Introductory Information for Subroutines No. F2-38a and F2-38b.

These two related but operatively independent subroutines are useful for two types of jointly distributed variables, depending on their mean values. Subroutine No. F2-38a is appropriate if either mean value is non-zero, and the generated least squares regression line has the form,

$$Y_c = a + bX$$

The point defined by the mean values of the variables, $(\bar{X},\bar{Y})$ lies on the line, either (or both) of the mean values are non-zero, and the line does not pass through the origin (or what is the same thing, the constant "a" is non-zero).

If, on the other hand, the bivariate distribution's variables happen to have mean values which are both zero or, if a transformation is made on X and Y such that the transformed variables, x and y, meet this condition, then Subroutine No. F2-38b is fitting. It may be used to fit the regression line

$$y_c = bx$$

to the joint distribution.

The following symbol-conventions are used:

1. Capital alphabetical symbols, X, and Y, represent "observed" data or "scores" measured from an untampered origin; i.e., no transformation of the type listed below has been made on such X, Y data.

2. Small alphabetical symbols, x, and y, are used to represent deviations or measurements from sample mean-values $\bar{X}$, and $\bar{Y}$, and may be obtained from initial X, and Y measurements or scores by the relations:

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$

3. The Y (or y) symbols with subscripts "c" represent "estimated" or computed values supplied by the least squares equations.

Both No. F2-38a and F2-38b use the original floating point interpretive routine, No. 24.0 and its input-output companion, No. 11.6-12.6.

These regression programs come equipped with tapes having three short examples and a control program for directing the computer through its operations while executing the computations. Information is punched in the following order with one tape each for No. F2-38a, and No. F2-38b:

1. Bivariate Linear Regression (Either F2-38a or F2-38b)
2. Short Control Program
3. Three examples

The Flexowriter printing done while the computer executes these sets of examples appears on pages 4 and 12 of this program description. In practice, regression-correlation calculations would not be made on such small sets of data, but these serve well to illustrate program operation. It will also be noted that the usual equal-spacing requirement for simplified data-transformation has not been met; this condition was down-graded when an interest developed in applying No. F2-38b to No. F2-38a's previously constructed examples.

Program Title: "Bivariate Linear Regression"
(Straight Line Fit to Observed Data)

Author: R. L. Stearman

Installation: Booz-Allen Applied Research, Inc., Bethesda, Maryland.

Purpose: To compute the least squares regression line, of the form $Y_c = a + bX$, to a set of points in floating point and to compute the correlation coefficient, r, for this set of points. (See list of additional statistics computed, as given below under "Output").

Input: (Note: This is the same for both F2-38a and F2-38b)

1. Location of the floating point interpretive subroutine, No. 24.0.
2. Location of N, the size of sample (N must be given in floating point).
3. Location of the first value of X (values of X must be in floating point and stored sequentially).
4. Location of the first value of Y (values of Y must also be in floating point and stored sequentially according to the order of the corresponding values of X).

Subroutine must be entered in fixed point and exits in fixed point.

Output:

The following statistics are stored within the program:

Statistic	Storage Location
$\sum X$	L_o + 0103
$\sum X^2$	L_o + 0104
$\sum x^2 = s.s._x = \sum X^2 - \frac{(\sum X)^2}{N}$	L_o + 0106
$\sum Y$	L_o + 0107
$\sum Y^2$	L_o + 0108
$\sum y^2 = s.s._y = \sum Y^2 - \frac{(\sum Y)^2}{N}$	L_o + 0110
$\sum XY$	L_o + 0111
$\sum xy = s.p._{xy} = \sum XY - \frac{(\sum X)(\sum Y)}{N}$	L_o + 0113
$b = s.p._{xy} / s.s._x$	L_o + 0114
$a = \frac{(\sum Y - b\sum X)}{N}$	L_o + 0115

Output: Continued

Statistic * (See definitions below)	Storage Location
$r = s.p._{xy} / \sqrt{s.s._x s.s._y}$	L_o + 0102
$s.s._{reg} = s.p._{xy}^2 / s.s._x$	L_o + 0116
$s.s._{y.x} = s.s._y - s.s._{reg}$	L_o + 0117
$d.f. = n - 2$	L_o + 0118
$s^2_{y.x} = s.s._{y.x} / d.f.$	L_o + 0119
$F = s.s._{reg} / s^2_{y.x}$	L_o + 0120

Linkage & Calling Sequence: (Note: this is the same for both Subroutines F2-38a and F2-38b)

To be entered while in fixed point.

Location	Instruction	Address
α	R	L_o
α + 1	U	L_o
α + 2	Z	Location of floating point
α + 3	Z	Location of n
α + 4	Z	Location of first x
α + 5	Z	Location of first y
α + 6	etc.	

Storage:

2 tracks and 24 sectors

DI 1

Program Stops:

None

---

* Further Definition of Expressions.

r : Correlation Coefficient.

$s.s._{reg}$: Sum of Squares due to regression

$s.s._{y.x}$: Sum of Squares due to residual

d.f. : Degrees of Freedom

$s^2_{y.x}$: Mean Square due to residual

LGP-30 USERS' ORGANIZATION - POOL
"Bivariate Linear Regression" Program No. F2-38a
(Straight Line Fit to Observed Data)

CONTROL PROGRAM FOR OPERATING No. 38a and THREE SHORT EXAMPLES

CONTROL PROGRAM

```
;0003400'/0003400'
xr1200'          lo. floating point
xu1200'
xi0000'          input data
xe0000'          exit f.p.
xr3000'          L-zero Bivariate Linear Regression
xu3000'
xz1200'          floating point lo.
xz4000'          lo. n
xz4001'          lo. X s
xz4032'          lo. Y s
xz0000'          stop
xp1600'          c.r.
xz0000'          delay
xr1200'
xu1200'
xb3115'          a
xp0000'          print a
xb3114'          b
xp0000'          print b
xb3102'          r
xp0000'          print r
xb3120'          F
xp0000'          print F
xe0000'          exit f.p.
xz0000'          stop
```

EXAMPLE NO. I

```
.0003400'
+004000'      5'
0'1'3'6'8'       X s    -0000000'
+004032'
1'3'2'5'4'       Y s    -0000000'  '

.1646018  01     .3761061  00     .7996127  00     .5319016  01
```

EXAMPLE NO. II

```
.0003400'
+004000'
5'               n
0'1'2'3'4'       X s    -0000000'
+014032'
0'19'40'59'81'   Y s    -0000000'  '

.6000023- 01-    .2020000  01     .9997060  00     .5101681  04
```

EXAMPLE NO. III

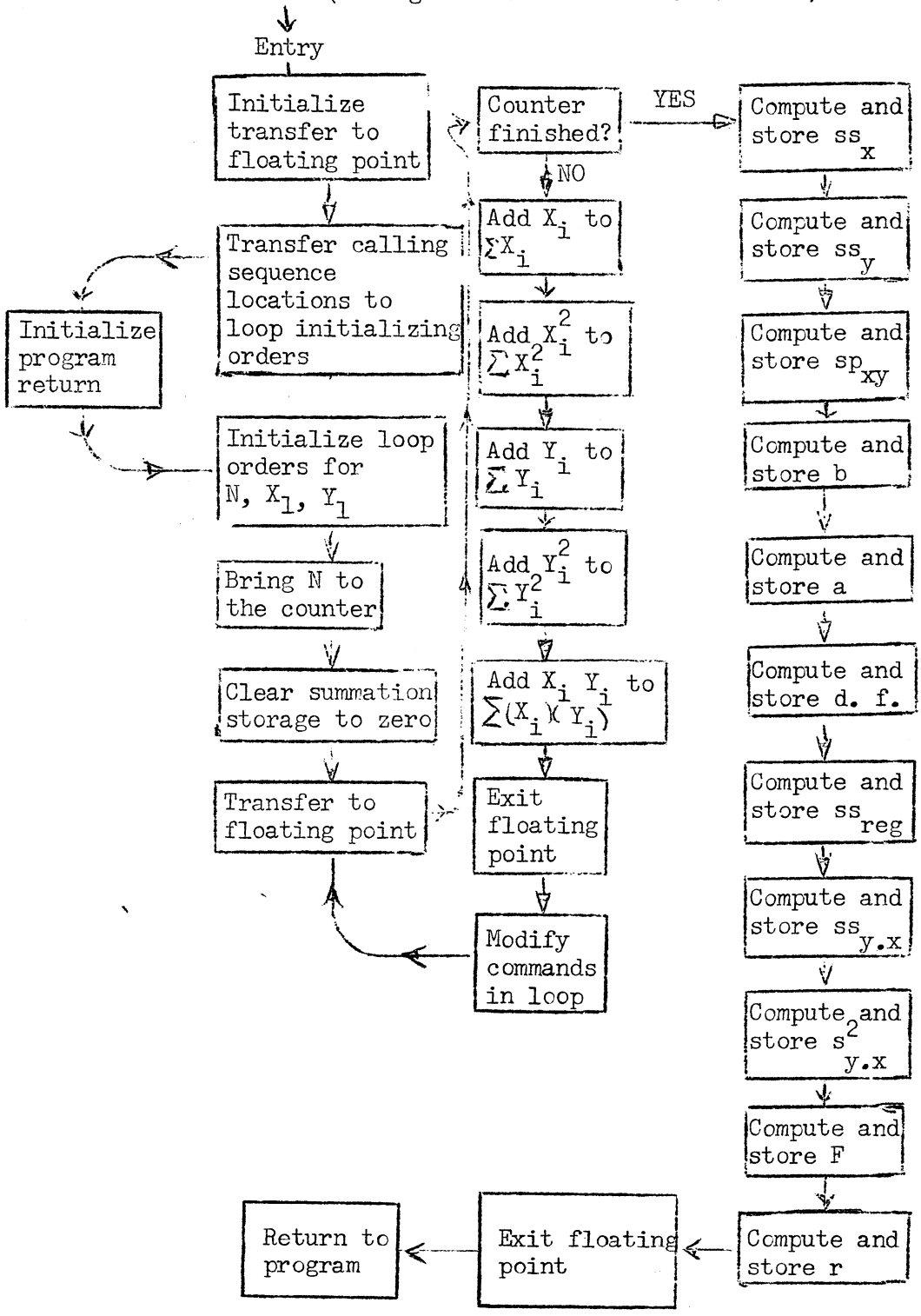
```
.0003400'
+004000'      6'
52'110'205'377'400'500'      -0000000'
+004032'
252'280'360'400'450'520'     -0000000'  +0000000'

.2236380  03     .5597153  00     .9806940  00     .1006043  03
```

Program No. F2-38a
Dec. 1, 1959.

Author: R. L. Stearman

Installation: Booz-Allen Applied Research Inc., Bethesda 14, Maryland.

Floating Point
"BIVARIATE LINEAR REGRESSION"
(Straight Line Fit to Observed Data)



PREPARED FOR: LGP-30 USERS' ORGANIZATION - POOL | PAGE 1 / 5

JOB NO. | PROGRAM NO. F2-38a | PROGRAM PREPARED BY: R.L.Stearman | PROGRAM CHECKED BY: POOL Review | DATE Dec.1,1959.

PROBLEM: "Bivariate Linear Regression Straight Line Fit to Observed Data: y = a + bx | TRACK

PROGRAM INPUT CODES	STOP	LOCATION	INSTRUCTION OPERATION	INSTRUCTION ADDRESS	STOP	CONTENTS OF ADDRESS	NOTES
000	/					The symbols for x and y may be considered capitalized, since ~~they represent Observed Data.~~	
000	/	☒					
		00 00	X B	[ ]	/	Fl. pt.	
		00 01	Y	0036	/		
		00 02	Y	0037	/		
		00 03	B	0000	/	☒ loc loc n	
		00 04	A	0121	/		
		00 05	Y	0011	/		
		00 06	A	0121	/	loc loc x_1	
		00 07	Y	0019	/	☒	
		00 08	A	0121	/	loc loc y_1	
		00 09	Y	0024	/		
		00 10	A	0121	/	Prog.	
		00 11	Y	0223	/	☒ return	initialize
		00 12	X B	[ ]	/		program
		00 13	Y	0029	/		for locations
		00 14	Y	0135	/		
		00 15	Y	0142	/	☒ loc n	
		00 16	Y	0149	/		
		00 17	Y	0161	/		
		00 18	Y	0163	/		
		00 19	X B	[ ]	/	☒	
		00 20	Y	0042	/	loc x_1	
		00 21	Y	0045	/		
		00 22	Y	0046	/		
		00 23	Y	0056	/	☒	
		00 24	X B	[ ]	/		
		00 25	Y	0049	/		
		00 26	Y	0052	/	loc y_1	
		00 27	Y	0053	/	☒	
		00 28	Y	0057	/		
		00 29	X B	[ ]	/	n	n counter
		00 30	C	0100	/	counter	
		00 31	C	0103	/	☒ Σx	clear sum into storage

FORM LP-10 | Royal McBee Corporation DATA PROCESSING DIV. PORT CHESTER, NEW YORK | PRINTED IN U.S.A.

☒ = CARRIAGE RETURN

/ = CONDITIONAL STOP CODE

PREPARED FOR: LGP-30 USERS' ORGANIZATION - POOL | PAGE 2 OF 15

JOB NO. | PROGRAM NO. F2-38a | PROGRAM PREPARED BY: R.L.Stearman | PROGRAM CHECKED BY: POOL Review | DATE Dec. 1, '59

PROBLEM: "Bivariate Linear Regression Straight Line Fit to Observed Data: $y = a + bx$" | TRACK

PROGRAM INPUT CODES	STOP	LOCATION	INSTRUCTION OPERATION	INSTRUCTION ADDRESS	STOP	CONTENTS OF ADDRESS	NOTES
	/					The symbols for x and y may be considered capitalized, since they represent Observed Data.	
	/	☒					
		0 0 3 2	C	0 1 0 4	/	Σx^2	clear
		3 3	C	0 1 0 7	/	Σy	summation
		3 4	C	0 1 0 8	/	Σy^2	storage
		3 5	C	0 1 1 1	/	☒ Σxy	
		3 6	y R	[ ]	/		→ fl. pt.
		3 7	X U	[ ]	/		
		3 8	B	0 1 0 0	/	counter	
		3 9	S	0 1 0 1	/	☒ 1	counter
		4 0	T	0 1 3 3	/	out of loop	
		4 1	C	0 1 0 0	/	counter	
		4 2	X B	[ ]	/	x_i	
		4 3	A	0 1 0 3	/	☒ Σx_i	Σx_i
		4 4	C	0 1 0 3	/	Σx_i	
		4 5	X P	[ ]	/	x_i	
		4 6	X M	[ ]	/	x_i	Σx_i^2
		4 7	A	0 1 0 4	/	☒ Σx_i^2	
		4 8	C	0 1 0 4	/	Σx_i^2	
		4 9	X B	[ ]	/	y_i	
		5 0	A	0 1 0 7	/	Σy_i	Σy_i
		5 1	C	0 1 0 7	/	☒ Σy_i	
		5 2	X P	[ ]	/	y_i	
		5 3	X M	[ ]	/	y_i	Σy_i^2
		5 4	A	0 1 0 8	/	Σy_i^2	
		5 5	C	0 1 0 8	/	☒ Σy_i^2	
		5 6	X P	[ ]	/	x_i	
		5 7	X M	[ ]	/	y_i	$\Sigma x_i y_i$
		5 8	A	0 1 1 1	/	$\Sigma x_i y_i$	
		5 9	C	0 1 1 1	/	☒ $\Sigma x_i y_i$	
		6 0	X E	0 0 0 0	/		exit fl. pt.
		6 1	B	0 0 4 2	/		modify commands
		6 2	A	0 1 2 1	/	x_i	
		6 3	U	0 1 2 2	/	☒	

☒ CARRIAGE RETURN

/ = CONDITIONAL STOP CODE

PREPARED FOR: LGP-30 USERS' ORGANIZATION - POOL | PAGE 3 OF 15

JOB NO. | PROGRAM NO. F2-38a | PROGRAM PREPARED BY: R.L.Stearman | PROGRAM CHECKED BY: POOL Review | DATE Dec. 1, 1959.

PROBLEM: "Bivariate Linear Regression Straight Line Fit to Observed Data": $y = a + bx$ | TRACK

PROGRAM INPUT CODES	STOP	LOCATION	INSTRUCTION OPERATION	INSTRUCTION ADDRESS	STOP	CONTENTS OF ADDRESS	NOTES
	/					The symbols for x and y may be considered capitalized, since ~~they represent Observed Data.~~	
	/	☒					
0 0 0 0 0 2 1		0 1 0 0	[	]	/	counter	
		0 1	4 0 0 0	0 0 0 2	/	1	
		0 1 2			/	r	
		0 1 3			/	$\sum x$	
		0 1 4			/	$\sum x^2$	
		0 1 5			/	$(\sum x)^2/n$	
		0 1 6			/	ss_x	
		0 1 7			/	$\sum y$	
		0 1 8			/	$\sum y^2$	
		0 1 9			/	$(\sum y)^2/n$	storage
		1 1 0			/	ss_y	
		1 1 1			/	$\sum xy$	
		1 1 2			/	$(\sum x)(\sum y)/n$	
		1 1 3			/	$s.p._{xy}$	
		1 1 4			/	b	
		1 1 5			/	a	
		1 1 6			/	$s.s._{reg}$	
		1 1 7			/	$s.s._{y.x}$	
		1 1 8			/	d.f.	
		1 1 9			/	$s^2_{y.x}$	
		1 2 0	[	]	/	F	
		1 2 1	X Z 0 0 0 1		/		
		1 2 2	Y 0 0 4 2		/		
		1 2 3	Y 0 0 4 5		/	x_i	
		1 2 4	Y 0 0 4 6		/		
		1 2 5	Y 0 0 5 6		/		modify command
		1 2 6	B 0 0 4 9		/		
		1 2 7	A 0 1 2 1		/		
		1 2 8	Y 0 0 4 9		/	y_i	
		1 2 9	Y 0 0 5 2		/		
		1 3 0	Y 0 0 5 3		/		
		1 3 1	Y 0 0 5 7		/		

PREPARED FOR: LGP-30 USERS' ORGANIZATION - POOL

PAGE 4 OF 5

JOB NO. | PROGRAM NO. F2-38a | PROGRAM PREPARED BY: R.L.Stearman | PROGRAM CHECKED BY: POOL Review | DATE Dec. 1, '59

PROBLEM: "Bivariate Linear Regression Straight Line Fit to Observed Data": $y = a + bx$ | TRACK

PROGRAM INPUT CODES	STOP	LOCATION	INSTRUCTION OPERATION	INSTRUCTION ADDRESS	STOP	CONTENTS OF ADDRESS	NOTES
	/					The symbols x and y may be considered capitalized, since they represent Observed Data.	
	/	☒					
		0 1 3 2	U	0 0 3 6	/	start loop	
		1 3 3	P	0 1 0 3	/	Σx	from ~~0040~~
		1 3 4	M	0 1 0 3	/	Σx	
		1 3 5	X D	[ ]	/	n	
		1 3 6	C	0 1 0 5	/	$(\Sigma x)^2/n$	ss_x
		1 3 7	B	0 1 0 4	/	Σx^2	
		1 3 8	S	0 1 0 5	/	$(\Sigma x)^2/n$	
		1 3 9	C	0 1 0 6	/	ss_x	
		1 4 0	P	0 1 0 7	/	Σy	
		1 4 1	M	0 1 0 7	/	Σy	
		1 4 2	X D	[ ]	/	n	
		1 4 3	C	0 1 0 9	/	$(\Sigma y)^2/n$	s.s.y
		1 4 4	B	0 1 0 8	/	Σy^2	
		1 4 5	S	0 1 0 9	/	$(\Sigma y)^2/n$	
		1 4 6	C	0 1 1 0	/	s.s.y	
		1 4 7	P	0 1 0 3	/	Σx	
		1 4 8	M	0 1 0 7	/	Σy	
		1 4 9	X D	[ ]	/	n	sp_{xy}
		1 5 0	C	0 1 1 2	/	$(\Sigma x)(\Sigma y)/n$	
		1 5 1	B	0 1 1 1	/	Σxy	
		1 5 2	S	0 1 1 2	/	$(\Sigma x)(\Sigma y)/n$	
		1 5 3	H	0 1 1 3	/	sp_{xy}	
		1 5 4	D	0 1 0 6	/	$s.s._x$	b
		1 5 5	C	0 1 1 4	/	b	
		1 5 6	P	0 1 1 4	/	b	
		1 5 7	M	0 1 0 3	/	Σx	
		1 5 8	C	0 1 1 5	/	temp bΣx	
		1 5 9	B	0 1 0 7	/	Σy	a
		1 6 0	S	0 1 1 5	/	temp bΣx	
		1 6 1	X D	[ ]	/	n	
		1 6 2	C	0 1 1 5	/	a	
		1 6 3	X B	[ ]	/	n	d.f.

RM LP-12

PRINTED IN U.S.A.

Royal McBee Corporation
DATA PROCESSING DIV.
PORT CHESTER, NEW YORK

/ = CONDITIONAL STOP CODE

PREPARED FOR: LGP-30 USERS' ORGANIZATION - POOL | PAGE 5 OF 15

JOB NO. | PROGRAM NO. F2-38a | PROGRAM PREPARED BY: R.L.Stearman | PROGRAM CHECKED BY: POOL Review | DATE Dec. 1, 1959.

PROBLEM: "Bivariate Linear Regression Straight Line Fit to Observed Data": $y = a + bx$ | TRACK

PROGRAM INPUT CODES	STOP	LOCATION	INSTRUCTION OPERATION	INSTRUCTION ADDRESS	STOP	CONTENTS OF ADDRESS	NOTES
	/					The symbols for x and y may be considered capitalized, since ~~they represent Observed Data.~~	
	/	☒					
		02 00	S	01 01	/	1	d.f.
		00 01	S	01 01	/	1	
		00 02	C	01 18	/	d.f.	
		00 03	P	01 13	/	☒ $s.p._{xy}$	
		00 04	M	01 13	/	$s.p._{xy}$	→ $s.s_{reg}$
		00 05	D	01 06	/	ss_x	
		00 06	C	01 16	/	ss_{reg}	
		00 07	B	01 10	/	☒ ss_y	→ $s.s_{yx}$
		00 08	S	01 16	/	ss_{reg}	
		00 09	H	01 17	/	$s.s._{yx}$	
		00 10	D	01 18	/	d.f.	
		00 11	C	01 19	/	☒ $s^2_{y.x}$	→ s^2_{yx}
		00 12	B	01 16	/	ss_{reg}	→ F
		00 13	D	01 19	/	$s^2_{y.x}$	
		00 14	C	01 20	/	F	
		00 15	P	01 06	/	☒ ss_x	
		00 16	M	01 10	/	ss_y	
		00 17	XR	00 00	/	sq. rt.	
		00 18	C	01 02	/	temp	r
		00 19	B	01 13	/	☒ $s.p._{xy}$	
		01 00	D	01 02	/	$s.s_x$ $s.s_y$	
		01 01	C	01 02	/	r	
		01 02	XE	00 00	/		~~exit fl. pt.~~
		01 03	XU	[ ] [ ]	/	☒	return to program
		01 04			/		
		01 05			/		
		01 06			/		
		01 07			/	☒	
		01 08			/		
		01 09			/		
		01 10			/		
		01 11			/	☒	