1 SUMMARY

VC04 fits a straight line to data by the least squares method.

The subroutine takes data values x_i, y_i, i=1, 2,..., m ($m{\geq}2$) and optionally weights w_i, i=1, 2,..., m, and finds the line y=a+bx where the parameters a and b are such that

$$\sum_{i=1}^{m} w_i \{y_i-(a+bx_i)\}^2$$

is minimised.

The weights may be omitted (which is as if w_i=1, i=1, 2,..., m). There are options for obtaining information relevant to the problem of fitting x as a function of y, i.e. $x(y)$=p+qy, and information is also available for the general analysis of the correlation problem.

ATTRIBUTES — **Version:** 1.0.0. **Types:** VC04A, VC04AD. **Calls:** FD05, OA03. **Original date:** August 1985. **Origin:** J.Ledger, Harwell.

2 HOW TO USE THE PACKAGE

2.1 Argument list

To fit a straight line to data using **no weights** and returning only the parameter values:

The single precision version

        CALL VC04A(X,Y,M,A,B,IPRINT)

The double precision version

        CALL VC04AD(X,Y,M,A,B,IPRINT)

To fit a straight line to data using **weights** and returning only parameter values:

The single precision version

        CALL VC04B(X,Y,W,M,A,B,IPRINT)

The double precision version

        CALL VC04BD(X,Y,W,M,A,B,IPRINT)

To fit a straight line using **weights** and returning, in addition to the parameter values, their **standard deviations**, the residual standard deviation and the correlation coefficient:

The single precision version

        CALL VC04C(X,Y,W,M,PAR,RHSTD,CORR,IPRINT)

The double precision version

        CALL VC04CD(X,Y,W,M,PAR,RHSTD,CORR,IPRINT)

X,Y   are REAL (DOUBLE PRECISION in the D version) arrays which the user must set to the data values x_i,y_i, i=1, 2,..., m.

W   is a REAL (DOUBLE PRECISION in the D version) array which must be set by the user to the weights w_i i=1, 2,.., m on a call to VC04B/BD or VC04C/CD. **Restriction**: W(I)$\geq$0, I=1, 2,..., M.

All use is subject to licence.

M   is an INTEGER set by the user to the number, m, of data points. **Restriction**: M$\geq$2.

A,B   are REAL (DOUBLE PRECISION in the D version) variables which are set to the values of the parameters a and b defining the fitted straight line $y=a+bx$ on a call to VC04A/AD or VC04B/BD.

PAR   is a REAL (DOUBLE PRECISION in the D version) array of length at least 4 which need not be set by the user. On return from VC04C/CD it will contain the parameters a and b defining the fitted straight line $y=ax+b$, and their estimated standard deviations, i.e. PAR(1)=a, PAR(2)=b, PAR(3)=$s(a)$, PAR(4)=$s(b)$.

RHSTD is a REAL (DOUBLE PRECISION in the D version) variable which need not be set by the user. On return from VC04C/CD, contains an estimate of the residual standard deviation (see section 4).

CORR is a REAL (DOUBLE PRECISION in the D version) variable which need not be set by the user. On return from VC04C/CD, it contains an estimate of the coefficient of correlation (see section 4).

IPRINT is an INTEGER variable which must be set by the user to specify one of the following print options:

>> if IPRINT=0, no printing is done,
>>
>> if IABS(IPRINT)=1, the parameters a and b, their estimated standard deviations, the residual standard deviation and the correlation coefficient are printed,
>>
>> if IABS(IPRINT)=2, the corresponding details for the fit $x=p+qy$ and the means and variances of the data are also printed,
>>
>> if IPRINT<0, a table of the data points, and weights if appropriate, is also printed.

2.2 Common

In the single precision version:
```
      COMMON/VC04D/LP,NDF,CCORR,AA,BB,ASTD,BSTD,ABCOV,RSTD,P,Q,
    *       PSTD,QSTD,PQCOV,RHOSTD,XBAR,YBAR,S2XX,S2YY,S2XY
      COMMON/VC04H/IERR
```

In the double precision version:
```
      COMMON/VC04DD/LP,NDF,CCORR,AA,BB,ASTD,BSTD,ABCOV,RSTD,P,Q,
    *       PSTD,QSTD,PQCOV,RHOSTD,XBAR,YBAR,S2XX,S2YY,S2XY
      COMMON/VC04HD/IERR
```

LP   is an INTEGER variable, with default value 6, set by BLOCK DATA VC04G/GD, which holds the stream number for printing. This may be reset by the user before calling the subroutine.

NDF   is an INTEGER variable which is set by the subroutine to the number of degrees of freedom (in fact this will always be $m-2$).

CCORR is a REAL (DOUBLE PRECISION in the D version) variable set by the subroutine to the coefficient of correlation.

IERR is an INTEGER variable which specifies one of the following error conditions:

> if IERR = 0, no errors,
>
> if IERR > 0, a weight is negative,
>
> if IERR = –1, the weights sum to zero,
>
> if IERR = –2, there is an insufficient number of points,
>
> if IERR = –3, the variance of x is zero,
>
> if IERR = –4, the variance of y is zero and y=a.

Regression *y=a+bx*

`AA,BB` are `REAL` (`DOUBLE PRECISION` in the D version) variables set by the subroutine to the parameters *a* and *b*.

`ASTD,BSTD,ABCOV` are `REAL` (`DOUBLE PRECISION` in the D version) variables set by the subroutine to the standard deviations of *a* and *b* and to the covar(*ab*) respectively.

`RSTD` is a `REAL` (`DOUBLE PRECISION` in the D version) variable set by the subroutine to the residual standard deviation.

Regression *x=p+qy*

`P,Q` are `REAL` (`DOUBLE PRECISION` in the D version) variables set by the subroutine to the parameters *p* and *q*.

`PSTD,QSTD,PQCOV` are `REAL` (`DOUBLE PRECISION` in the D version) variables set by the subroutine to the standard deviations of *p* and *q* and to the covar(*pq*) respectively.

`RHOSTD` is a `REAL` (`DOUBLE PRECISION` in the D version) variable set by the subroutine to the residual standard deviation.

Data statistics

`XBAR,YBAR` are `REAL` (`DOUBLE PRECISION` in the D version) variables set by the subroutine to the data means $\bar{x}$ and $\bar{y}$.

`S2XX,S2YY,S2XY` are `REAL` variables set by the subroutine to the variances s_{xx}^2 and s_{yy}^2 and the covariance s_{xy}^2, see section 4.

3 GENERAL INFORMATION

Use of common: `VC04D/DD, BLOCK DATA VC04G/GD, VC04H/HD`.

Other routines called directly: `FD05, _DOT, _AMAX, OA03A/AD` are called.

Input/output: Output is under the control of arguments `IPRINT` and `LP`.

Restrictions: $m \geq 2$ and $w_i \geq 0$, $i=1, 2,..., m$.

4 METHOD

The subroutine uses the following formulae:

$$\bar{w} = \sum_{i=1}^{m} w_i \qquad\qquad s_{xx}^2 = \frac{m}{m-1} \sum_{i=1}^{m} \frac{w_i(x_i-\bar{x})^2}{\bar{w}}$$

$$\bar{x} = \sum_{i=1}^{m} \frac{w_i x_i}{\bar{w}} \qquad\qquad s_{yy}^2 = \frac{m}{m-1} \sum_{i=1}^{m} \frac{w_i(y_i-\bar{y})^2}{\bar{w}}$$

$$\bar{y} = \sum_{i=1}^{m} \frac{w_i y_i}{\bar{w}} \qquad\qquad s_{xy}^2 = \frac{m}{m-1} \sum_{i=1}^{m} \frac{w_i(x_i-\bar{x})(y_i-\bar{y})}{\bar{w}}$$

The regression parameters are calculated simply by:

$$b = s_{xy}^2/s_{xx}^2, \quad a = \bar{y}-b\bar{x} \quad \text{and} \quad q = s_{xy}^2/s_{yy}^2, \quad p = \bar{x}-q\bar{y}$$

The coefficient of correlation is

$$r = \sqrt{bq}$$

Estimates of the weighted residual variances are

$$s_{ab}^2 = \frac{1}{m-2}\sum_{i=1}^{m} w_i\{y_i-(a+bx_i)\}^2 \quad \text{and} \quad s_{pq}^2 = \frac{1}{m-2}\sum_{i=1}^{m} w_i\{y_i-(p+qx_i)\}^2$$

The estimates of the parameter standard deviations are obtained from the variances

$$\text{var}(b) = \frac{m-1}{m}\frac{s_{ab}^2}{\bar{w}\,s_{xx}^2}, \text{var}(a) = \text{var}(b)\{\frac{m-1}{m}s_{xx}^2+\bar{x}^2\}$$

$$\text{var}(q) = \frac{m-1}{m}\frac{s_{pq}^2}{\bar{w}\,s_{yy}^2}, \quad \text{var}(p) = \text{var}(q)\{\frac{m-1}{m}s_{yy}^2+\bar{y}^2\}$$

The covariances are

$$\text{covar}(ab) = -\bar{x}\,\text{var}(b), \quad \text{covar}(pq) = -\bar{y}\,\text{var}(q).$$

5   EXAMPLE OF USE

The use of the subroutine can be illustrated by the following test program:

```
C
C     TEST FOR VC04 USING THREE DATA POINTS
C
C

      DIMENSION X(10),Y(10)
      DOUBLE PRECISION AA,BB,X,Y
      DOUBLE PRECISION XX
      COMMON/VC04DD/LP,NDF,XX(18)
      LP=9
C
C     M EQUALS THE NUMBER OF DATA POINTS
C
      M=3
      X(1)=1.1
      X(2)=1.9
      X(3)=3.05
      Y(1)=0.9
      Y(2)=1.95
      Y(3)=2.99
      IPRINT=1
      CALL VC04AD(X,Y,M,AA,BB,IPRINT)
      STOP
      END
```

This produces the following output

```
VC04AD: STRAIGHT LINE FIT.
   DEG. FREEDOM  1;  CORRELATION COEFF.  0.994385

VC04AD: REGRESSION Y = A + B*X.
   A= -1.9120988725065002D-01    B=   1.0601040763226364D+00
   RES. S.T.D.= 1.563905D-01

   STD(A)=  2.447746D-01      STD(B)=   1.128161D-01
   COVAR(AB)= -2.566709D-02
```