



Tutorials and worked examples for simulation,
 curve fitting, statistical analysis, and plotting.
<http://www.simfit.org.uk>

For a box and whisker plot you can use the exhaustive analysis of a matrix or 1-way ANOVA options from program **simstat**, or the simple and statistical plotting using program **simplot**. The case to be considered is where there are 5 groups of sizes 5, 8, 6, 8, and 8 using the library file `anova1.tf1`.

23	29	38	30	31
27	25	31	27	33
26	33	28	28	31
19	36	35	22	28
30	32	33	33	30
	28	36	34	24
	30		34	29
	31		32	30

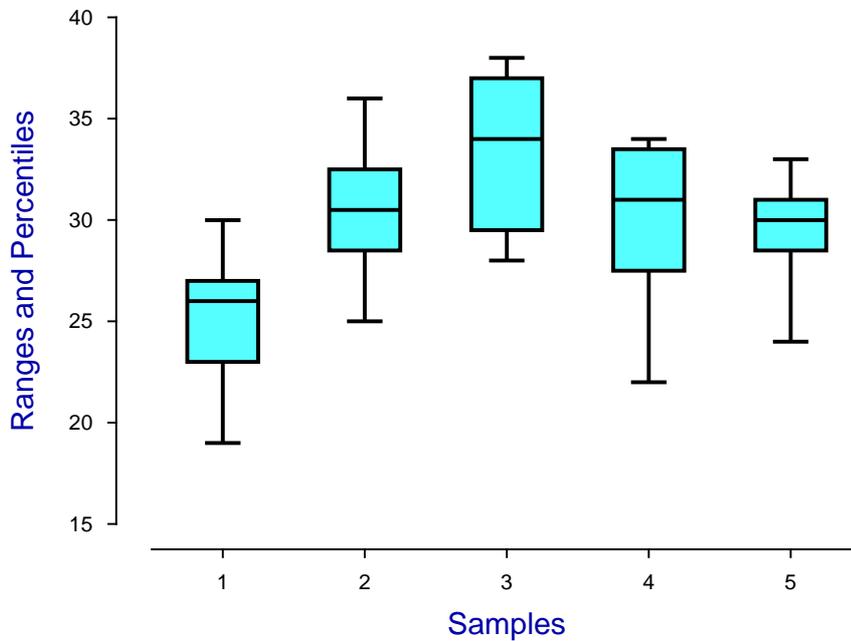
If sample sizes differ, data can be entered as an incomplete matrix with missing values, e.g., `incomplete.tf1`, individual column vectors, from a project archive, or as a library file referencing data files for each of the columns. Results from 1-way ANOVA are as follows.

1-Way Analysis of Variance: Grand Mean 29.89
 Transformation: x (untransformed data)

Source	SSQ	NDOF	MSQ	F	p
Between Groups	202.0	4	50.51	3.931	0.0111
Residual	385.5	30	12.85		
Total	587.5	34			

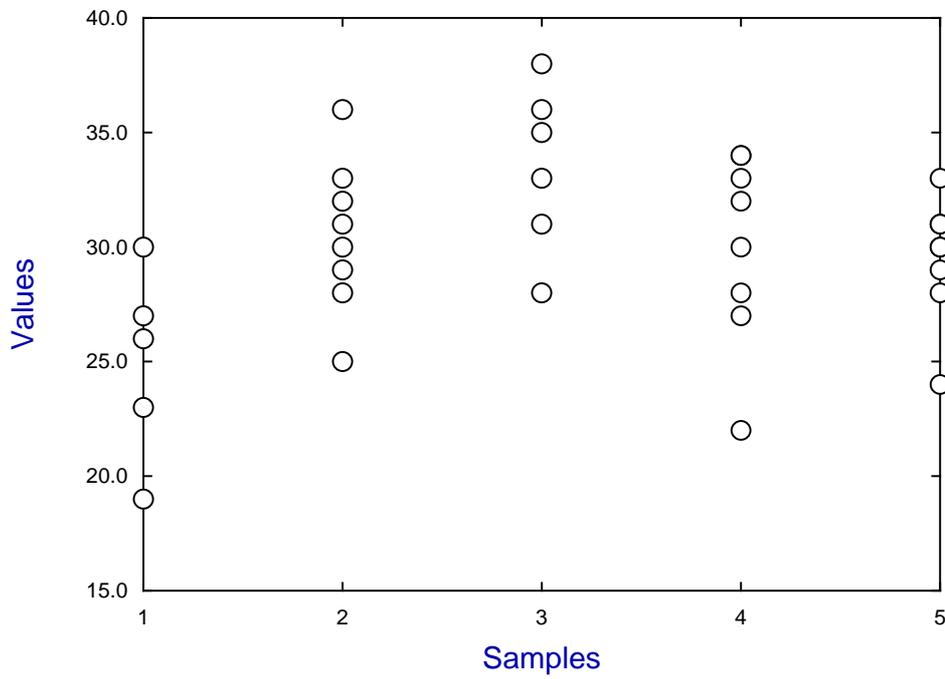
The the next plot shows sample ranges with medians and quartiles as a box and whisker plot.

Box and Whisker Plot

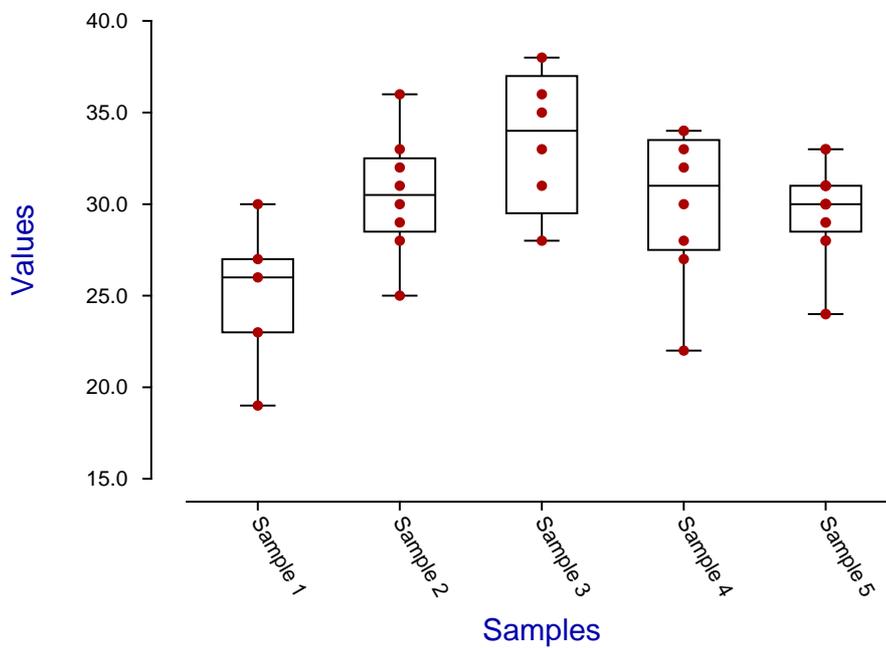


The results suggest rejecting the null hypothesis of equal means at the 5% significance level as $p < 0.05$, but not the 1% significance level as $p > 0.01$. The data can also be shown with error bars or as scattered points or even points scattered on the box and whisker plot as shown below.

Scatter Plot for Multiple Samples

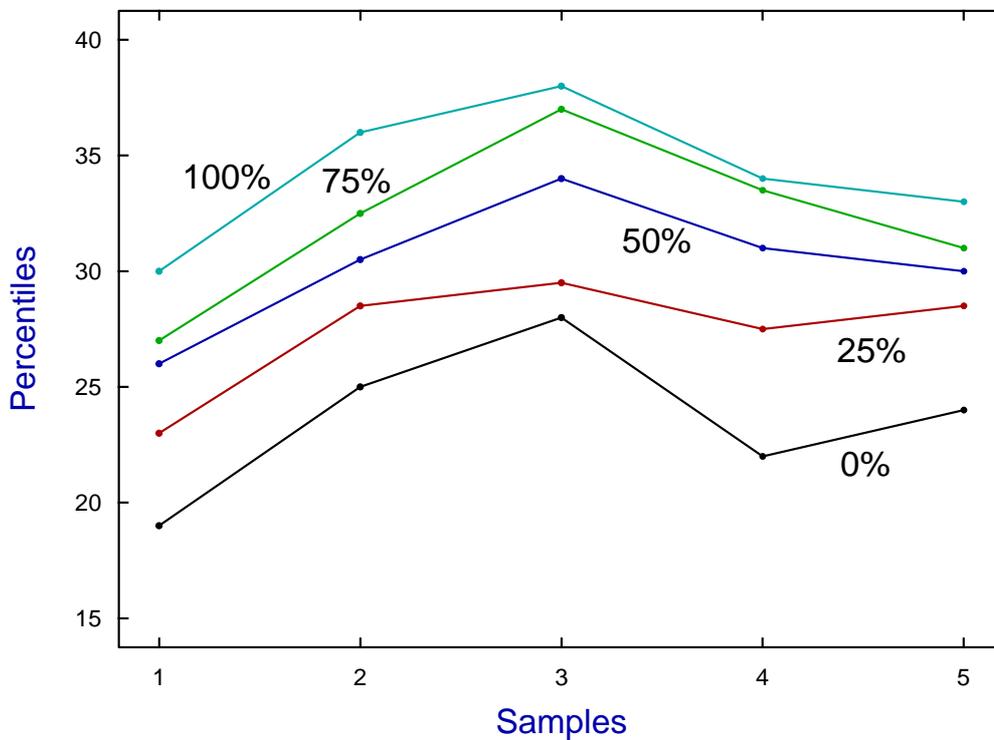


Box and Whisker Plot with Scattered Data



Another way to display the data is a range and percentiles plot as illustrated next. Here the lowest line segments join the lowest sample value for the corresponding groups, the upper line segments join the largest sample values, while between them the line segments join the points corresponding to the 25%, 50%, and 75% levels.

Range and Percentiles



Actually box and whisker plots can be created using advanced bar chart files like `barchart.tfl` which has the following format, derived from the data in `anova.tfl`.

```

Box and whisker plot
5 9
1 -2 -1 0 2 3 1 1 15
3 -1 0 2 4 5 1 1 15
5 1 2 3 5 6 1 1 15
7 0 1 2 4 5 1 1 15
9 1 3 5 6 7 1 1 15
7
begin{labels}
January
February
March
April
May
end{labels}

```

Note that, by constructing such advanced bar chart files, it is possible to create a great number of specialized plots, for instance: plots with overlapping bars, bars with error bars, mixed bars and box and whiskers, hanging bars, etc.

An explanation of the above three sections of data values follows.

1. **Header:** line 1: title, line 2: m = number of rows, n = number of columns
2. **Data:** line 3 onwards. Each line of data looks like this:-
 $x, y1, y2, y3, y4, y5, f, w, c$, and has the following interpretation:
 - x = x coordinate for the bar (x in nondecreasing order)
 - $y1$ = y for bottom of range (i.e. bottom error bar)
 - $y2$ = y for lower quartile (i.e. bottom of box)
 - $y3$ = y for median of data (i.e. divider for box)
 - $y4$ = y for upper quartile (i.e. top of box)
 - $y5$ = y for top of range (i.e. top error bar)
 - f = fill-style (between 0 and 10)
 - w = width (between 0 and 1)
 - c = colour (between 0 and 71)
3. **Labels:** line $m + 3$: number of important trailing text lines, the first $m + 2$ of these being
 - a) first of all `begin{labels}` to indicate the start of the labels,
 - b) then the consecutive labels for data rows 1 to m ,
 - c) then finally `end{labels}` to indicate the end of the labels.

For example, reading `barchart.tfl` into program **simplot** followed by minor editing then installing a PostScript special created the next plot.

SIMPLI Perspective Effects In Bar Charts

