical functions

No	Function	syntax	Description	Examples
1	Average	average(a1; a2; … an)	Calculate the mean of a1, a2, ….an	AVERAGE(1,8,9,2)=5
2	normdist	Normdist(x; mean; standard deviation; cumulative)	Calculate the cumulative probability function on normal distribution with given mean and standard derivation)	Normdist(1.96; 0; 1; True) = 0.9750
3	Norminv			

3-Logical statements

No	Function	syntax	Description	Examples
1	If	If(logical test; statment1 if true; statement 2 if false)	If statement; do statement 1 if true, do statement 2 if false	Let x be a column of values 1, 2, 3. write the following statement in the first cell of the column y if(x2<=1; 1; 2). The output of the column y will be 1,2,3.
2				

Some other examples on the statistical functions

Examples

1- Calculate the mean, median, variance, standard deviation and mode(s) for the following data

2, 3, 2, 2, 0, 2, 3, 6, 3, 2, 2, 4, 3, 1, 3, 3, 3, 2, 1, 5

To achieve this in excel put the data in a column (A1:A20) or raw

and use the following commands

=AVERAGE(A1:A20)

=MEDIAN(A1:A20)

=STDEV.S(A1:A20)

=STDEV.P(A1:A20)

=VAR.P(A1:A20)

=VAR.S(A1:A20)

2- Frequency Table (Frequency-example)

3- Probability distributions

If $X \sim N(2,7)$ **,** ***then Calculate***

$$P(X < 1.5), \quad f(3), \quad P(X < k) = 0.25, \; k \; ?$$

Some other examples on the mathematical functions

Example

Calculate exp(-1.5), Log(25) and ln(25)

Example on the logical functions

Use if statement to print the status of the student $(Pass \geq 60, \; fail < 60)$

In creation exam using the following marks

Marks
70
85
83
25
80
98
80
72
42

If(A1<60,"F","P")

Some mathematical, statistical and logical functions in excel

Some Examples on the Mathematical Function

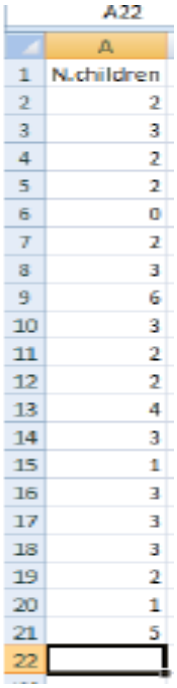
$e^{-1.5}$	**=EXP(-1.5)**	0.22313016
$\log(25)$	**=LOG(25)**	1.397940009
$\ln(25)$	**=LN(25)**	3.218875825
$\binom{10}{2}$	**=COMBIN(10,2)**	45
$\sqrt{25}$	**=SQRT(25)**	5
$\lvert -10 \rvert$	**=ABS(-10)**	10
3^2	**=POWER(3,2)** **=3^2**	9

Some Examples on the Statistical Function

Example

Suppose we are interested in the number of children that a Saudi woman has and we take a sample of 20 women and obtain the following data on the number of children

2, 3, 2, 2, 0, 2, 3, 6, 3, 2, 2, 4, 3, 1, 3, 3, 3, 2, 1, 5



	A
1	N.children
2	2
3	3
4	2
5	2
6	0
7	2
8	3
9	6
10	3
11	2
12	2
13	4
14	3
15	1
16	3
17	3
18	3
19	2
20	1
21	5
22	

- Calculate the mean, median, variance, standard deviation, total, maximum, minimum and mode(s)

To achieve this in excel put the data in a column (A2:A21) or raw and use the following commands

Total	=SUM(A2:A21)		52
Mean	=AVERAGE(A2:A21)		2.6
Median	=MEDIAN(A2:A21)		2.5
Max	=MAX(A2:A21)		6
Min	=MIN (A2:A21)		0
Standard	=STDEV(A2:A21)	New in Excel 2010	1.353358

deviation for sample		=STDEV.S(A2:A21)	
Standard deviation for population	= STDEV P(A2:A21)	New in Excel 2010 =STDEV.P(A2:A21)	1.319091
Variance for sample	=VAR(A2:A21)	New in Excel 2010 =VAR.S(A2:A21)	1.831579
Variance for population	=VARP (A2:A21)	New in Excel 2010 =VAR.S(A2:A21)	1.74

- Frequency Table

N.children		Frequency
0	=COUNTIF(A2:A21,"0")	1
1	=COUNTIF(A2:A21,"1")	2
2	=COUNTIF(A2:A21,"2")	7
3	=COUNTIF(A2:A21,"3")	7
4	=COUNTIF(A2:A21,"4")	1
5	=COUNTIF(A2:A21,"5")	1
6	=COUNTIF(A2:A21,"6")	1

- Probability distributions

If $X \sim N(2,7)$, then Calculate

$P(X < 1.5)$	=NORMDIST(1.5,2,7,TRUE)	0.471528
$P(X < k) = 0.25, k?$	=NORMINV(0.25,2,7)	-2.72143

Example on the Logical Functions

Use if statement to print the status of the student $(Pass \geq 60, Fail < 60)$. In creation exam using the following marks

	A		
1	Marks	Function	Grade
2	70	=IF(A2<60,"Fail","Pass")	Pass
3	85	=IF(A3<60,"Fail","Pass")	Pass
4	83	=IF(A4<60,"Fail","Pass")	Pass
5	25	=IF(A5<60,"Fail","Pass")	Fail
6	80	=IF(A6<60,"Fail","Pass")	Pass
7	98	=IF(A7<60,"Fail","Pass")	Pass

8	80	=IF(A8<60,”Fail”,”Pass”)	Pass
9	72	=IF(A9<60,”Fail”,”Pass”)	Pass
10	42	=IF(A10<60,”Fail”,”Pass”)	Fail
11	32	=IF(A11<60,”Fail”,”Pass”)	Fail

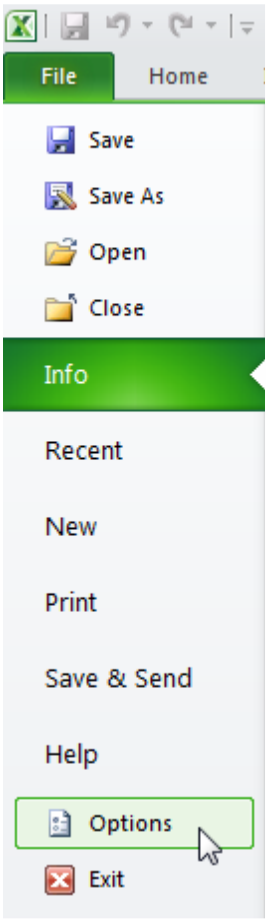
Descriptive statistics using excel

- Data Analysis

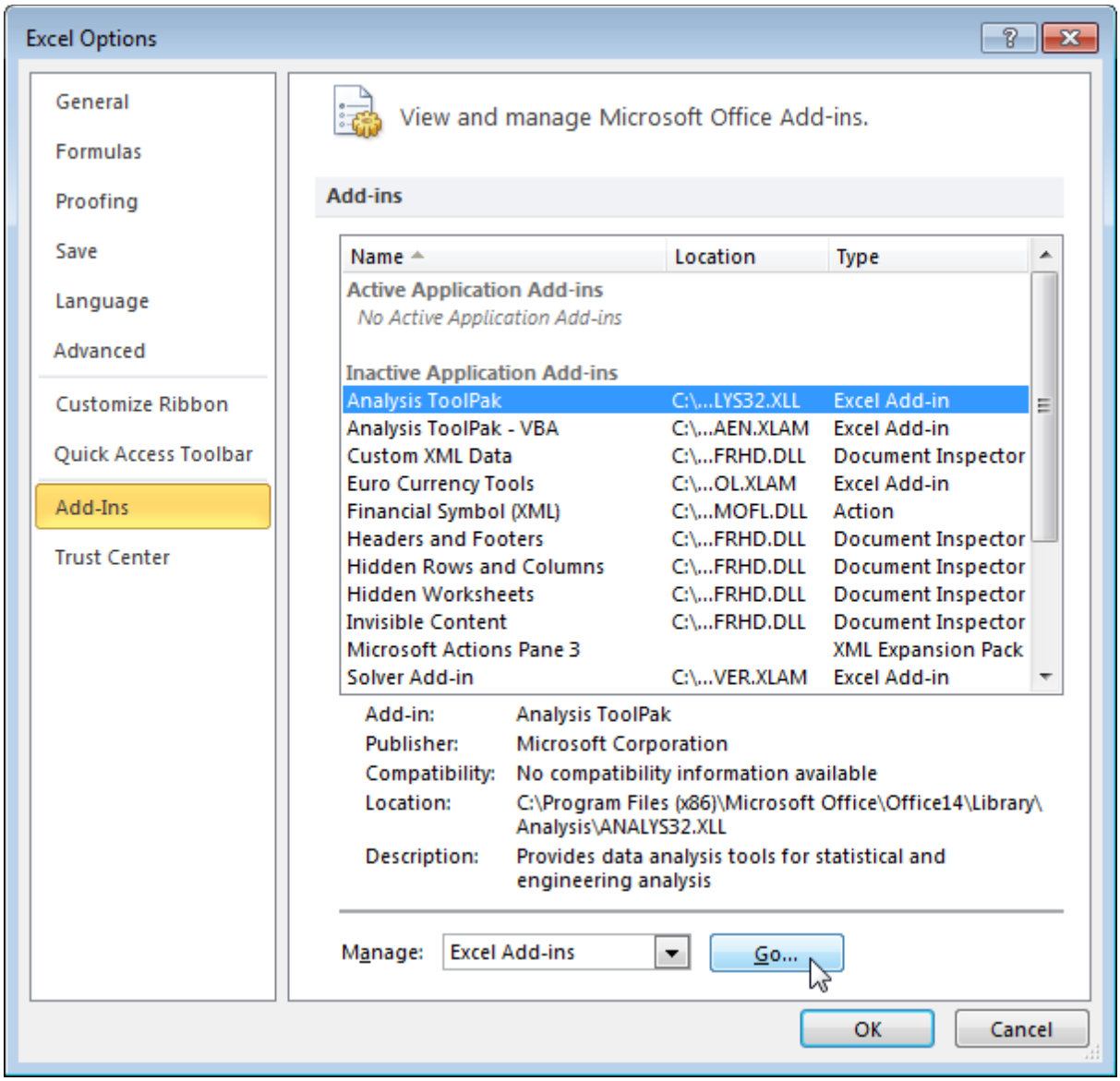
The Analysis ToolPak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis.

To load the Analysis ToolPak add-in, execute the following steps.

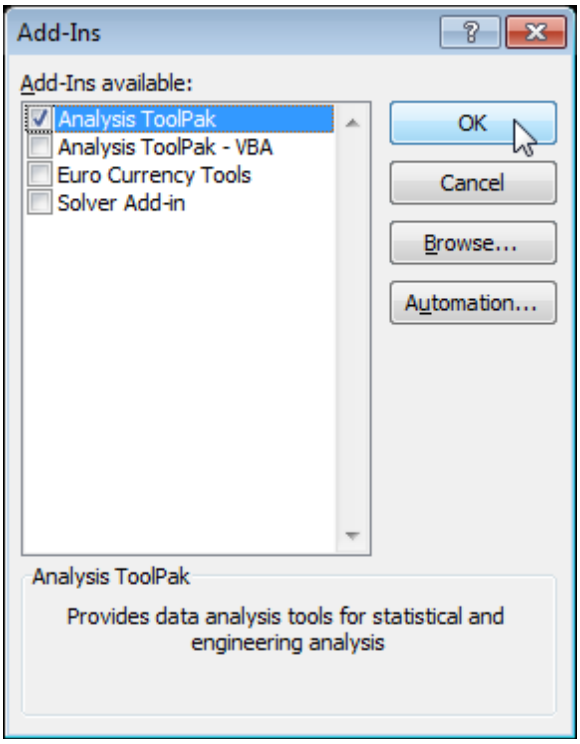
1. Click on Excel Options.



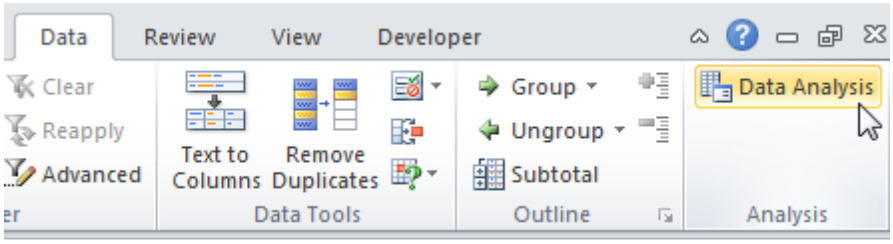
2. Under Add-ins, select Analysis ToolPak and click on the Go button.



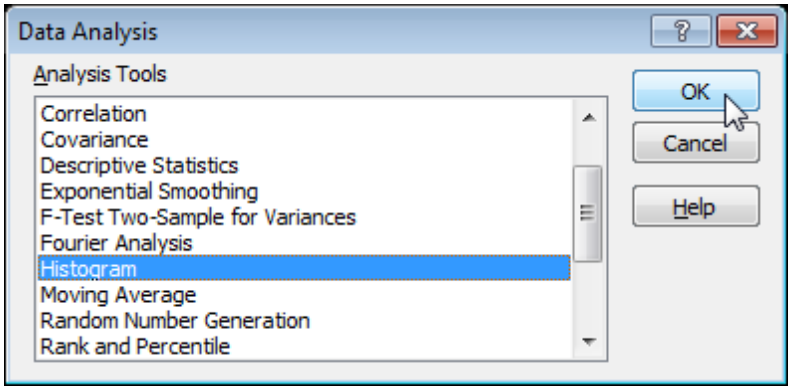
3. Check Analysis ToolPak and click on OK.



4. On the Data tab, you can now click on Data Analysis.



The following dialog box below appears.



Examples on Data analysis

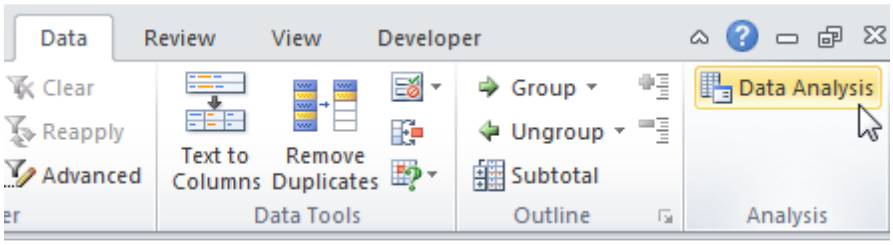
Descriptive Statistics

You can use the Analysis Toolpak add-in to generate descriptive statistics. For example, you may have the scores of 14 participants for a test.

M26

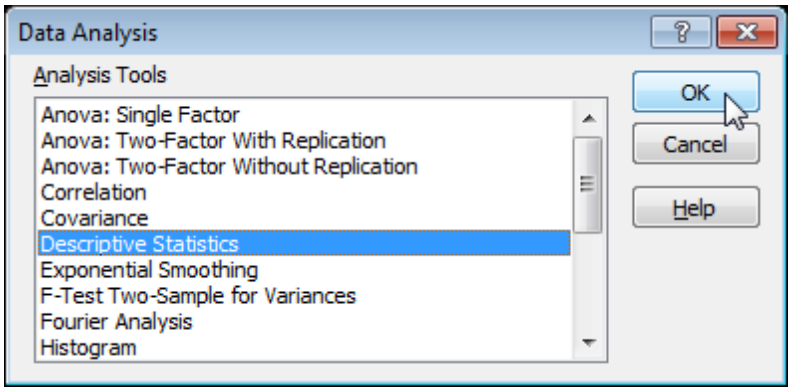
	A	B
1	Scores	
2	82	
3	93	
4	91	
5	69	
6	96	
7	61	
8	88	
9	58	
10	59	
11	100	
12	93	
13	71	
14	78	
15	98	
16		
17		

To generate descriptive statistics for these scores, execute the following steps.
1. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

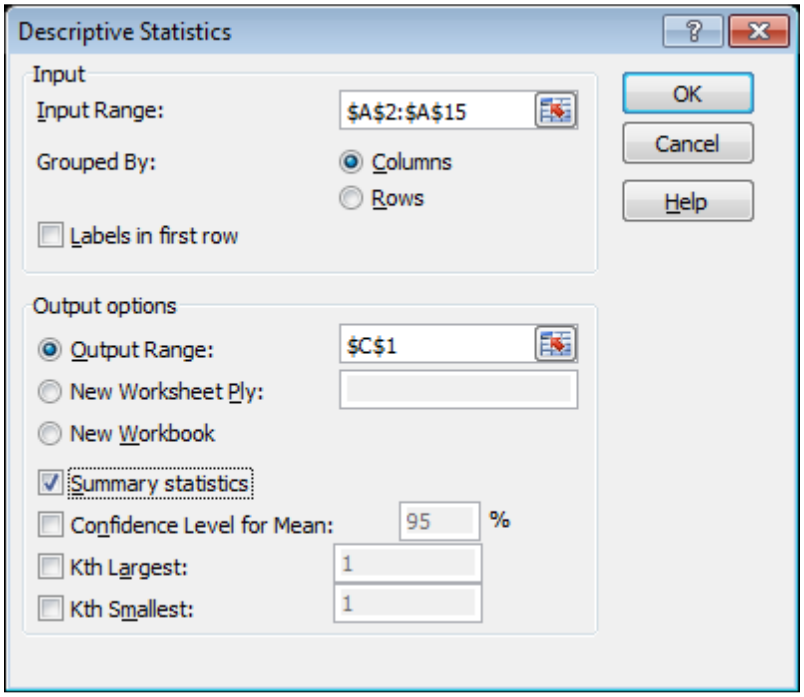
2. Select Descriptive Statistics and click OK.



3. Select the range A2:A15 as the Input Range.

4. Select cell C1 as the Output Range.

5. Make sure Summary statistics is checked.



6. Click OK.

Result:

L28

	A	B	C	D	E
1	Scores		*Column1*		
2	82				
3	93		Mean	81.21428571	
4	91		Standard Error	4.045318243	
5	69		Median	85	
6	96		Mode	93	
7	61		Standard Deviation	15.13619489	
8	88		Sample Variance	229.1043956	
9	58		Kurtosis	-1.426053506	
10	59		Skewness	-0.402108004	
11	100		Range	42	
12	93		Minimum	58	
13	71		Maximum	100	
14	78		Sum	1137	
15	98		Count	14	
16					
17					

Statistical tests using excel

1- Analysis of variance (ANOVA)

- One Way ANOVA

$\alpha = 0.05$.

Hypotheses testing:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$

H_1: at least one of the means is different.

Test statistic:

$$F = \frac{MSA}{MSE}$$

Critical region:

Reject H_0 if $F > F_{1-\alpha,(v_1,v_2)}$

Or

$p - value \leq \alpha$

Decision:

This example teaches you how to perform a single factor ANOVA (analysis of variance) in Excel. A single factor or one-way ANOVA is used to test the null hypothesis that the means of several populations are all equal.

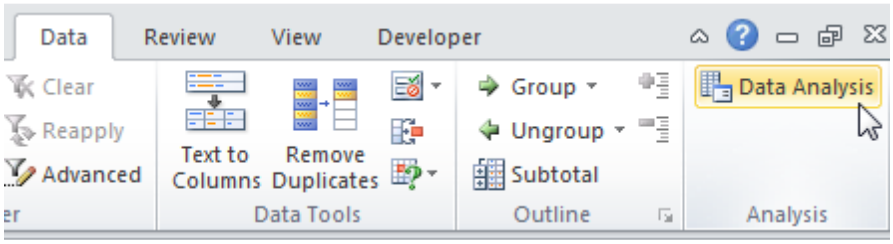
Below you can find the salaries of people who have a degree in economics, medicine or history.

H18

	A	B	C	D
1	economics	medicine	history	
2	42	69	35	
3	53	54	40	
4	49	58	53	
5	53	64	42	
6	43	64	50	
7	44	55	39	
8	45	56	55	
9	52		39	
10	54		40	
11				
12				

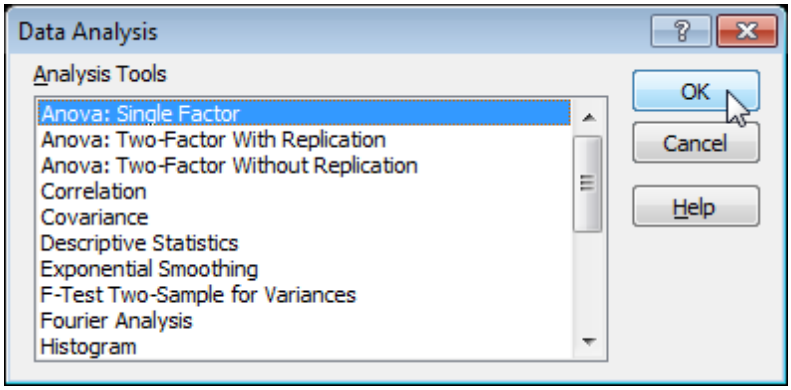
To perform a single factor ANOVA, execute the following steps.

1. On the Data tab, click Data Analysis.



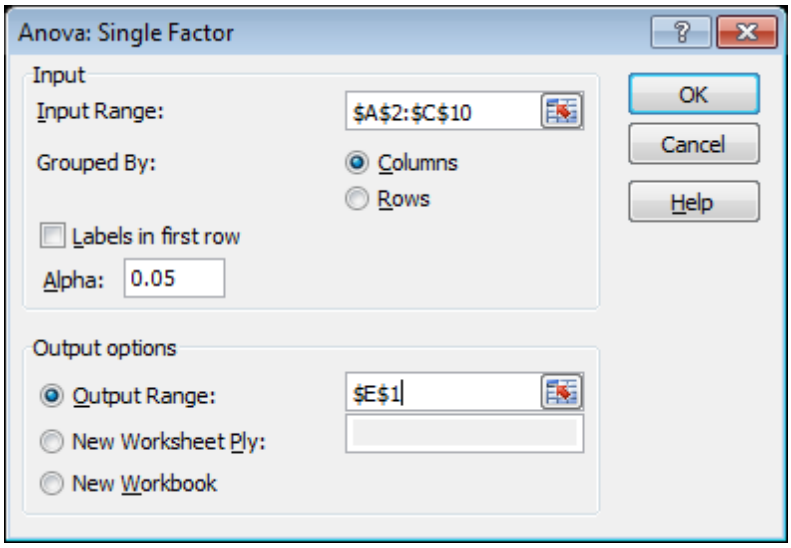
Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

2. Select Anova: Single Factor and click OK.



3. Click in the Input Range box and select the range A2:C10.

4. Click in the Output Range box and select cell E1.



5. Click OK.

Result:

E	F	G	H	I	J	K
Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	9	435	48.33333	23.5		
Column 2	7	420	60	32.33333		
Column 3	9	393	43.66667	50.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1085.84	2	542.92	15.19623	7.16E-05	3.443357
Within Groups	786	22	35.72727			
Total	1871.84	24				

$\alpha = 0.05.$

Hypotheses testing:

$H_0: \mu_1 = \mu_2 = \mu_3$

H_1: at least one of the means is different.

Test statistic:

$$F = \frac{MSA}{MSE} = 15.19623$$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.443357$

$p - value = 0.00007163$

Reject H_0 if $F > 3.443$

Or

$p - value \leq 0.05$

Decision:

Therefore, we reject the null hypothesis. The means of the three populations are not all equal. At least one of the means is different. However, the ANOVA does not tell you where the difference lies. You need a t-Test to test each pair of means.

- Two-Factor Analysis of Variance

$\alpha = 0.05$.

The three hypotheses to be tested are as follows:

Hypotheses testing:

1. H_0: $a_1 = a_2 = \cdots = a_a = 0$

   H_1: At least one of the a_i is not equal to zero.

2. H_0: $\beta_1 = \beta_2 = \cdots = \beta_b = 0$

   H_1: At least one of the β_j is not equal to zero.

3. H_0: $(a\beta)_{11} = (a\beta)_{12} = \cdots = (a\beta)_{ab} = 0$

   H_1: At least one of the $(a\beta)_{ij}$ is not equal to zero.

Critical region:

$p - value \leq \alpha$

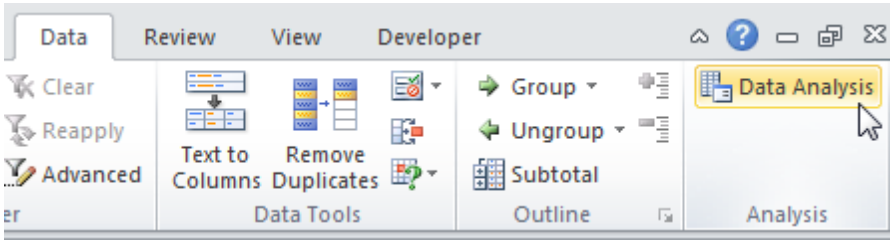
Decision:

- ANOVA two ways with Replication

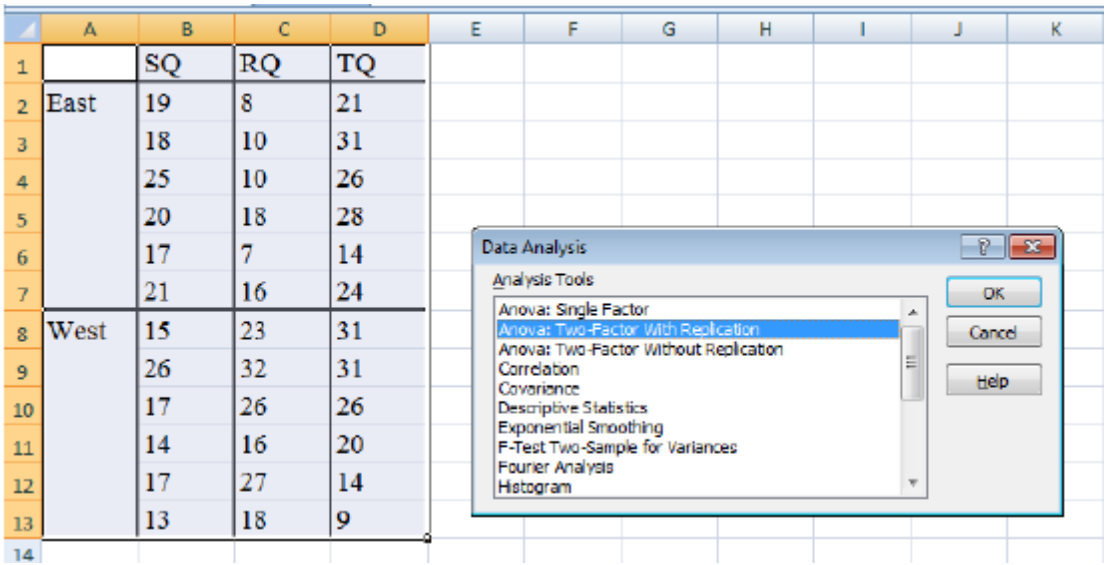
In the following example, the cost of three swimming programs in two regions in the East and the West

	SQ	RQ	TQ
East	19	8	21
	18	10	31
	25	10	26
	20	18	28
	17	7	14
	21	16	24
West	15	23	31
	28	32	31
	17	26	26
	14	16	20
	17	27	14
	13	18	9

1. On the Data tab, click Data Analysis.



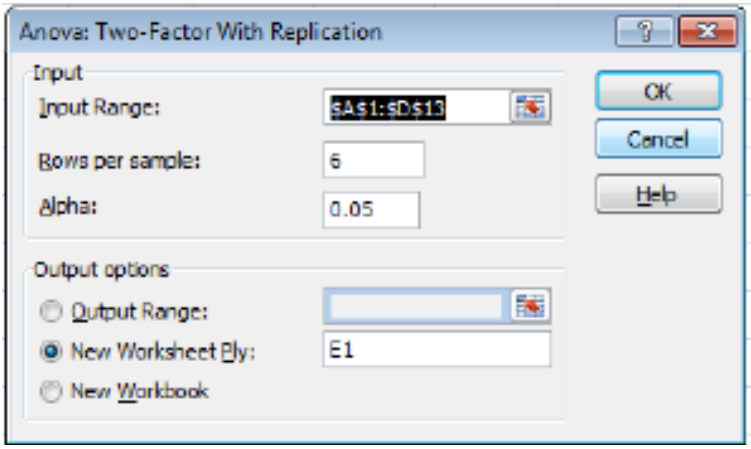
2. Select Anova: Two-Factor With Replication and click OK.



	A	B	C	D
1		SQ	RQ	TQ
2	East	19	8	21
3		18	10	31
4		25	10	26
5		20	18	28
6		17	7	14
7		21	16	24
8	West	15	23	31
9		26	32	31
10		17	26	26
11		14	16	20
12		17	27	14
13		13	18	9

3. Click in the Input Range box and select the range A1:D13.

4. Click in the Output Range box and select cell E2.



5. Click OK.

Result:

Anova: Two-Factor With Replication						
SUMMARY	SQ	RQ	TQ	Total		
East						
Count	6	6	6	18		
Sum	120	69	144	333		
Average	20	11.5	24	18.5		
Variance	8	19.9	35.6	47.44118		
West						
Count	6	6	6	18		
Sum	102	142	131	375		
Average	17	23.66667	21.83333	20.83333		
Variance	22	35.46667	82.96667	49.67647		
Total						
Count	12	12	12			
Sum	222	211	275			
Average	18.5	17.58333	22.91667			
Variance	16.09091	65.53788	55.17424			
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	49	1	49	1.441648	0.239268	4.170877
Columns	195.1667	2	97.58333	2.871036	0.072297	3.31583
Interaction	436.1667	2	218.0833	6.416313	0.004788	3.31583
Within	1019.667	30	33.98889			
Total	1700	35				

$\alpha = 0.05.$

1) Hypotheses testing:

H_0: $a_{East} = a_{West} = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 1.4416$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 4.1708$

$p - value = 0.2392$

Reject H_0 if $F > 4.1708$ Or $p - value \leq 0.05$

Decision:

Therefore, we not reject the null hypothesis. There are no difference in the means of the cost between the two regions in the east and the west.

2) Hypotheses testing:

H_0: $\beta_{SQ} = \beta_{RQ} = \beta_{TQ} = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 2.871$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.3158$

$p - value = 0.0722$

Reject H_0 if $F > 3.3158$ Or $p - value \leq 0.05$

Decision:

Therefore, we not reject the null hypothesis. There was no significant difference between the mean cost of the three programs.

3) Hypotheses testing:

H_0: $a\beta_{East,SQ} = a\beta_{East,RQ} = a\beta_{East,TQ} = a\beta_{West,SQ} = a\beta_{West,RQ} = a\beta_{West,TQ} = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 6.4163$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.3158$

$p - value = 0.0047$

Reject H_0 if $F > 3.3158$ Or $p - value \leq 0.05$

Decision:

Therefore, we reject the null hypothesis. There is interaction between the two factors (region, type of program).

- ANOVA two ways without Replication

In the following example, the price of three swimming programs in six areas

	SQ	RQ	TQ
1	19	8	21
2	18	10	31
3	25	10	26
4	20	18	28
5	17	7	14
6	21	16	24

$\alpha = 0.05$.

The two hypotheses to be tested are as follows:

Hypotheses testing:

1. H_0: $a_1 = a_2 = \cdots = a_a = 0$

   H_1: At least one of the a_i is not equal to zero.

2. H_0: $\beta_1 = \beta_2 = \cdots = \beta_b = 0$

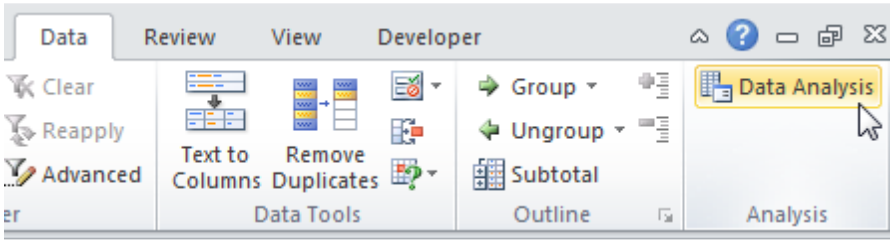
   H_1: At least one of the β_j is not equal to zero.

Critical region:

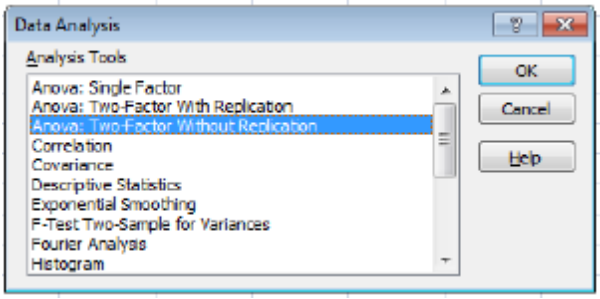
$p - value \leq \alpha$

Decision:

1. On the Data tab, click Data Analysis.

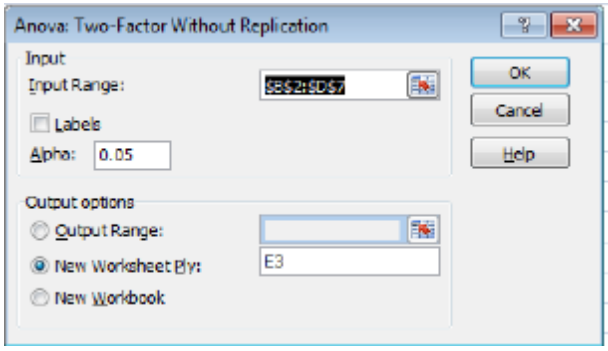


2. Select Anova: Two-Factor With Replication and click OK.



3. Click in the Input Range box and select the range B2:D7.

4. Click in the Output Range box and select cell E3.



5. Click OK.

Result:

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
Row 1	3	48	16	49
Row 2	3	59	19.66667	112.3333
Row 3	3	61	20.33333	80.33333
Row 4	3	66	22	28
Row 5	3	38	12.66667	26.33333
Row 6	3	61	20.33333	16.33333
Column 1	6	120	20	8
Column 2	6	69	11.5	19.9
Column 3	6	144	24	35.6

ANOVA

ce of Varic	SS	df	MS	F	P-value	F crit
Rows	181.8333	5	36.36667	2.68059	0.086618	3.325835
Columns	489	2	244.5	18.02211	0.000483	4.102821
Error	135.6667	10	13.56667			
Total	806.5	17				

$\alpha = 0.05.$

1) Hypotheses testing:

H_0: $a_1 = a_2 = \cdots = a_6 = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 2.6805$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.3258$

$p - value = 0.0866$

Reject H_0 if $F > 3.3258$ Or $p - value \leq 0.05$

Decision:

Therefore, we not reject the null hypothesis. There are no difference in the means of the cost between the six areas.

2) Hypotheses testing:

H_0: $\beta_{SQ} = \beta_{RQ} = \beta_{TQ} = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 18.022$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 4.1028$

$p - value = 0.0004$

Reject H_0 if $F > 4.1028$ Or $p - value \leq 0.05$

Decision:

Therefore, we reject the null hypothesis. There is a significant difference between the mean prices of the three programs.

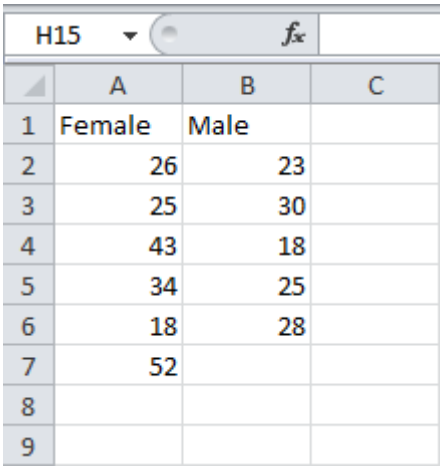
3- F-Test

This example teaches you how to perform an F-Test in Excel. The F-Test is used to test the null hypothesis that the variances of two populations are equal.

Below you can find the study hours of 6 female students and 5 male students.

H0: $\sigma 12 = \sigma 22$

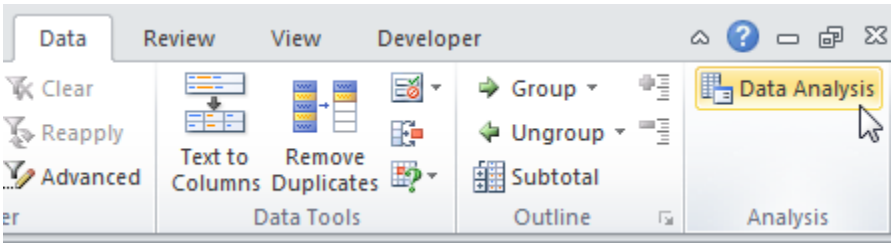
H1: $\sigma 12 \neq \sigma 22$



	A	B	C
1	Female	Male	
2	26	23	
3	25	30	
4	43	18	
5	34	25	
6	18	28	
7	52		
8			
9			

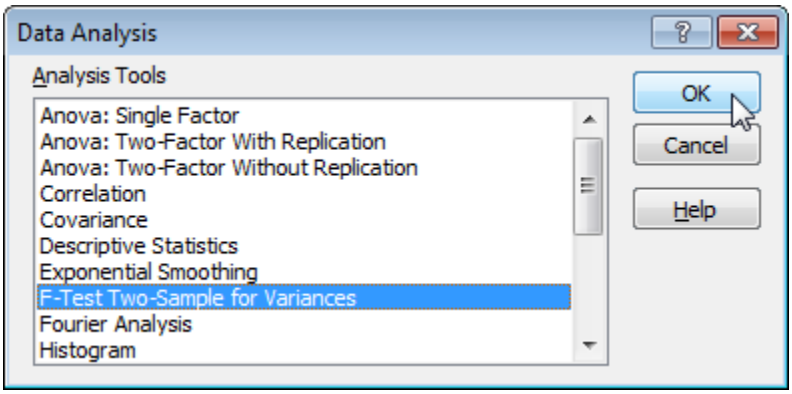
To perform an F-Test, execute the following steps.

1. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

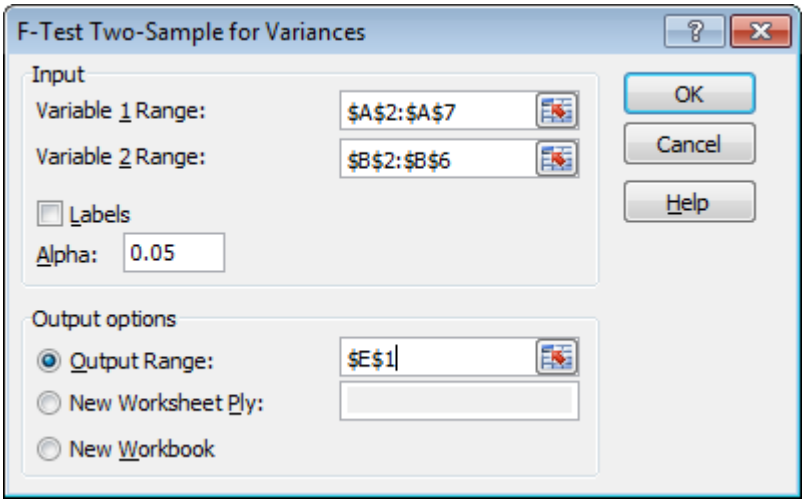
2. Select F-Test Two-Sample for Variances and click OK.



3. Click in the Variable 1 Range box and select the range A2:A7.

4. Click in the Variable 2 Range box and select the range B2:B6.

5. Click in the Output Range box and select cell E1.



6. Click OK.

Result:

E	F	G
F-Test Two-Sample for Variances		
	Variable 1	*Variable 2*
Mean	33	24.8
Variance	160	21.7
Observations	6	5
df	5	4
F	7.373271889	
P(F<=f) one-tail	0.037888376	
F Critical one-tail	6.256056502	

Important: be sure that the variance of Variable 1 is higher than the variance of Variable 2. This is the case, 160 > 21.7. If not, swap your data. As a result, Excel calculates the correct F value, which is the ratio of Variance 1 to Variance 2 (F = 160 / 21.7 = 7.373).

Conclusion: if F > F Critical one-tail, we reject the null hypothesis. This is the case, 7.373 > 6.256. Therefore, we reject the null hypothesis. The variances of the two populations are unequal.

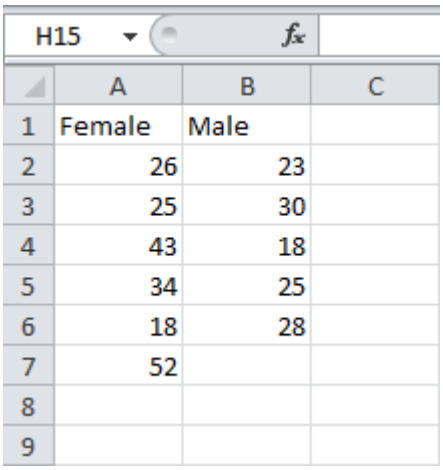
4- t-Test

This example teaches you how to perform a t-Test in Excel. The t-Test is used to test the null hypothesis that the means of two populations are equal.

Below you can find the study hours of 6 female students and 5 male students.

H0: $\mu 1 - \mu 2 = 0$

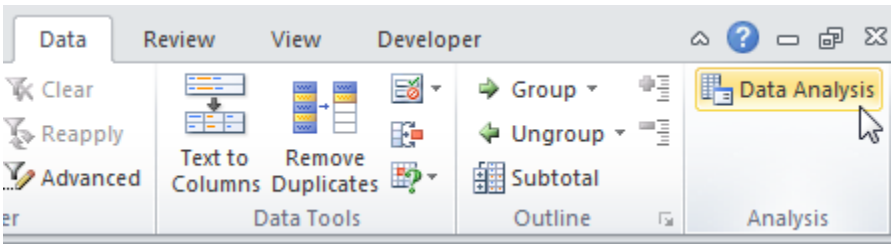
H1: $\mu 1 - \mu 2 \neq 0$



To perform a t-Test, execute the following steps.

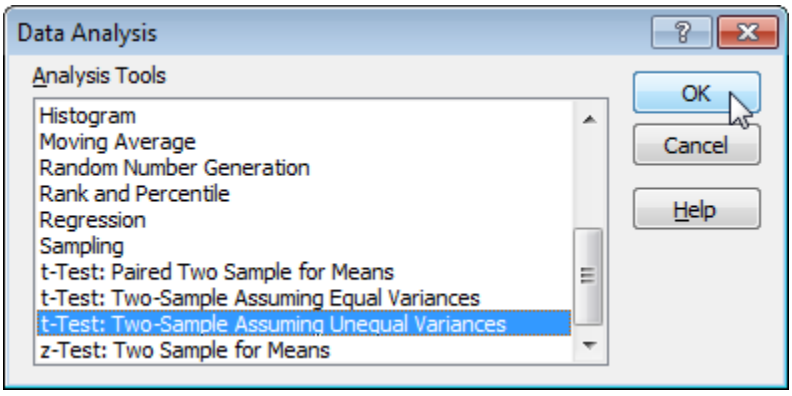
1. First, perform an F-Test to determine if the variances of the two populations are equal. This is not the case.

2. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

3. Select t-Test: Two-Sample Assuming Unequal Variances and click OK.

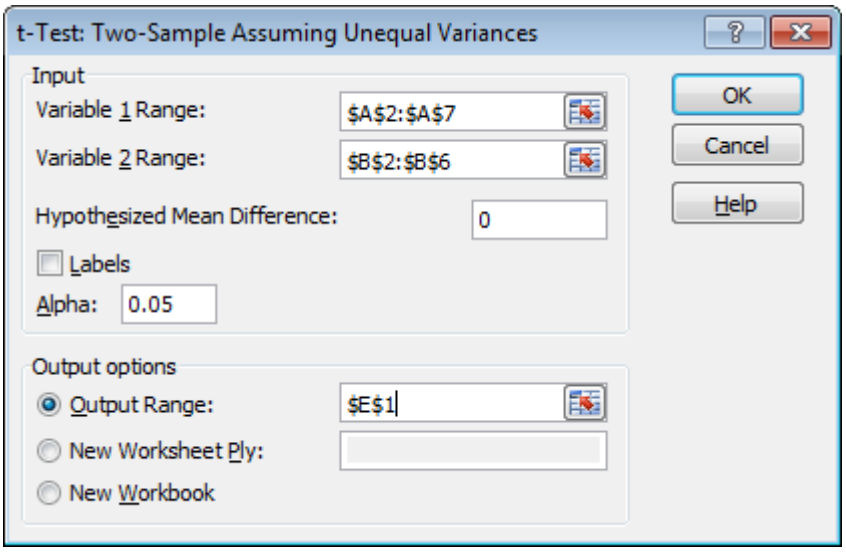


4. Click in the Variable 1 Range box and select the range A2:A7.

5. Click in the Variable 2 Range box and select the range B2:B6.

6. Click in the Hypothesized Mean Difference box and type 0 (H0: μ1 - μ2 = 0).

7. Click in the Output Range box and select cell E1.



8. Click OK.

Result:

E	F	G
t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	33	24.8
Variance	160	21.7
Observations	6	5
Hypothesized Mean Difference	0	
df	7	
t Stat	1.47260514	
P(T<=t) one-tail	0.092170202	
t Critical one-tail	1.894578605	
P(T<=t) two-tail	0.184340405	
t Critical two-tail	2.364624252	

Conclusion:

We do a two-tail test (inequality). lf t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis. This is not the case, -2.365 < 1.473 < 2.365. Therefore, we do not reject the null hypothesis. The observed difference between the sample means (33 - 24.8) is not convincing enough to say that the average number of study hours between female and male students differ significantly.

5- Paired Sample t Test

In paired sample hypothesis testing, a sample from the population is chosen and two measurements for each element in the sample are taken. Each set of measurements is considered a sample. Unlike the hypothesis testing studied so far, the two samples are not independent of one another. Paired samples are also called matched samples or repeated measures.

Example:

A clinic provides a program to help their clients lose weight and asks a consumer agency to investigate the effectiveness of the program. The agency takes a sample of 15 people, weighing each person in the sample before the program begins and 3 months later to produce the table below.

	A	B	C	D
1	**Two sample t test with paired samples**			
2				
3	Person	Before	After	Difference
4	1	210	197	13
5	2	205	195	10
6	3	193	191	2
7	4	182	174	8
8	5	259	236	23
9	6	239	226	13
10	7	164	157	7
11	8	197	196	1
12	9	222	201	21
13	10	211	196	15
14	11	187	181	6
15	12	175	164	11
16	13	186	181	5
17	14	243	229	14
18	15	246	231	15
19				
20	mean			10.933333
21	std dev			6.3298236

Determine whether the program is effective?

Solution

Let x = the difference in weight 3 months after the program starts. The null hypothesis is:

H0: $\mu = 0$; i.e. any differences in weight is due to chance

We use the Excel's t-Test: Paired Two Sample for Means data analysis tool or the T Test. The output from the Excel data analysis tool is shown below:

t-Test: Paired Two Sample for Means		
	Before	*After*
Mean	207.9333	197
Variance	815.781	595
Observations	15	15
Pearson Correlation	0.98372	
Hypothesized Mean Difference	0	
df	14	
t Stat	6.6897	
P(T<=t) one-tail	5.14E-06	
t Critical one-tail	1.76131	
P(T<=t) two-tail	1.03E-05	
t Critical two-tail	2.144787	

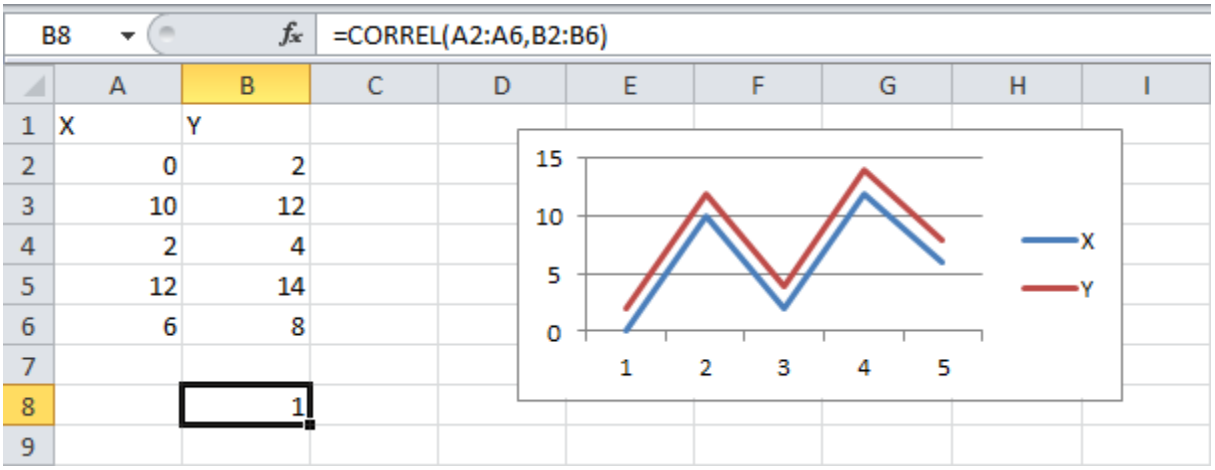
Note that the mean differences are the same, but the standard deviation for the paired sample case is lower, which results in a higher t-stat and a lower p-value. This is generally true.

Observation: Since the two sample paired data case is equivalent to the one sample case.

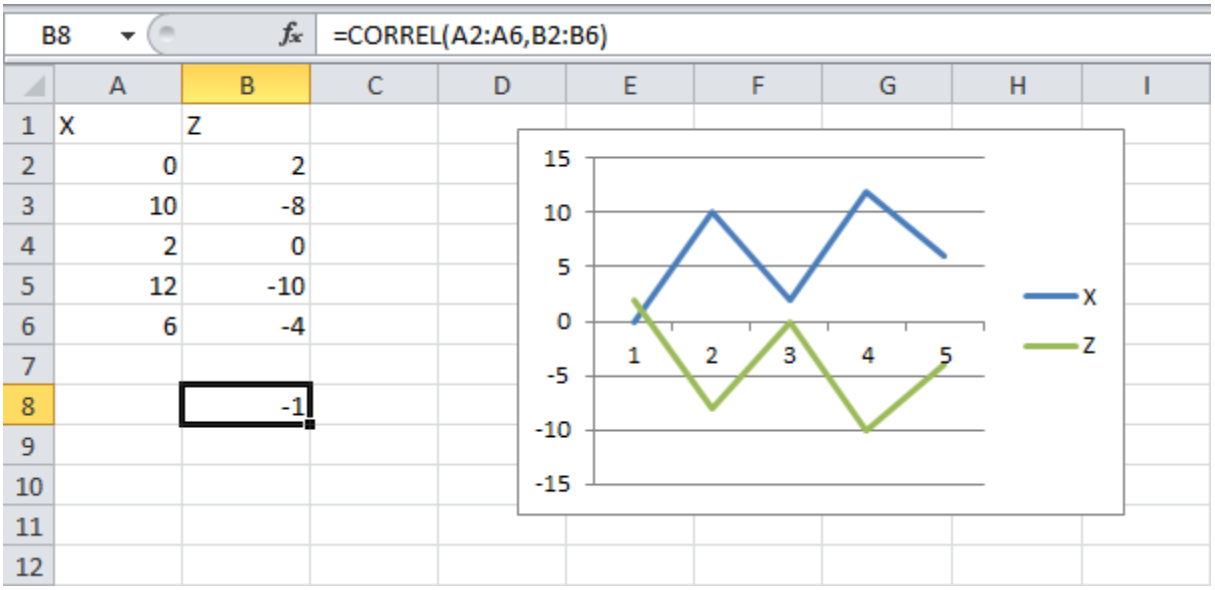
6- Correlation

The correlation coefficient (a value between -1 and +1) tells you how strongly two variables are related to each other. We can use the CORREL function or the Analysis Toolpak add-in in Excel to find the correlation coefficient between two variables.

- A correlation coefficient of +1 indicates a perfect positive correlation. As variable X increases, variable Y increases. As variable X decreases, variable Y decreases.



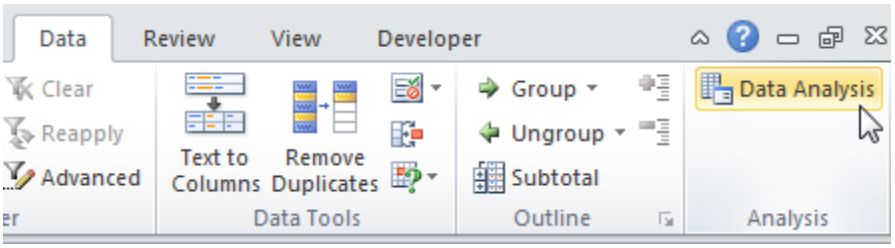
- A correlation coefficient of -1 indicates a perfect negative correlation. As variable X increases, variable Z decreases. As variable X decreases, variable Z increases.



- A correlation coefficient near 0 indicates no correlation.

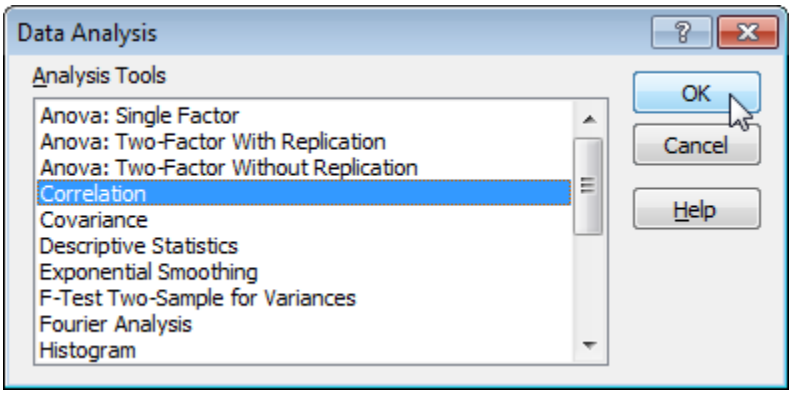
To use the Analysis Toolpak add-in in Excel to quickly generate correlation coefficients between multiple variables, execute the following steps.

1. On the Data tab, click Data Analysis.

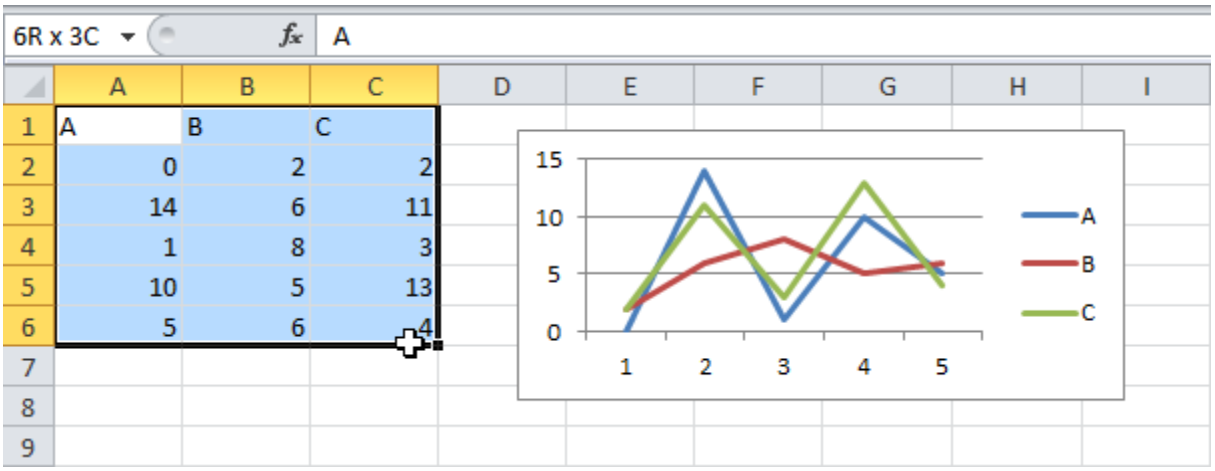


Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

2. Select Correlation and click OK.



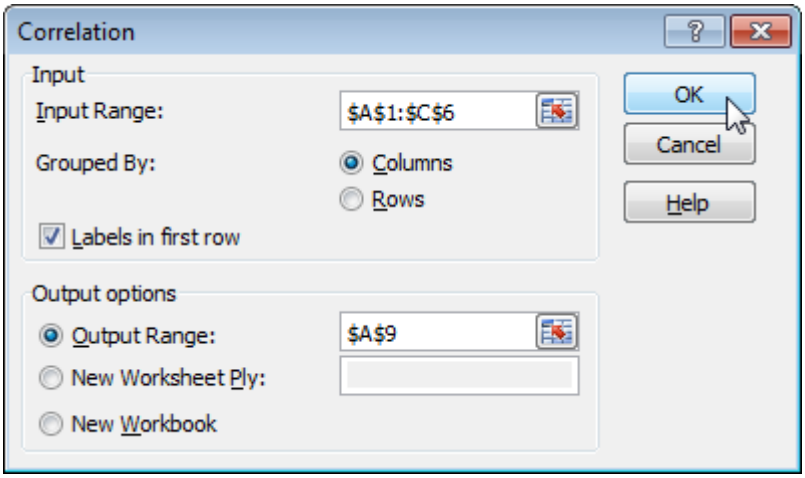
3. For example, select the range A1:C6 as the Input Range.



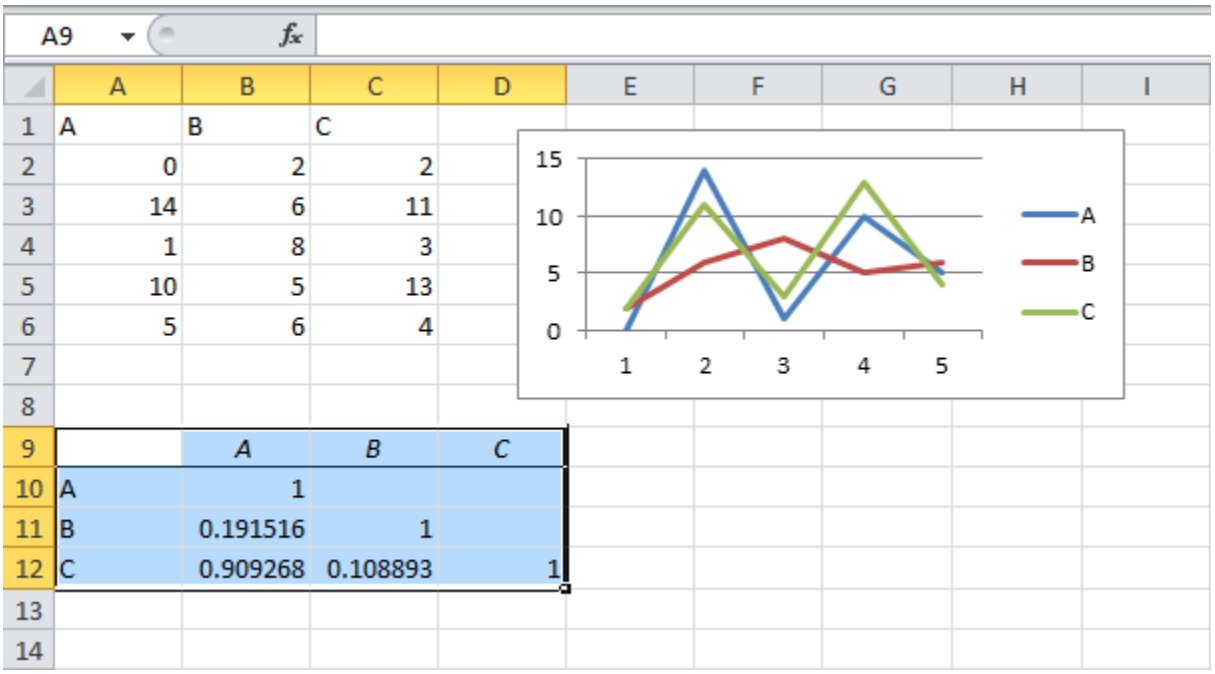
4. Check Labels in first row.

5. Select cell A9 as the Output Range.

6. Click OK.



Result.



Conclusion: variables A and C are positively correlated (0.91). Variables A and B are not correlated (0.19). Variables B and C are also not correlated (0.11) . You can verify these conclusions by looking at the graph.

7-Regression

This example teaches you how to perform a regression analysis in Excel and how to interpret the Summary Output.

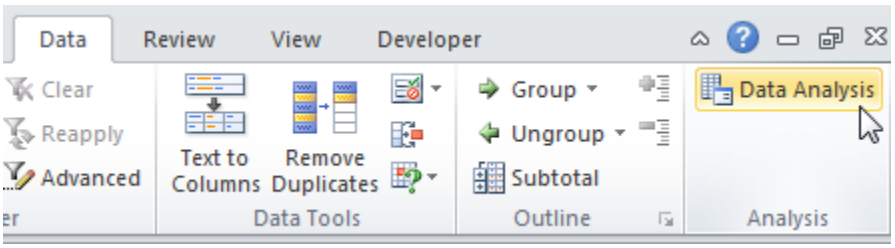
In this study a random sample of service call records for a computer repair operation were examined and the length of each call (in minutes) and the number of components repaired or replaced were recorded. The data is given below.

Minutes	Units
23	1
29	2
49	3
64	4
74	4
87	5
96	6
97	6
109	7
119	8
149	9
145	9
154	10
166	10

Below you can find our data. The big question is:

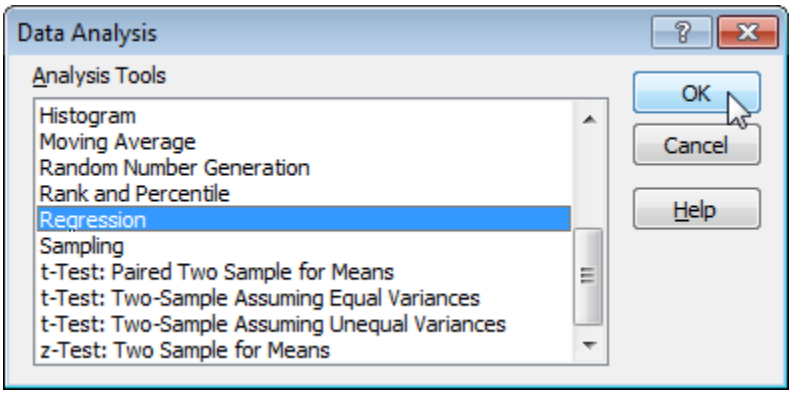
is there a relation between Minutes (Output) and Units (Input). In other words: can we predict Minutes if we know the unites?

1. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

2. Select Regression and click OK.



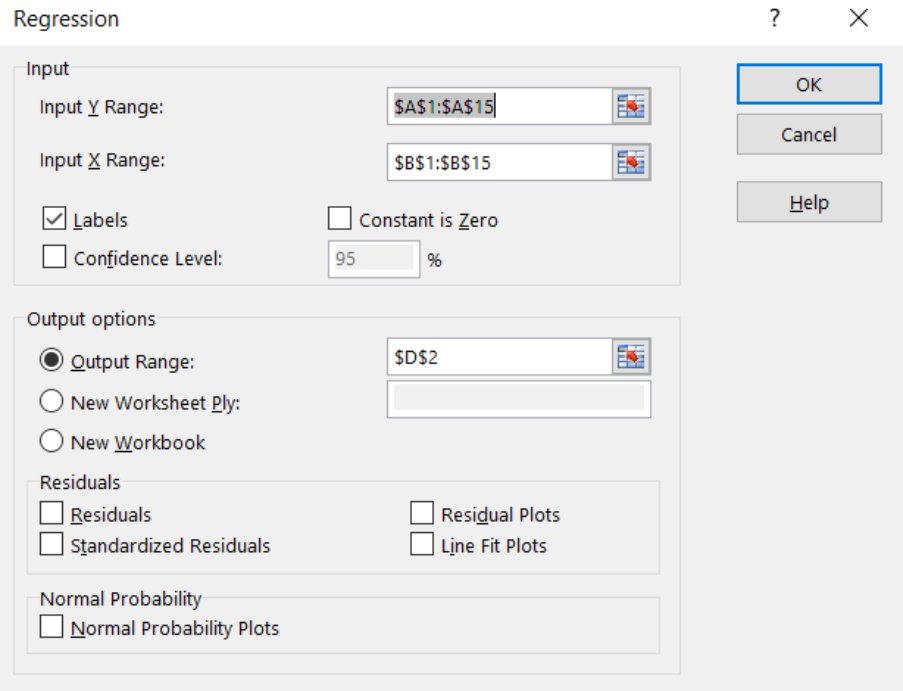
3. Select the Y Range (A2:A15). This is the predictor variable (also called dependent variable).

4. Select the X Range(B2:B15). These are the explanatory variables (also called independent variables). These columns must be adjacent to each other.

5. Check Labels.

6. Select an Output Range.

8. Click OK.



Excel produces the following Summary Output (rounded to 3 decimal places).

Minutes	Units
23	1
29	2
49	3
64	4
74	4
87	5
96	6
97	6
109	7
119	8
149	9
145	9
154	10
166	10

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.994
R Square	0.987
Adjusted R	0.986
Standard E	5.392
Observatic	14.000

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressior	1.000	27419.509	27419.509	943.201	0.000
Residual	12.000	348.848	29.071		
Total	13.000	27768.357			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4.162	3.355	1.240	0.239	-3.148	11.472	-3.148	11.472
Units	15.509	0.505	30.712	0.000	14.409	16.609	14.409	16.609

Results

The regression line is: y = Minutes=4.162+15.509*Units. In other words, for increasing the units by one, the Time Minutes increases by 15.509, while there is 4.162 minutes does not depend on the unites.

Part 3

Some Statistical Charts in Excel

Excel provides fairly extensive capabilities for creating graphs, what Excel calls **charts**. You can access Excel's charting capabilities by selecting **Insert > Charts**. We will describe how to create bar and line charts here. Elsewhere on the website we describe how to create scatter charts. Other types of charts are created in a similar manner. Once a chart is created three new ribbons are accessible, namely **Design**, **Layout** and **Format**. These are used to refine the chart created.

1- Bar charts

To create a bar chart, execute the following steps:

1. Enter the data that you are charting into a worksheet.
2. Highlight the data range and select **Insert > Charts|Column**. A list of bar chart types is displayed. As usual you can place the mouse pointer over the picture of any chart type to get a brief description of that chart type. E.g. the first type is a 2-dimensional side-by-side bar chart while second choice is a 2-dimensional stacked bar chart.
3. Use the **Design**, **Layout** and **Format** ribbons to refine the chart. At any time, you can click on the chart to get access to these ribbons.

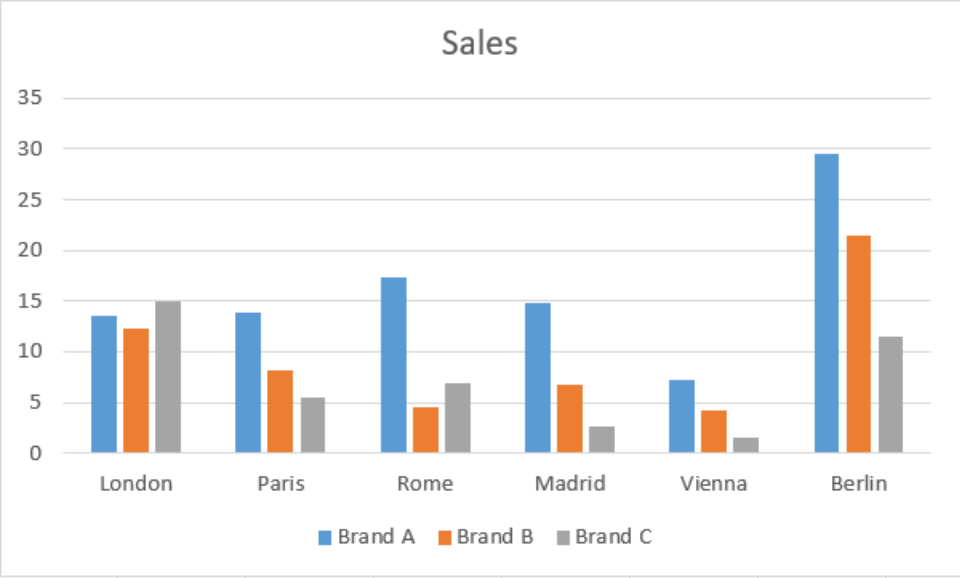
We now demonstrate how to create a bar chart via the following example.

Example

The following data represents the sales in millions of Euros in for a certain product has three brands A, B and C

City	Brand A	Brand B	Brand C
London	13.5	12.3	15
Paris	13.8	8.1	5.5
Rome	17.3	4.5	6.9
Madrid	14.8	6.8	2.7
Vienna	7.2	4.2	1.6
Berlin	29.5	21.4	11.5

The first step is to enter the data into the worksheet. We next highlight the range the data including the row and column headings, and select **Insert > Charts|Column**.

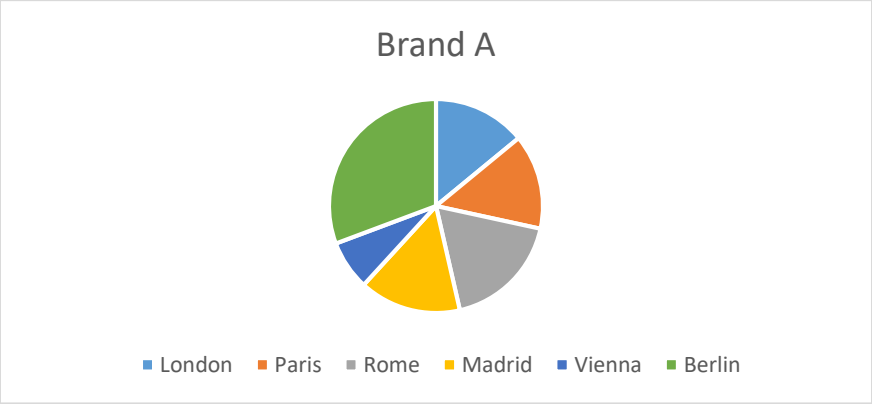


The resulting chart, although initially the chart does not contain a chart title or axes titles. To add a chart title click on the chart, select Layout > Labels|Chart Title and then choose Above Chart and enter the title sales. The title of the horizontal axis can be added in a similar manner by selecting Layout > Labels|Axis Titles > Primary Horizontal Axis Title > Title Below Axis and entering the word City. Finally, the

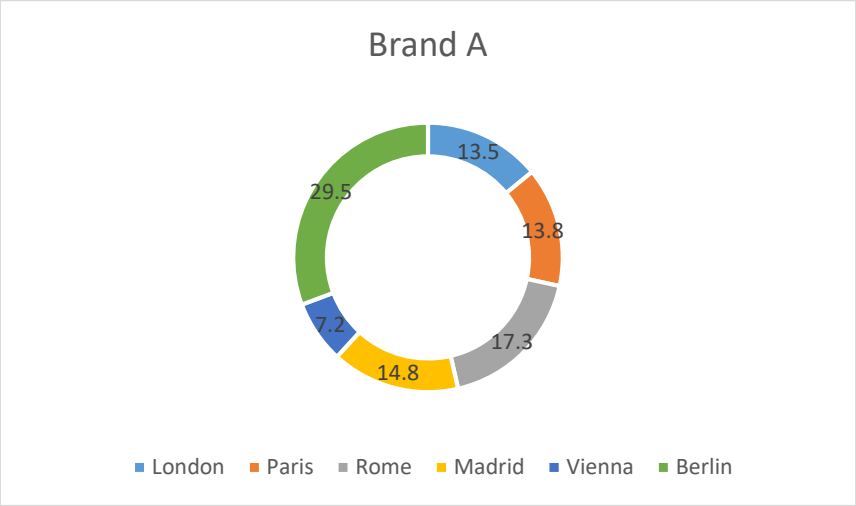
title of the vertical axis is added by selecting Layout > Labels|Axis Titles > Primary Vertical Axis Title > Rotated Title.

2-Pie charts

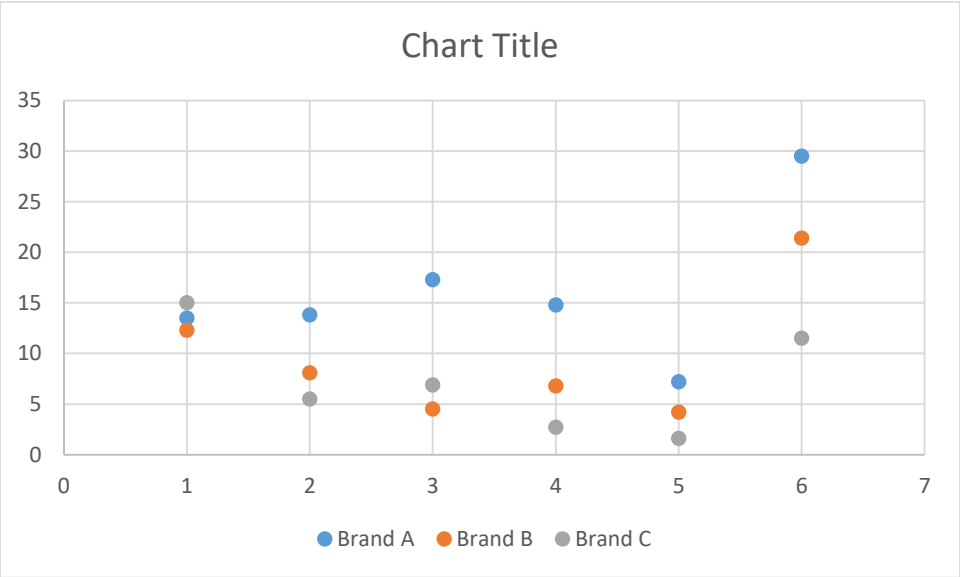
For the previous example draw the pie chart for brand A



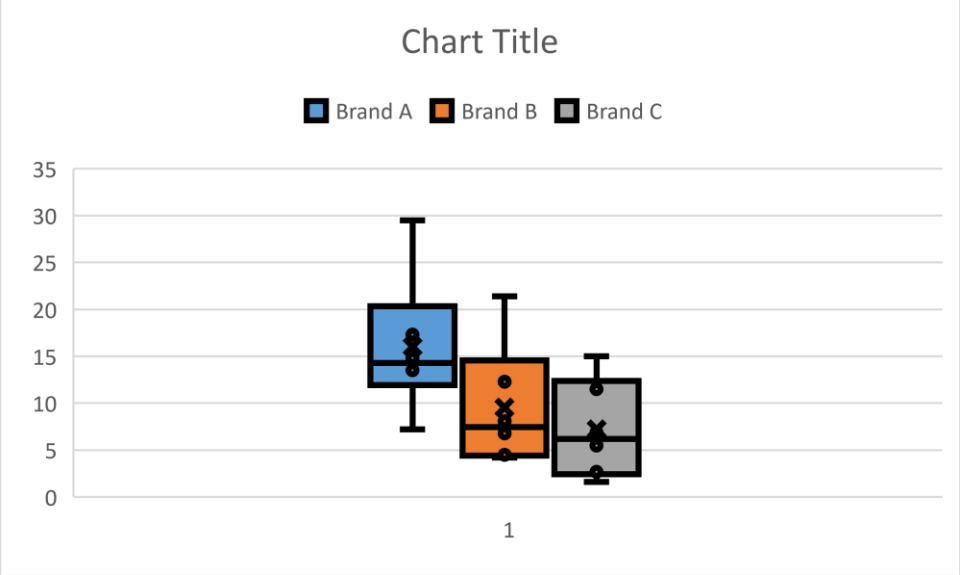
Also,



3-scatter plot



4-Boxplot



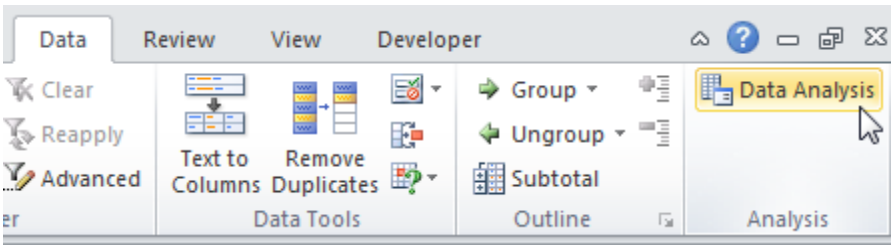
4-Histogram

We explain how to create the Histogram in Excel through the following example given in the file Histogram.xls. This example teaches you how to create a histogram in Excel.

1. First, enter the bin numbers (upper levels) in the range C3:C7.

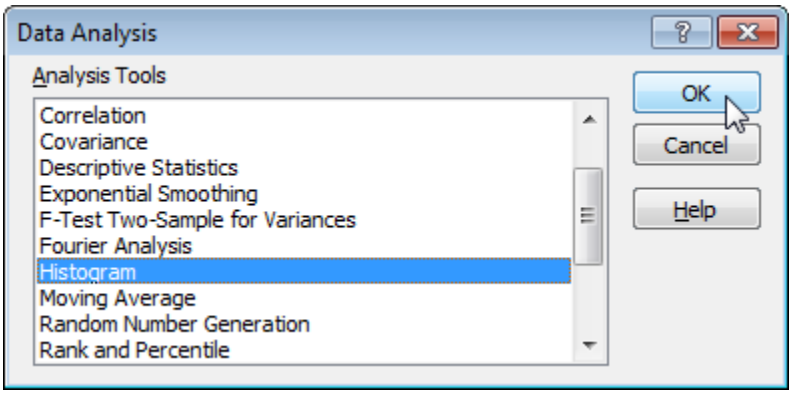
	A	B	C	D
1	Number of students			
2	22			
3	29		20	
4	40		25	
5	30		30	
6	48		35	
7	24		40	
8	21			
9	19			
10	24			
11	22			
12	25			
13	52			
14	35			
15	40			
16	31			
17	37			
18	21			
19	23			
20				
21				

2. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

3. Select Histogram and click OK.

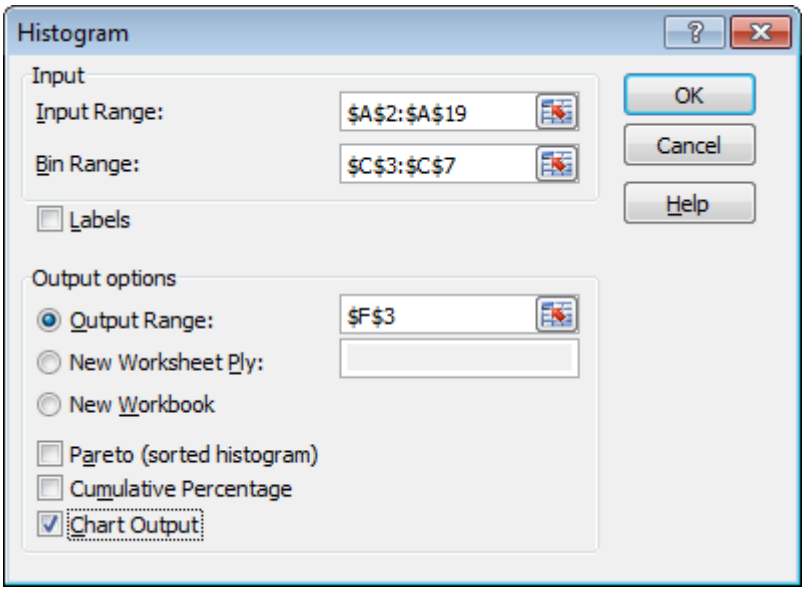


4. Select the range A2:A19.

5. Click in the Bin Range box and select the range C3:C7.

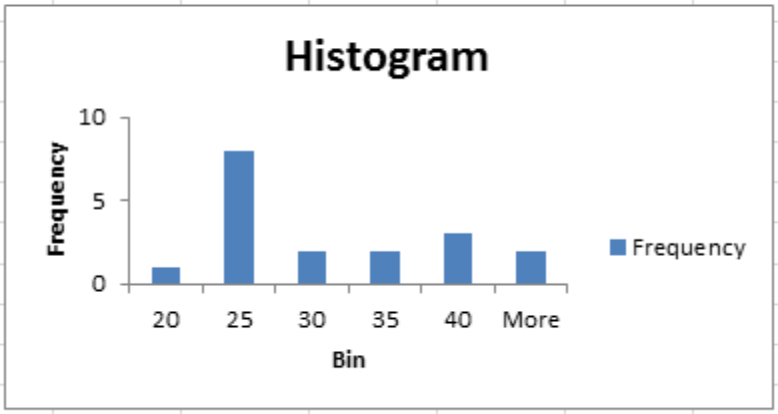
6. Click the Output Range option button, click in the Output Range box and select cell F3.

7. Check Chart Output.



8. Click OK.

Bin	Frequency
20	1
25	8
30	2
35	2
40	3
More	2



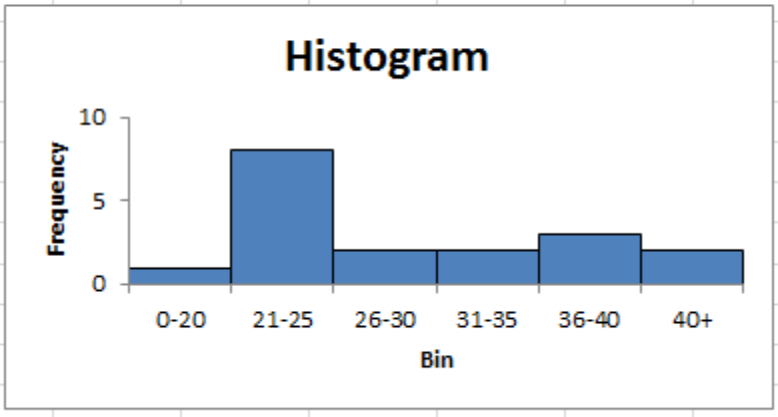
9. Click the legend on the right side and press Delete.

10. Properly label your bins.

11. To remove the space between the bars, right click a bar, select Format Data Series and change the Gap Width to 0%. Select Border Color to add a border.

Result:

Bin	Frequency
0-20	1
21-25	8
26-30	2
31-35	2
36-40	3
40+	2



Working With Matrices In Excel

To work with matrices in excel, especially with the multiplications and inverse, you have to follow the following steps

1- Select suitable (m x n) block anywhere else in your spreadsheet to receive the expected results on them.

2- Press the “F2” key. • Press the “=” sign

3- Type the matrix function with its arguments.

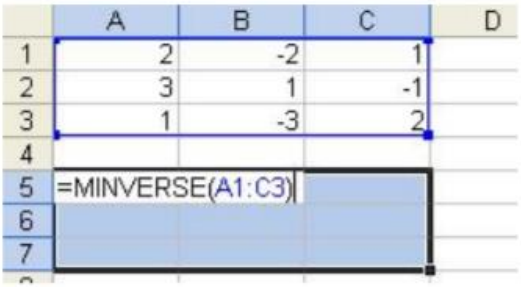
4- Hold down the keys “Ctrl” and “Shift” at the same time and press the “Enter” key while holding down the other two.

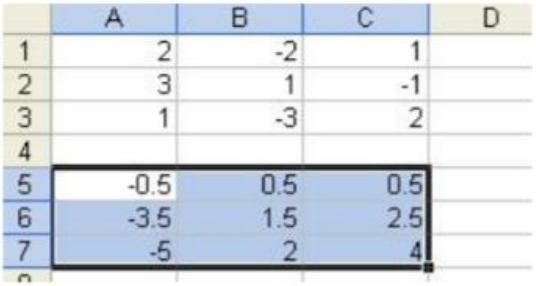
Example:

Find the inverse of

$$A = \begin{pmatrix} 2 & -2 & 1 \\ 3 & 1 & -1 \\ 1 & -3 & 2 \end{pmatrix}$$

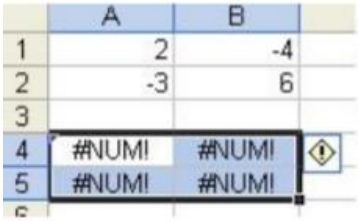
A^{-1} should now appear in the block of cells you had selected (see screen shot on the right below).





If something weird appears (like only one number appearing), you may have pressed the **Enter** key without holding down the **Ctrl** and **Shift** keys. If this is the case re-select the cells you had chosen for your A^{-1} and press the **Delete** key, then repeat the above steps.

If A^{-1} does not exist, e.g. the matrix $B = \begin{pmatrix} 2 & -4 \\ -3 & 6 \end{pmatrix}$ has no inverse since $\det(B) = 0$, you make get something that looks like this:



Example

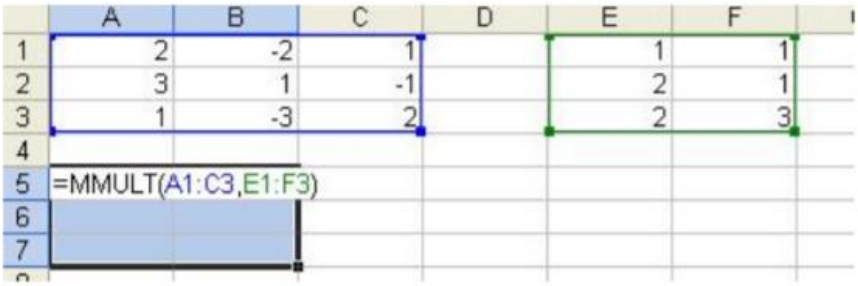
Example:

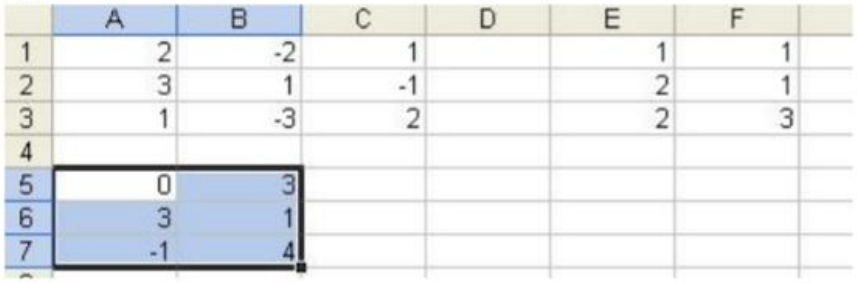
Find the A.C, such that

$$A = \begin{pmatrix} 2 & -2 & 1 \\ 3 & 1 & -1 \\ 1 & -3 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 2 & 3 \end{pmatrix}$$

Solution

As an illustration, let A be as given above and $C = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 2 & 3 \end{pmatrix}$. Since A is a 3×3 matrix and C is a 3×2 matrix, the product AC is a 3×2 matrix. Performing the above steps will give the following screen shots:





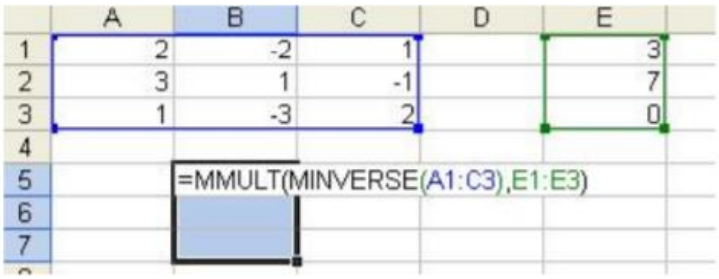
We are now ready to find the solution to our system of equations.

Solving the system of Equations

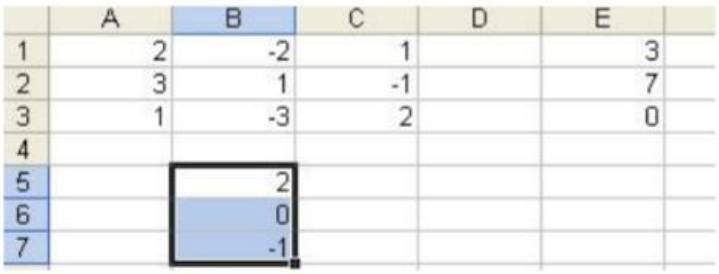
Enter the right-hand-side column vector $\mathbf{b} = \begin{pmatrix} 3 \\ 7 \\ 0 \end{pmatrix}$ into a 3×1 block of cells in your spreadsheet. Also enter the matrix A into a 3×3 block of cells in your spreadsheet.

Since we will get the solution from the matrix product $A^{-1}\mathbf{b}$, we first determine the order of the resultant matrix. In this case, we are multiplying a 3×3 matrix by a 3×1 vector, we will get a 3×1 vector.

Select a 3×1 block of empty cells. Type the following: **F2 = MMULT(MINVERSE(** and select the cells containing matrix A. Close bracket **)**. Type a comma and select the cells containing the vector **b** and close bracket **)**. You should have something that looks like this:



Hold down the **Ctrl** and **Shift** keys and press the **Enter** key while still holding the other two keys. The values of x, y and z should now appear (see screen shot below).



Note that we did not have to calculate A^{-1} separately. We simply asked Excel to calculate it for us and then multiply the inverse by **b**. Most Excel functions can be nested in this way.

Matrix determinant

Example 1 - 2x2 Matrix

The spreadsheet on the right shows a simple 2x2 matrix. The determinant of this matrix can be calculated using the Excel Mdeterm function as follows:

	A	B
1	5	2
2	7	1

=MDETERM(A1:B2)

This gives the result **-9**.

Example 2 - 3x3 Matrix

The determinant of the simple 3x3 matrix on the right can be calculated using the following Excel Mdeterm function:

	A	B	C
1	6	4	2
2	3	5	3
3	2	3	4

=MDETERM(A1:C3)

This gives the result **40**.

Generation Random samples

- Generation from uniform distribution between 0 and 1

  To generate a random sample between 0 and 1 follow the steps below

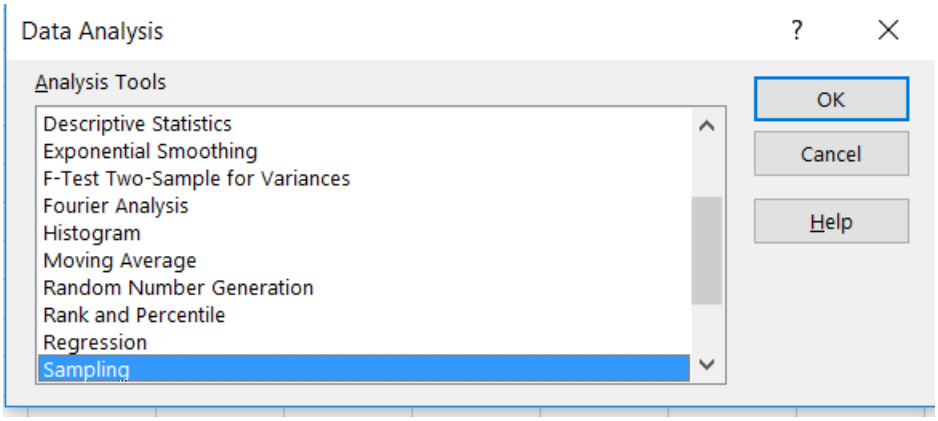
1- Highlight number the cells (in column A, for example) equal the sample size.

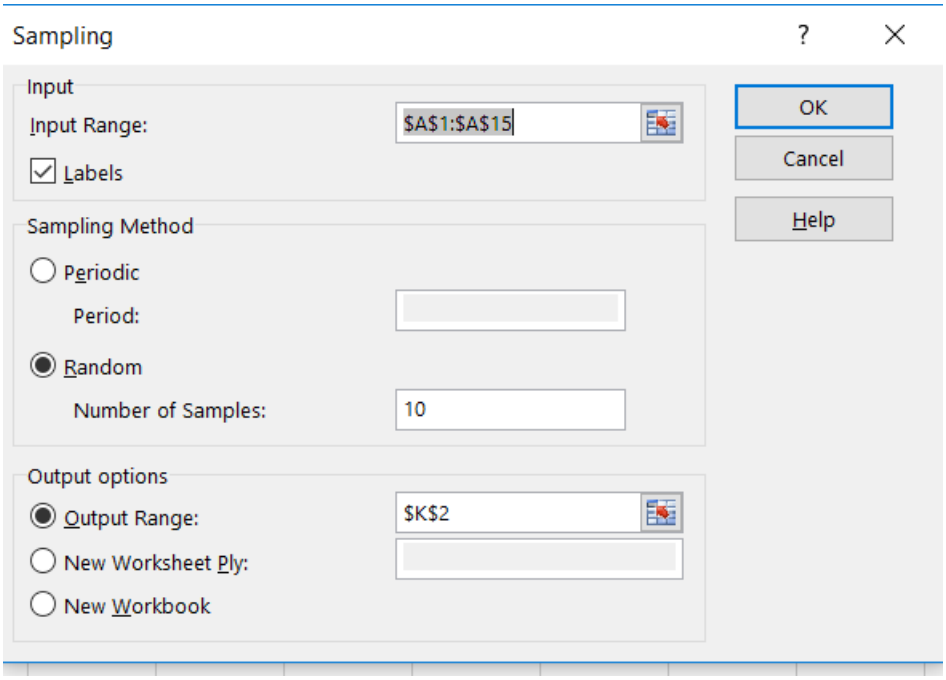
2- Type "=RAND()" without the quotes into the Formula bar, and then press "Ctrl-Enter" to assign random numbers to your selection.

- Random between two numbers

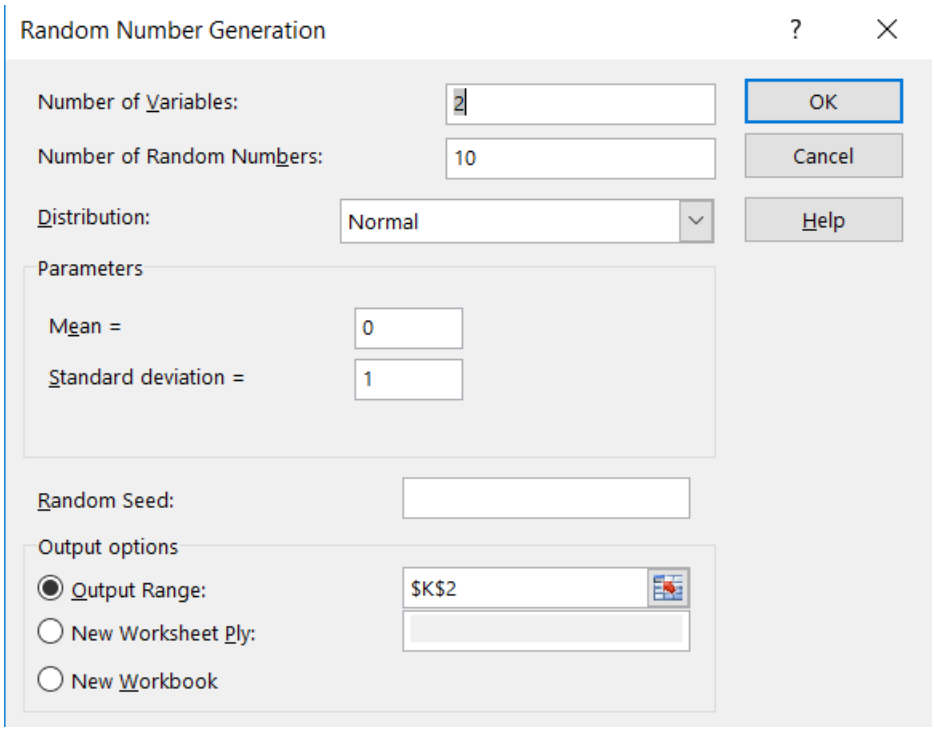
  =RANDBETWEEN(a,b)

- Random sample

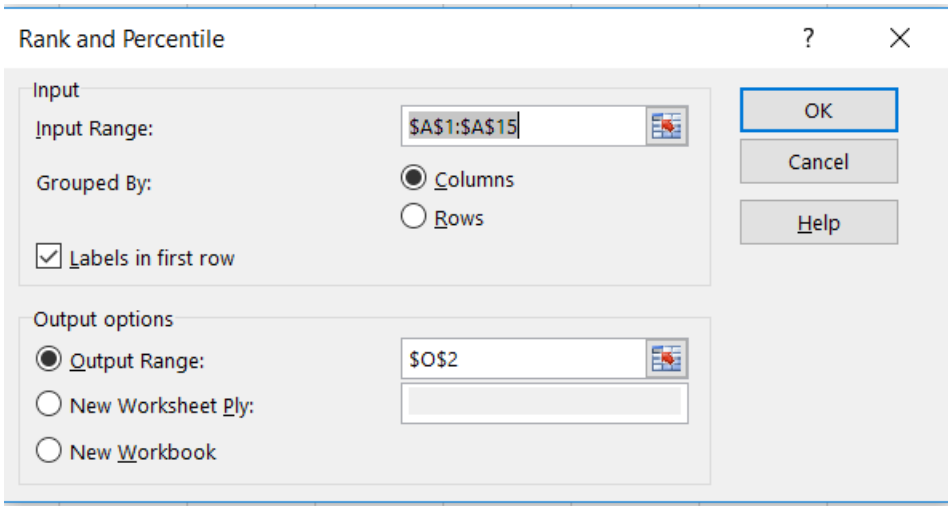




- Random number generation from distributions



- Rank and percentile



x		*Point*	*x*	*Rank*	*Percent*
5		9	9	1	100.00%
7		6	8	2	92.30%
4		2	7	3	84.60%
5		14	6	4	76.90%
5		1	5	5	53.80%
8		4	5	5	53.80%
2		5	5	5	53.80%
2		3	4	8	30.70%
9		10	4	8	30.70%
4		11	4	8	30.70%
4		13	3	11	23.00%
2		7	2	12	0.00%
3		8	2	12	0.00%
6		12	2	12	0.00%