Using Excel for statistical analysis

Spreadsheet formulas: Enter these formulas, as they appear, into a cell in a worksheet. Where a formula specifies “data range,” enter the addresses of the beginning and ending cells of the range that contains the data. If some other argument is needed, enter either the argument or a cell address that contains the value for the argument. (Note: an “argument” here means a value that must be entered into the formula and that will change what the formula calculates. Examples appear below.)

=\text{AVERAGE}(\text{data range})
\text{arithmetic mean of the data in specified range}
\text{EX:} =\text{average}(a4:a18) \text{ gives the mean of the data in column A, rows 4 through 18}

=\text{STDEV}(\text{data range})
\text{sample standard deviation of data in specified range}

=\text{STDEVP}(\text{data range})
\text{population standard deviation}

=\text{MEDIAN}(\text{data range})

=\text{QUARTILE}(\text{data range}, \text{number})
“number” is the number of quartile desired: 1, 2, 3, 4. “Number” is an argument in the quartile formula.
\text{EX:} =\text{quartile}(a4:a18,3) \text{ gives the third quartile of the data in range a4:a18}

=\text{MAXIMUM}(\text{data range})
=\text{MINIMUM}(\text{data range})

=\text{SMALL}(\text{data range}, \text{number})
the nth smallest value in the data range, where n is specified as “number”
\text{EX:} =\text{small}(a4:a18,2) \text{ gives the second smallest number in the range a4:a18}

\text{NOTE:} When inserting formulas into a cell, the equals sign “=” must appear as the first character in the cell, but thereafter one need not enter “=” again to indicate a formula.
\text{EX:} The expressions 4-13 and \text{average}(a4:a18) will be treated as labels and show as entered
\text{EX:} =\text{maximum}(a4:a18)-\text{minimum}(a4:a18) \text{ will give the range of the data in a4:a18}
The Data Analysis Tool Pack

- From the tool bar click on “TOOLS” then “DATA ANALYSIS”
  - If you haven’t previously used the Data Analysis facility, you may have to initiate access to it: Click on “TOOLS” then click on “ADD INS”. When the Add Ins menu appears, place checks in the boxes for Analysis ToolPak and Analysis ToolPak VBA
  - The Data Analysis Tool Pack offers a menu of procedures, each of which is guided by a dialog box
  - **Examples:**
    - Descriptive Statistics
    - Histogram

More spreadsheet formulas:

- NORMDIST($x_0$, $\mu$, $\sigma$, cumulative)
  - in this formula, $\mu$ is the mean of a normal distribution and $\sigma$ its standard deviation. “cumulative” is an argument which has values “false” and “true”: false returns height of normal curve at $x_0$
  - true returns probability $x \leq x_0$

**EXAMPLE:** Suppose $\mu = 100$, $\sigma = 15$, find probability $x \leq 85$; $x \leq 105$; $x > 130$,

- $x > 145$
  - Solution: enter =NORMDIST(85, 100, 15, true)
  - For $x > 130$, enter =1 − NORMDIST(130,100,15,true)
  - General solution: In cell A5: =NORMDIST(A3,A1,A2,true)
    - In cell A6: =(1 − A5)

- NORMINV(probability, $\mu$, $\sigma$)
  - returns $x_0$ such that $x \leq x_0$ has the given probability

**EXAMPLE:** As above, what is an $x_0$ such that only 20% of population are less than or equal to $x_0$?
  - Solution: enter =NORMINV(.2,100,15)
  - Find $x_0$ such that only 10% of distribution is more than $x_0$:
    - =NORMINV(.9,100,15)
General solution?

- **NORMSDIST(z value):** proportion of distribution $\leq$ given z value

- **NORMSINV(probability):** value $z_0$ such that $P(z \leq z_0)$ has the probability entered

**EXAMPLES:**

- =normsdist(1.96) will give the result 0.975
- =normsdist(-1.96) yields 0.025
- =normsinv(0.05) yields -1.64485
- =normsinv(0.95) yields 1.64485

- **TDIST(t value, degrees of freedom, tails)**
  Here t value is calculated from the data, degrees of freedom are specified by the nature of the problem, and the argument tails has values 1 or 2. If tails = 1, the formula gives the area in one tail of a t distribution beyond the specified t value; if tails = 2, the formula gives the area in both tails. The tdist function is used to find the p-value in hypothesis tests; it gives the probability of a t value as large or larger than the one actually obtained.

- **TINV(probability, degrees of freedom)** gives the two-tailed t value for the specified probability.
  The tinv function is used for finding t values to be used in calculating confidence intervals and for finding critical t values in hypothesis tests.

**EXAMPLES:**

- =tdist(2,9,2) will yield 0.076553
- =tdist(2,9,1) will yield 0.038276
  note that 0.076553 = 2 $\times$ 0.038276
- =tinv(0.05,9) will show 2.262159, the value that would be used in constructing a 95% confidence interval
- =tinv(0.1,9) will show 1.833114, the critical t-value for an upper one tail test at 10% significance level