

Handy EXCEL statistical functions

All EXCEL functions begin with an equal sign and all EXCEL descriptive statistical functions have at least one range of cells as its argument.

Example: Numerical data is in cells C3:C209. This is the *cell range*.

=average(C3:C209) will compute the average of all the numbers in C3:C209. The *cell range* in this example is C3:C209. The colon is used to denote the cell range.

Univariate descriptive

=average(<i>cell range</i>)	computes the arithmetic mean
=geomean(<i>cell range</i>)	computes the geometric mean -antilog of log data mean
=median(<i>cell range</i>)	computes the median
=stdev(<i>cell range</i>)	computes the standard deviation
=count(<i>cell range</i>)	computes the sample size (n)
=quartile(<i>cell range</i> , <i>q</i>)	computes 1 st , 2 nd , 3 rd or 4 th quartile (q=1,2,3 or 4). will not work for q > 4.
=percentile(<i>cell range</i> , <i>p</i>)	Gives the 100 ^p th percentile for the data in cell range $0 \leq p \leq 1$
=max(<i>cell range</i>)	gives the maximum in the cell range
=min(<i>cell range</i>)	gives the minimum in the cell range

Bivariate descriptive

=correl(<i>y cell range</i> , <i>x cell range</i>)	computes the Pearson correlation (r) between y & x
=slope(<i>y cell range</i> , <i>x cell range</i>)	computes the slope (b) for $y = a + b X$
=intercept(<i>y cell range</i> , <i>x cell range</i>)	computes the intercept (a) for $y = a + b X$

Gaussian

=normsdist(<i>cell</i>)	gives the percentile (as a decimal) for a Z score. <i>cell</i> =Z score Ex: normsdist(1.96)= 0.975.
=normsinv(<i>cell</i>)	gives the Z score for a percentile. <i>Cell</i> =percentile (as a decimal) Ex: normsinv(0.975) = 1.96

=normdist(*y*, *mean*, *SD*, *cumulative indicator*) If *cumulative indicator* equals 1, gives the percentile (as a decimal) for y where y has a Gaussian distribution with *mean* and *SD*. If *cumulative indicator* equals 0, gives the Gaussian density for y, f(y).

EXCEL mathematics functions

=log10(<i>cell</i>)	computes the base 10 log value of <i>cell</i>
=ln(<i>cell</i>)	computes the base e log value of <i>cell</i>
=exp(<i>cell</i>)	computes the antilog (base e) of <i>cell</i> (e^{cell})
=10^ <i>cell</i>	computes the antilog (base 10) of <i>cell</i> (10^{cell})
=sqrt(<i>cell</i>)	computes the square root of <i>cell</i>

There are many other functions in EXCEL – this is only a small sample.

functions that produce p values

=TDIST(*t*, *df*, 2) – two sided p value for *t* statistic where *df* is the degrees of freedom

=CHIDIST(χ^2 , *df*) Chi-Square Distribution, where χ^2 is the chi square value and *df* is the degrees of freedom.

=CHITEST (*observed freq range*, *expected freq range*) returns p-value for testing observed versus expected frequencies. Does NOT compute expected frequencies

=FDIST(*F*, *numerator df*, *denominator df*) - Gives p value for the F Distribution. Can also use the get a p value for a t statistic by squaring the t. For numerator $df=1$, $t^2 = F$.

other probability density functions

=BINOMDIST(*num successes*, *num trials*, *prob of success*, *cumulative*) – If cumulative is set to zero, gives probability of the number of successes out of the number of trials assuming the probability of each success is “prob of success”. If cumulative is zero to one, givesn probability of 0, 1, ... num of successes)

=POISSON(*x*, *mean*, *cumulative*) - Gives probability of observing “x” given a mean rate if cumulative is zero. If cumulative is one, gives probability of 0,1, ...x. Rounds “x” down to the nearest integer.

Functions that produce statistic percentiles when given a p value

=FINV(*upper tail p value*, *numerator df*, *denominator df*) – Gives F statistics for the given p value

=CHIINV(*p value*, *df*) – Gives chi-square statistic for the given p value.

Logical functions

=if(*logical expression*, *result if true*, *result if false*)

Example: Imagine that column AC contains date of surgery and column AD contains last follow up date. If AD is not later than AC, this is a mistake. In a new column, we type

If(AD1 – AC1 < 0, 1, 0)

The values in the new column are equal to 1 if there is a mistake (if AD is less than AC) and are equal to zero otherwise.

Regression and correlation

=slope(*y data, x data*) - Gives the slope (b) for the linear regression $y = a + b X$

=intercept(*y data, x data*) - Gives the intercept (a) for the linear regression $y = a + b X$

=correl(*y data, x data*) - Give the Pearson correlation coefficient, r

There are many other functions in EXCEL – this is only a small sample.