Sample 3: Statistics and Curve-Fitting

Excel is loaded with statistical tools but TK Solver provides a handy collection of its own and can be used as a supplement in several areas. Data can be copied from Excel into TK lists or tables.

Example 1: Checking the Distribution of a Sample Data Set

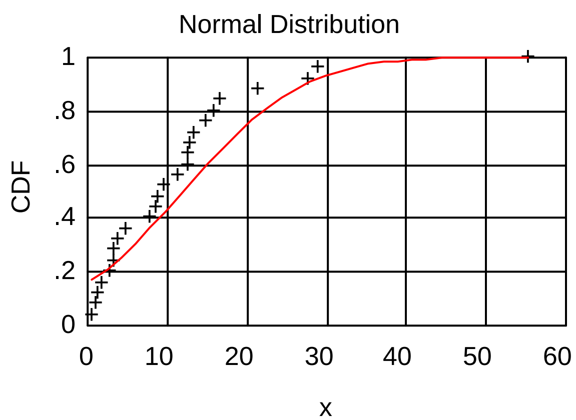
Many areas of statistics rely on an assumption that the data come from a particular underlying distribution. For example, we often assume that data come from a Gaussian Normal distribution. The TK Solver Library provides a collection of routines for checking the distribution of a sample data set.

- Launch the TK Library Menu and select the Statistics folder. Open the Curve Fitting folder and select Fitting a Normal Distribution. When the selection loads you will see a window with comments guiding you in the use of the routine.

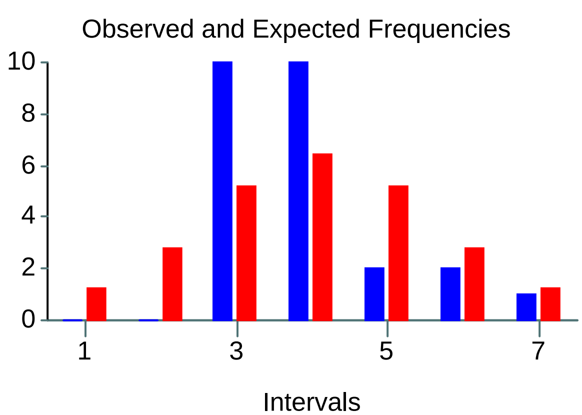
The file loads with sample data in a list called sample. This list name is required as an input on the Variable Sheet. The rest of the Variable Sheet is a summary of the results.

- Solve the model and observe the plots.

The Normal plot displays the data points along with the best-fit cumulative distribution function curve.



The test plot displays a frequency histogram with the expected frequencies overlaid. Based on the results, we should not assume that this data come from a normal distribution.



Example 2: Nonlinear Curve-Fitting

The TK Solver Library includes an easy-to-use curve-fitting routine for estimating parameters of nonlinear equations. In many cases, the form of the equation used to fit the data is already known from past experience. The task is to determine the coefficients which make the equation fit as well as possible. The following example shows how this is done in TK.

- Use the Menu to access the Statistics section of the Library. From the Curve Fitting area, load the routine labeled Nonlinear Curve-Fitting, One Predictor Variable. The instructions are presented in the form of a Comment Sheet.

The equation we need to fit to our data is

$$y = D + \frac{A}{1 + e^{(B \cdot (x - C))}}$$

- According to the instructions, we need to enter this as the definition in the procedure function called FUNCTION. Following the prescribed syntax conventions for the function, we enter it as shown below.

Statement
for i=1 to n
y[i] = b[4] + b[1]/(1+exp(b[2]*(x[i]-b[3])))
next i

All of the variables from our original equation have been replaced by references to list elements which are processed by the curve-fitting routine. A is now b[1], B is b[2], C is b[3] and D is b[4]. Each data point is represented by the ith element of a list, y[i] and x[i]. The names of the lists are passed into the function as input variables.

- The next step is to enter the data in the table called spec. We also enter initial guesses and bounds for the unknown coefficients in the b0, bmin and bmax columns. The b and SSE columns are reserved for output.

If you already have data stored in a file or another application, you must get it into TK lists called x and y (lower-case) in order for it to be processed by this routine. You can copy and paste it directly from a spreadsheet into the table. If you are importing the data, go to the List Sheet and change or blank the names of the existing x and y lists and then rename your lists.

Element	x	y	b0	bmin	bmax	b	SSE
1	20	158.81	100	0	1000		
2	30	159.5	.1	-1	1		
3	40	157.91	100	0	1000		
4	50	151.38	100	-1000	1000		
5	60	151.88					
6	70	140.69					
7	80	109.74					
8	90	62.12					
9	100	42.95					
10	110	44.44					
11	120	43.2					
12	130	41					
13	140	43.73					

- Solve and watch the status bar at the bottom of the TK Window for a progress report in the form of the sum of squared errors between the actual and predicted y values as TK tries new b coefficients.

After several steps, the results are added to the table. The SSE value decreased from over 9000 to 76.6 during the process. The top portion of the table is shown here.

Element	x	y	b0	bmin	bmax	b	SSE
1	20	158.81	100	0	1000	114.2769	76.60439
2	30	159.5	.1	-1	1	.1762495	
3	40	157.91	100	0	1000	81.63323	
4	50	151.38	100	-1000	1000	41.99055	

- Now let’s see how well our equation fits the data by viewing the plot called fit.