

EXCEL と R による代表値の求め方一覧

求める代表値など	EXCEL の関数または手順 (範囲 A1:Y1 にデータがあるとして)	R の関数または手順 ($x < -c(\dots)$ などのやり方で変数 x にデータを入れたとして)
最頻値	離散データなら=MODE(A1:Y1) で良いが、連続量なら、ツール>分析ツール>ヒストグラムでヒストグラムを書いて最大度数のデータ区間を探し、その区間の中点を最頻値とする。	hist(x) でヒストグラムを書いて最大度数のデータ区間を探し、その区間の中点を最頻値とする。 hist(x,c(min(x),5,8,max(x))) などとすれば、 x の最小値から 5 まで、5 から 8 まで、8 から x の最大値までという 3 つの区間で度数を計算させることができる。本来は hist(x,5) とすれば 5 つの区間という形の指定ができるはずなのだが、区間の数によってうまくいったりいかなかったりする。 なお、hist(x,plot=F) とすれば、グラフを書く代わりに数値を表示させられる。
中央値	=MEDIAN(A1:Y1)	median(x) 但し、 x の中に NA (欠損値) を含む場合は、median(x,na.rm=T) または median(x[is.na(x)]) とする。以下同様。
平均値	=AVERAGE(A1:Y1) 調和平均は=HARMEAN(A1:Y1)、幾何平均は=GEOMEAN(A1:Y1) で求められる。	mean(x) 調和平均は $1/\text{mean}(1/x)$ 、幾何平均は $\exp(\text{mean}(\log(x)))$ で求められる。
範囲	=MAX(A1:Y1)-MIN(A1:Y1)	max(x)-min(x)
四分位範囲	=QUARTILE(A1:Y1,3) - QUARTILE(A1:Y1,1)	IQR(x) または、 $y < -\text{quantile}(x); y[4] - y[2]$ または、fivenum(x)[4] - fivenum(x)[2] でも良い。
四分位偏差	=(QUARTILE(A1:Y1,3) - QUARTILE(A1:Y1,1))/2	IQR(x)/2 または、 $y < -\text{quantile}(x); (y[4] - y[2])/2$ または、(fivenum(x)[4] - fivenum(x)[2])/2 でも良い。
平均偏差	=AVEDEV(A1:Y1)	組み込み関数にはないが、 sum(abs(x-mean(x)))/NROW(x) で得られる。
不偏分散	=VAR(A1:Y1) (不偏でない分散は=VARP(A1:Y1) で得られる)	var(x) 不偏でない分散は組み込み関数にはないが、 sum((x-mean(x))^2)/NROW(x) で得られる。
不偏標準偏差	=STDEV(A1:Y1) (不偏でない標準偏差は=STDEVP(A1:Y1) で得られる)	sd(x) 不偏でない標準偏差は、 sqrt(sum((x-mean(x))^2)/NROW(x)) で得られる。 (*)
タブ区切りデータファイルの読み込み	そのままドラッグ&ドロップ	1 行目に変数名が入っているなら、 $x < -\text{read.delim}("C:/My Documents/solomon.dat", \text{header}=T)$ などとする (**)。 それぞれの変数は、例えば x\$age のようにして参照できる。1 行目が変数名でなくすぐにデータである場合は、 $x < -\text{read.delim}("C:/My Documents/solomon.dat", \text{header}=F)$ とする。この場合、変数名は x\$V1, x\$V2, ... として参照できる。 一々 x\$ とつけるのが面倒なら、attach(x) とすれば V1 とか V21 だけで参照できる。
カンマ区切りデータファイルの読み込み	そのままドラッグ&ドロップ	1 行目に変数名が入っているなら、 $x < -\text{read.csv}("C:/My Documents/solomon.dat", \text{header}=T)$ とする (**)。1 行目が変数名でなくすぐにデータである場合は、 $x < -\text{read.csv}("C:/My Documents/solomon.dat", \text{header}=F)$ とする。
データの編集	表にそのまま打ち込む	de(x\$V1,x\$V5) などとすれば表形式で指定した変数の値を編集できる。表の上でマウスを右クリックすると操作メニューがでる。
データの書き出し	ファイルから保存を選ぶ	コンマ区切りでデータフレーム x をマイドキュメントの sample.dat に書き出すには、 write.table(x,"C:/My Documents/sample.dat",sep=",") とする。タブ区切りなら sep="¥t" とすればよい。

(*) もちろん、不偏でない分散を出すときに、 $Vx < -\text{sum}((x-\text{mean}(x))^2)/\text{NROW}(x)$ などとして値を保存しておいて、sqrt(Vx) とするのがエレガントである。

(**) ¥ を / に置き換えたファイル名をフルパスで書く。ただし、2 バイトコードが入ったディレクトリ名やファイル名は、文字化けするので使いにくい (半角英数字のファイル名に書き換えておくべきである)。また、Windows2000 の場合、マイドキュメントフォルダのフルパスは、普通、C:/Documents and Settings/nakazawa/My Documents/ のようになる。

(注) 表記の都合上全角で書いた文字もあるが、実際に入力するのはすべて半角である。