

Computing and comparing correlation and regression coefficients using a pocket calculator

D. V. M. BISHOP

*Neuropsychology Unit, Department of Clinical Neurology, Churchill Hospital
Headington, Oxford OX3 7LJ, England*

Programs suitable for pocket calculators using reverse Polish notation are described. Program 1 computes regression coefficients, correlation coefficient, and standard error of estimate for paired data. Program 2 performs a t test to compare the slopes of two regression lines. Program 3 computes F ratios to test the departure of a regression slope from zero and to test linearity of the regression. Programs 4 and 5 test the significance of differences between independent and correlated correlation coefficients, respectively.

Although the development of inexpensive microcomputers has brought considerable computing power within the reach of most scientists, the portability and ready availability of the pocket calculator remain a considerable advantage for those who work some distance from the nearest computer or those who have to queue to use it. Most manuals for pocket calculators provide several elementary statistical programs, usually including computation of correlation and/or regression coefficients, but, to my knowledge, none go beyond these simple statistics, and most are rather clumsy, requiring, for example, that separate programs be used for correlation and regression, so that the data have to be keyed in twice. The programs described here overcome some of these limitations. The programs were devised for a Hewlett-Packard 29C programmable calculator but can be adapted for use with other calculators that use reverse Polish notation. Users of other calculators who wish to modify the programs should be aware of the following features of the HP-29C.

(1) The summation key (marked $\Sigma+$) is used to automatically compute and store sums, sums of squares, and cross products for a set of paired data. The first value of a pair, (x_1), is keyed in, followed by the ENTER key. The second value, (y_1), is then keyed in, and the summation key is pressed. This sequence of operations is repeated until all paired data are entered. The number of data pairs (N) is stored in Register .0, sum of X values (ΣX) in Register .1, sum of Y values (ΣY) in Register .3, sum of squares for X (ΣX^2) in Register .2, sum of squares for Y (ΣY^2) in Register .4, and sum of cross products (ΣXY) in Register .5. Note that storage registers prefixed by a decimal point differ from those without such a prefix in the HP-29C. Programs 1 and 2 use these data in Stores .0 through .5, so users of other calculators will need to substitute the appropriate store numbers.

This work was supported by Medical Research Council Grant G 973/144/C.

(2) Program 1 involves computing the SD for y values. The HP-29C achieves this by Steps 47 and 48. Users of other calculators should insert the appropriate sequence of operations at this point.

(3) The HP-29C allows the user to label a particular part of a program using the LBL instruction. The user can then go directly to that part of the program using a GSB or GTO instruction. The number following a GTO or GSB refers to the number of a labeled part of the program and not to a step number. For example, in Program 2, Step 08 (GTO 2) instructs the machine to go to Step 04 (labeled 2).

(4) The instruction STO n results in the displayed value being placed in Store n, deleting any previous number from that store. To add a value to one already in store, the instruction STO + n is used. For example, in Program 2, Step 36 results in the value currently in the X register being added to the value in Store 3, and the sum is stored in Store 3. Subtraction, multiplication, and division with stored values are achieved by analogous commands (i.e., STO - n, STO \times n, STO \div n).

(5) The instruction RCL n recalls the contents of Store n to the display, without deleting this value from Store n.

(6) The instruction f Σ selectively clears Storage Registers .0 through .5, leaving other stores intact. Users of other calculators will need to modify Programs 2 (Step 63) and 3 (Step 30) accordingly.

PROGRAM 1: REGRESSION COEFFICIENTS, CORRELATION COEFFICIENT, AND STANDARD ERROR OF ESTIMATE

This program (Table 1) uses the data in Stores .0 through .5 (see above) to consecutively compute b_{yX} , a_{yX} , r_{xy} , and σ_{yX} (Guilford, 1973). Two additional stores are used in the computations.

Example

Paired data from Table 2 are entered so that N (5)

Table 1
Program 1

S	K	C	S	K	C	S	K	C
01	<i>LBL 0</i>	<i>label program</i>	18	÷		35	×	
02	RCL .2		19	R/S	<i>displays b_{yx}</i>	36	\sqrt{x}	
03	RCL .0		20	RCL .1		37	1/x	
04	×		21	×		38	RCL 2	
05	RCL .1		22	CHS (change sign)		39	×	
06	x^2		23	RCL .3		40	R/S	<i>displays r_{xy}</i>
07	-		24	+		41	x^2	
08	STO 1		25	RCL .0		42	CHS	
09	RCL .0		26	÷		43	1	
10	RCL .5		27	R/S	<i>displays a_{yx}</i>	44	+	
11	×		28	RCL .4		45	\sqrt{x}	
12	RCL .1		29	RCL .0		46	STO 1	
13	RCL .3		30	×		47	s	} s.d. for y
14	×		31	RCL .3		48	R+	
15	-		32	x^2		49	RCL 1	
16	STO 2		33	-		50	×	
17	RCL 1		34	RCL 1		51	RTN	<i>displays σ_{yx} and returns</i>

Note—S = step; K = keystroke; C = comment. Keystrokes in italics apply to programmable calculators only.

Table 2
Example for Program 1

Subject	Y	X
1	8	5
2	2	3
3	9	10
4	17	11
5	8	6

is in Store .0, ΣX (35) is in Store .1, ΣX^2 (291) is in Store .2, ΣY (44) is in Store .3, ΣY^2 (502) is in Store .4, and ΣXY (371) is in Store .5. Running the program successively displays b_{yx} (1.37), a_{yx} (-.79), r_{xy} (.87), and σ_{yx} (2.67).

PROGRAM 2: COMPARING SLOPES OF TWO REGRESSION LINES

This program (see Table 3) follows the procedure described by Armitage (1971) to perform a t test to compare two values of b_{yx} . As in the previous program, data are first entered so that N, ΣX , ΣX^2 , ΣY , ΣY^2 , and ΣXY are in the five storage registers, .0 through .5, respectively. An optional subroutine is given to facilitate data entry in cases in which there are several observations of y for each value of x. Part of the main program is identical to Steps 02-19 for Program 1 (computing b_{yx}). If desired, further parts of Program 1 (e.g., computation of a_{yx}) may also be inserted. In addition to the five registers used for accumulating sums, six further storage registers are required.

Example

Data from Table 4 may be used to demonstrate use of this program.

(a) Data entry: If the raw data occur in blocks, as in Table 4, the optional subroutine for data entry may be used. The optional data entry subroutine in Program 2 was labeled using the instruction LBL 1, so that in RUN mode, pressing GSB 1 takes us to this part of the program. To enter data for Group A, first key in the first x value (1), followed by R/S (run/stop). Then enter all y values (10, 11, 12, 19, 11) corresponding to this x value, following each by R/S. Repeat this procedure (from GSB 1 onward) for the next x value until all data for Group A are entered. Proceed to Step b).

If data are not in blocks and the optional data entry subroutine is not used, start with accumulated sums in Stores .0 through .5 (as in Program 1) and proceed directly to Step b).

(b) Press GSB 3 (to go to Program Step 09). Display will show b_{yx} (5.27 for this example).

(c) Press R/S again (to go to Step 57).

(d) Press GSB 4 (to go to Step 58).

(e) Repeat Steps a, b, and c, but not Step d for Group B. (For this example, b_{yx} for Group B = 2.69.)

(f) Press GSB 5 (go to Step 68) to display t (4.17) and df (31).

PROGRAM 3: DEPARTURE OF SLOPE FROM ZERO AND TEST OF LINEARITY

Program 3 (Table 5) is suitable for data for which

Table 3
Program 2

S	K	C	S	K	C	S	K	C
01	<i>LBL 1</i>		39	RCL .3		63	FE	clears registers .0 through .5
02	<i>R/S</i>		40	×		64	<i>R/S</i>	
03	<i>STO 1</i>		41	RCL .0		65	<i>LBL 5</i>	
04	<i>LBL 2</i>	<i>optional subroutine to enter data in blocks</i>	42	÷		66	RCL 7	
05	<i>R/S</i>		43	RCL .5		67	RCL .0	
06	<i>RCL 1</i>		44	-		68	+	
07	<i>LT</i>		45	x^2		69	4	
08	<i>GTO 2</i>		46	RCL 4		70	-	
09	<i>LBL 3</i>		47	×		71	STO 1	
		<i>steps 10 to 27 as for steps 02 to 19 in program 1</i>	48	CHS		72	1/x	
			49	RCL .4		73	RCL 5	
			50	+		74	×	
27	<i>R/S</i>	<i>displays b_{yx}</i>	51	RCL .3		75	RCL 3	
28	<i>STO 2</i>		52	x^2		76	×	
29	RCL .2		53	RCL .0		77	\sqrt{x}	
30	RCL .1		54	÷		78	1/x	
31	x^2		55	-		79	RCL 6	
32	RCL .0		56	STO + 5		80	RCL 2	
33	÷		57	<i>R/S</i>		81	-	
34	-		58	<i>LBL 4</i>		82	×	
35	1/x		59	RCL 2		83	<i>R/S</i>	<i>displays t</i>
36	STO + 3		60	STO 6		84	RCL 1	
37	STO 4		61	RCL .0		85	<i>RTN</i>	<i>displays d.f. and returns</i>
38	RCL .1		62	STO 7				

Note—S = step; K = keystroke; C = comment. Keystrokes in italics apply to programmable calculators only.

Table 4
Example for Programs 2 and 3

Group A			Group B			
X			X			
1	2	3	1	2	4	5
10	15	20	6	10	12	19
11	13	21	7	10	12	19
Y 12	15	22	Y 8	11	13	20
10	19	20	9	12	14	21
11	17	20	12		20	
	16	25				

there are a number of replicate readings of y for each value of x. The formulas are taken from Armitage (1971).

Example

Data from Group A in Table 4 may be used to demonstrate use of this program.

(a) For data entry, press GSB 1 (to go to Program Step 01).

(b) Key in first x value (1) followed by R/S.

(c) Key in all y values corresponding to this x (10, 11, 12, 10, 11), following each with R/S.

(d) Press GSB 3 (to go to Program Step 09).

(e) Repeat Steps b, c, and d for next x value

(f) When all data are entered, press GSB 4 (to go to Program Step 32).

(g) Display shows F value for testing departure of b_{yx} from zero assuming linear regression (99.17).

(h) Press R/S twice for df for numerator and denominator (1,14).

(i) Press R/S for F value for testing the deviation from linearity (.07).

(j) Press R/S twice for df for numerator and denominator.

PROGRAM 4: COMPARISON OF INDEPENDENT CORRELATION COEFFICIENTS

This program (see Table 6) converts Pearson correlation coefficients to Fisher's Z before testing the difference between them by the formula:

$$z = (Z_a - Z_b) / \sqrt{1/(N_a - 3) + 1/(N_b - 3)}$$

Example

For Group A, let $N = 50$ and $r_{xy} = .82$. For Group B let $N = 60$ and $r_{xy} = .92$. Put .82 in Store 1, .92 in Store 2, 50 in Store 3, and 60 in Store 4 and run the program to obtain $z = -2.19$, which is evaluated with reference to the normal distribution.

Table 5
Program 3

S	K	C	S	K	C	S	K	C
01	<i>LBL 1</i>		29	RCL 8		57	RCL 9	
02	<i>R/S</i>		30	f Σ	clear stores	58	<i>R/S</i>	
03	<i>STO 0</i>		31	<i>GTO 1</i>	.0 through .5	59	RCL 0	
04	<i>LBL 2</i>	<i>data entry</i>	32	<i>LBL 4</i>		60	CHS	
05	<i>R/S</i>	<i>in blocks</i>	33	RCL 1		61	RCL 7	
06	<i>RCL 0</i>		34	RCL 3		62	+	
07	<i>E+</i>		35	x		63	RCL 3	
08	<i>GTO 2</i>		36	RCL 6		64	x ²	
09	<i>LBL 3</i>		37	\div		65	RCL 6	
10	RCL .0		38	RCL 5		66	\div	
11	STO + 6		39	-		67	-	
12	RCL .1		40	x ²		68	RCL 9	
13	STO + 1		41	RCL 1		69	<i>GSB 5</i>	<i>go to step 76</i>
14	RCL .2		42	x ²		70	RCL 8	
15	STO + 2		43	RCL 6		71	2	
16	RCL .3		44	\div		72	-	
17	STO +3		45	CHS		73	<i>R/S</i>	
18	RCL .4		46	RCL 2		74	RCL 9	
19	STO + 4		47	+		75	<i>R/S</i>	
20	RCL .5		48	\div		76	<i>LBL 5</i>	
21	STO + 5		49	STO 0		77	x	
22	RCL .3		50	RCL 6		78	RCL 4	
23	x ²		51	RCL 8		79	RCL 7	
24	RCL .0		52	-		80	-	
25	\div		53	STO 9		81	\div	
26	STO + 7		54	<i>GSB 5</i>	<i>go to step 76</i>	82	<i>R/S</i>	
27	1		55	1		83	<i>RTN</i>	
28	STO + 8		56	<i>R/S</i>				

Note—S = step; K = keystroke; C = comment. Keystrokes in italics apply to programmable calculators only.

Table 6
Program 4

S	K	C	S	K	C	S	K	C
01	<i>LBL 1</i>		13	3		25	STO 0	
02	RCL 1		14	-		26	1	
03	<i>GSB 0</i>	<i>go to step 24</i>	15	1/x		27	+	
04	STO 1		16	+		28	log _e	
05	RCL 2		17	\sqrt{x}		29	1	
06	<i>GSB 0</i>	<i>go to step 24</i>	18	1/x		30	RCL 0	
07	STO 2		19	RCL 1		31	-	
08	RCL 3		20	RCL 2		32	log _e	
09	3		21	-		33	-	
10	-		22	x		34	2	
11	1/x		23	<i>R/S</i>	<i>displays z</i>	35	\div	
12	RCL 4		24	<i>LBL 0</i>	<i>labels subroutine</i>	36	<i>RTN</i>	<i>return</i>

Note—S = step; K = keystroke; C = comment. Keystrokes in italics apply to programmable calculators only.

Table 7
Program 5

S	K	C	S	K	C
01	LBL 0		22	+	
02	RCL 1		23	2	
03	RCL 2		24	×	
04	×		25	1/x	
05	RCL 3		26	RCL 3	
06	×		27	1	
07	2		28	+	
08	×		29	RCL 0	
09	RCL 3		30	3	
10	x ²		31	-	
11	CHS		32	STO 4	
12	+		33	×	
13	RCL 2		34	×	
14	x ²		35	√x	
15	CHS		36	RCL 1	
16	+		37	RCL 2	
17	RCL 1		38	-	
18	x ²		39	×	
19	CHS		40	R/S	<i>displays t</i>
20	+		41	RCL 4	
21	1		42	RTN	<i>displays d.f. and returns</i>

Note—S = step; K = keystroke; C = comment. Keystrokes in italics apply to programmable calculators only.

PROGRAM 5: COMPARISON OF CORRELATED CORRELATION COEFFICIENTS

Guilford (1973) describes a formula devised by Hotelling that can be used when comparing two correlation coefficients from the same sample, when one of the correlated variables is the same in both cases. Program 5 (Table 7) uses this formula to give a value of t .

Example

In a sample of 200 children, a nonverbal IQ test correlates .45 with reading ability, a verbal IQ test correlates .55 with reading ability, and the correlation between the two IQ tests is .60. To test whether the verbal IQ test is significantly more strongly correlated with reading ability than is the nonverbal test, run Program 5 as follows. Put N (200) in Store 0, r_{12} (.45) in Store 1, r_{13} (.55) in Store 2, and r_{23} (.60) in Store 3. The program is now run to give t (-1.91) and df (197).

REFERENCES

- ARMITAGE, P. *Statistical methods in medical research*. Oxford: Blackwells, 1971.
 GUILFORD, J. P. *Fundamental statistics in psychology and education*. Tokyo: McGraw-Hill Kogakusha, 1973.

(Received for publication April 6, 1979;
revision accepted November 20, 1979.)