1 Ja, hoe groter het BNP per hoofd in euro's, hoe minder werkzaam in de agrarische sector.

2a Negatieve correlatie.
2b Geen correlatie.
2c Positieve correlatie.
2d Positieve correlatie.
2e Negatieve correlatie.
2f Geen correlatie.
2g Positieve correlatie.
2h Negatieve correlatie.
2i Positieve correlatie.

3a Ja, een positieve correlatie.
3b 75 inch; 73,6 inch.
3c 62,3 inch; 63,9 inch.
3d 69 inch.
3e 3 inch.
3f 68 inch; 62,8 inch.
3g 65,5 inch.
3h 1,5 inch.
3i 71,5 inch; 1,5 inch.

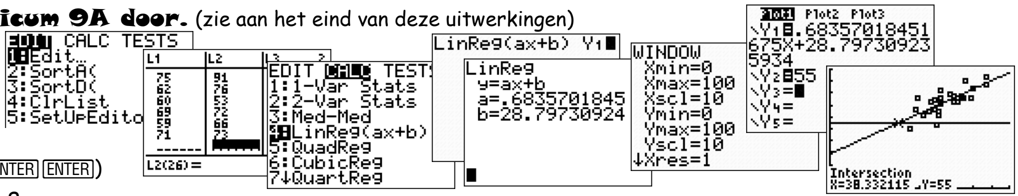
4a Sterke positieve correlatie.
4b Zwakke positieve correlatie.
4c Sterke negatieve correlatie.

5a Zwakke negatieve correlatie.
5b Sterke positieve correlatie.
5c Sterke positieve correlatie.
5d Geen correlatie.
5e Sterke of zwakke positieve correlatie.
5f Sterke positieve correlatie.
5g Sterke positieve correlatie.
5h Zwakke negatieve of geen correlatie.

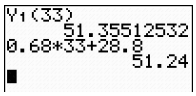
6a Door de verschillende eenheden langs de verticale assen.
6b De correlatiesterken zijn gelijk. (in beide figuren zijn dezelfde punten uitgezet)

*** **Neem GR - practicum 9A door.** (zie aan het eind van deze uitwerkingen)

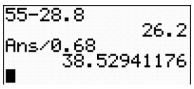
7a Zie de plot hiernaast.
7b Sterk positief.
7c LinReg(ax+b) Y1
(STAT ▸ 4 VARS ▸ ENTER ENTER ENTER)
geeft $\hat{y} \approx 0{,}68x + 28{,}8$.



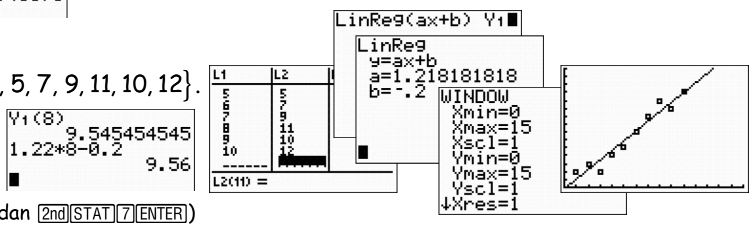
7d $x = 33 \Rightarrow \hat{y} = \hat{y}(33) \approx 0{,}68 \cdot 33 + 28{,}8 \approx 51$.



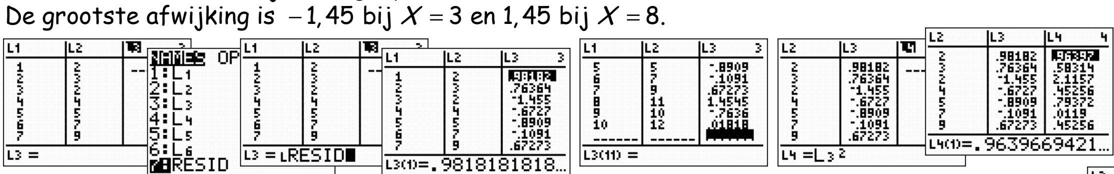
7e $\hat{y} > 55$ (intersect en een plot of)
$0{,}68x + 28{,}8 > 55$
$0{,}68x > 26{,}2$
$x > 38{,}5$. Dus vanaf CE-score 39.



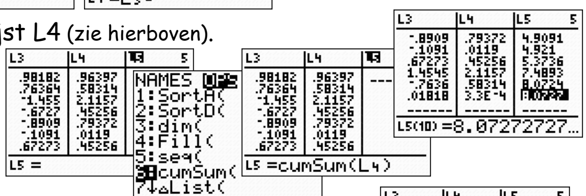
8a $L1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ en $L2 = \{2, 3, 2, 4, 5, 7, 9, 11, 10, 12\}$.
LinReg(ax+b) (STAT ▸ 4) geeft $\hat{y} \approx 1{,}22x - 0{,}2$.
8b $x = 8$ geeft $\hat{y} = \hat{y}(8) \approx 9{,}55$.



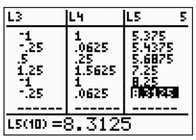
8c Voer de residuen in in lijst L3. (ga op L3 staan en dan 2nd STAT 7 ENTER)
De grootste afwijking is $-1{,}45$ bij $x = 3$ en $1{,}45$ bij $x = 8$.



8d Zet de kwadraten van lijst L3 (de kwadraten van de residuen) in lijst L4 (zie hierboven).
(ga op de naam L4 staan en toets L3[2] in met 2nd 3 x^2 ENTER)
Tel de kwadraten uit lijst L4 cumulatief op in L5.
(ga op L5 staan en toets cumSum(L4) in met 2nd STAT ▸ 6 2nd 4) ENTER)
De som van de kwadraten van de residuen is (ongeveer) 8,0727.



8e Voer $y_1 = 1{,}25x - 0{,}25$ in en (de residuen in) $L3 = L2 - y_1(L1)$.
(ga op de naam L3 staan en toets $L2 - y_1(L2)$ in met 2nd 2 − VARS ▸ ENTER ENTER (2nd 1) ENTER)
Zet de kwadraten van L3 in L4 en tel de kwadraten uit L4 cumulatief op in L5. (zoals in 8d)
De som van de kwadraten van de residuen is 8,3125 (ga dit zelf na) $\Rightarrow$ inderdaad groter dan in 8d.



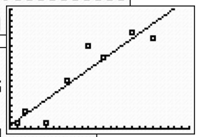
*** **Neem GR - practicum 9B door.** (zie aan het eind van deze uitwerkingen)

```
L1   L2
2    3
13   12
8    8
17   16
20   16
------
L2(9) =
```

```
EDIT CALC TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
```

```
LinReg(ax+b) Y1
LinReg
 y=ax+b
 a=.8564971751
 b=.5062146893
LinReg(ax+b) L2,
L1,Y2
LinReg
 y=ax+b
 a=1.002645503
 b=.8518518519
```

```
Plot1 Plot2 Plot3
\Y1=.85649717514
124X+.5062146892
656
\Y2=1.0026455026
455X+.8518518518
519
\Y3=
```



```
WINDOW
 Xmin=0
 Xmax=25
 Xscl=1
 Ymin=0
 Ymax=20
 Yscl=1
↓Xres=1
```

9a $\widehat{Y} \approx 0{,}86X + 0{,}51$ en $\widehat{X} \approx 1{,}00Y + 0{,}85$.

9b Zet alleen de regressielijn van Y op X aan.
De plot van het spreidingsdiagram met de regressielijn van Y op X zie je rechts hiernaast.

9c $Y = 10 \Rightarrow \widehat{X} = \widehat{X}(10) \approx 10{,}88$.

```
Y2(10)
      10.87830688
1.00*10+0.85
            10.85
```

9d $X = 10 \Rightarrow \widehat{Y} = \widehat{Y}(10) \approx 9{,}07$.

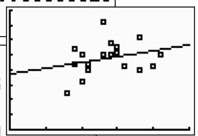
```
Y1(10)
      9.071186441
0.86*10+0.51
             9.11
```

```
L1   L2
20   21
16   21
11   20
18   19
10   15
8    12
L2(19) =
```

```
EDIT CALC TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
```

```
LinReg(ax+b) Y1
LinReg
 y=ax+b
 a=.389546139
 b=18.64228866
LinReg(ax+b) L2,
L1,Y2
LinReg
 y=ax+b
 a=.1946265449
 b=8.95088662
```

```
Plot1 Plot2 Plot3
\Y1=.38954613895
466X+18.64228866
4228
\Y2=.19462654486
835X+8.950886620
0967
\Y3=
```



```
WINDOW
 Xmin=0
 Xmax=25
 Xscl=5
 Ymin=0
 Ymax=40
 Yscl=5
↓Xres=1
```

10a $\widehat{Y} \approx 0{,}39X + 18{,}64$ en $\widehat{X} \approx 0{,}19Y + 8{,}95$.

10b Zet alleen de regressielijn van Y op X aan.
De plot van het spreidingsdiagram met de regressielijn van Y op X zie je rechts hiernaast.

10c $Y = 18 \Rightarrow \widehat{X} = \widehat{X}(18) \approx 12$.

10d $X = 18 \Rightarrow \widehat{Y} = \widehat{Y}(18) \approx 26$.

```
Y2(18)
      12.45416443
Y1(18)
      25.65411917
```

11 De verticale afstand tussen een punt en de regressielijn is $Y - \widehat{Y}$ of $\widehat{Y} - Y$.
Dus (verticale afstand)$^2 = (Y - \widehat{Y})^2 = (\widehat{Y} - Y)^2 = (ax + b - y)^2$.
Voor de regressielijnen moet $\sum$(verticale afstand)^2 minimaal zijn, dus $\sum_{i=1}^{n}(ax_i + b - y_i)^2$ moet minimaal zijn.

ERRATA: a op bladzijde 63 in de bovenste rode regel van Theorie A,
a_Y op bladzijde 65 in Theorie A de eerste rode regel,
a op bladzijde 65 in de derde regel van onder in het voorbeeld en
a op bladzijde 76 in terugblik
moet zijn $a = a_Y = \dfrac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2}$

12

X	Y	X^2	XY
1	12	1	12
2	8	4	16
2	10	4	20
3	7	9	21
6 +	4 +	36 +	24 +
$\sum X = 14$	$\sum Y = 41$	$\sum X^2 = 54$	$\sum XY = 93$

```
1+2+2+3+6
               14
1²+2²+2²+3²+6²
               54
```

```
12+8+10+7+4
               41
1*12+2*8+2*10+3*
7+6*4
               93
```

$n = 5$, $\overline{X} = \frac{\sum X}{n} = \frac{14}{5} = 2{,}8$ en $\overline{Y} = \frac{\sum Y}{n} = \frac{41}{5} = 8{,}2$;

$a_Y = \dfrac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2} = \dfrac{5 \cdot 93 - 14 \cdot 41}{5 \cdot 54 - 14^2} \approx -1{,}47$ en

$b = \overline{Y} - a\overline{X} \approx 8{,}2 + 1{,}473 \cdot 2{,}8 \approx 12{,}32$.

Dus $\widehat{Y} \approx -1{,}47X + 12{,}32$.

```
(5*93-14*41)/(5*
54-14²)
     -1.472972973
8.2-Ans*2.8
      12.32432432
8.2+1.473*2.8
          12.3244
```

13a

X	Y	X^2	XY
51	90	51^2	$51 \cdot 90$
46	96	46^2	$46 \cdot 96$
63	105	63^2	$63 \cdot 105$
51	110	51^2	$51 \cdot 110$
33	113	33^2	$33 \cdot 113$
28	120	28^2	$28 \cdot 120$
22	131	22^2	$22 \cdot 131$
31 +	124 +	31^2 +	$31 \cdot 124$ +
$\sum X = 325$	$\sum Y = 889$	$\sum X^2 = 14\,605$	$\sum XY = 35\,046$

```
51+46+63+51+33+2
8+22+31
              325
90+96+105+110+11
3+120+131+124
              889
```

```
51²+46²+63²+51²+
33²+28²+22²+31²
            14605
51*90+46*96+63*1
05+51*110+33*113
+28*120+22*131+3
1*124
            35046
```

```
325/8
           40.625
889/8
          111.125
```

$n = 8$, $\overline{X} = \frac{\sum X}{n} = \frac{325}{8} = 40{,}625$ en $\overline{Y} = \frac{\sum Y}{n} = \frac{889}{8} = 111{,}125$;

$a_Y = \dfrac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2} = \dfrac{8 \cdot 35046 - 325 \cdot 889}{8 \cdot 14605 - 325^2} \approx -0{,}763$ en

$b = \overline{Y} - a\overline{X} \approx 111{,}125 + 0{,}763 \cdot 40{,}625 \approx 142{,}122$.

Dus $\widehat{Y} \approx -0{,}76X + 142$.

```
(8*35046-325*889
)/(8*14605-325²)
     -.7629959875
111.125-Ans*40.6
25
       142.121712
```

13b $X = 42 \Rightarrow \widehat{Y} \approx -0{,}76 \cdot 42 + 142 \approx 110{,}1$. Dus (ongeveer) 110 uur zonneschijn.

```
-0.76*42+142
           110.08
```

14a

X	Y	X^2	Y^2	XY
5	7	5^2	7^2	$5 \cdot 7$
8	12	8^2	12^2	$8 \cdot 12$
7	6	7^2	6^2	$7 \cdot 6$
3	2	3^2	2^2	$3 \cdot 2$
12	15	12^2	15^2	$12 \cdot 15$
16	20	16^2	20^2	$16 \cdot 20$
19 +	19 +	19^2 +	19^2 +	$19 \cdot 19$ +
$\sum X = 70$	$\sum Y = 81$	$\sum X^2 = 908$	$\sum Y^2 = 1219$	$\sum XY = 1040$

```
5+8+7+3+12+16+19
               70
7+12+6+2+15+20+1
9
               81
5²+8²+7²+3²+12²+
16²+19²
              908
7²+12²+6²+2²+15²
+20²+19²
             1219
5*7+8*12+7*6+3*2
+12*15+16*20+19*
19
             1040
```

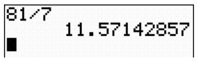
```
L1   L2   L3
5    7
8    12
7    6
3    2
12   15
16   20
19   19
L3(1)=
```

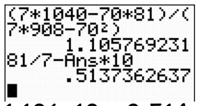
```
NAMES OPS MATH
1:min(
2:max(
3:mean(
4:median(
5:sum(
6:prod(
```

```
sum(L1)
               70
sum(L2)
               81
sum(L1²)
              908
```

```
sum(L2²)
             1219
sum(L1*L2)
             1040
mean(L1)
               10
```

14a $n = 7,\ \overline{X} = \frac{\sum X}{n} = \frac{70}{7} = 10$ en $\overline{Y} = \frac{\sum Y}{n} = \frac{81}{7} \approx 11{,}57;$



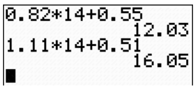


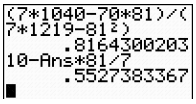
$a_Y = \frac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{7 \cdot 1040 - 70 \cdot 81}{7 \cdot 908 - 70^2} \approx 1{,}106$ en $b_Y = \overline{Y} - a_Y \overline{X} \approx \frac{81}{7} - 1{,}106 \cdot 10 \approx 0{,}514 \Rightarrow \widehat{Y} \approx 1{,}11X + 0{,}51.$

$a_X = \frac{n\sum XY - \sum X \cdot \sum Y}{n\sum Y^2 - (\sum Y)^2} = \frac{7 \cdot 1040 - 70 \cdot 81}{7 \cdot 1219 - 81^2} \approx 0{,}816$ en $b_X = \overline{X} - a_X \overline{Y} \approx 10 - 0{,}816 \cdot \frac{81}{7} \approx 0{,}553 \Rightarrow \widehat{X} \approx 0{,}82Y + 0{,}55.$

14b $Y = 14 \Rightarrow \widehat{X} \approx 0{,}82 \cdot 14 + 0{,}55 \approx 12{,}0.$

14c $X = 14 \Rightarrow \widehat{Y} \approx 1{,}11X + 0{,}51 \approx 16{,}0.$





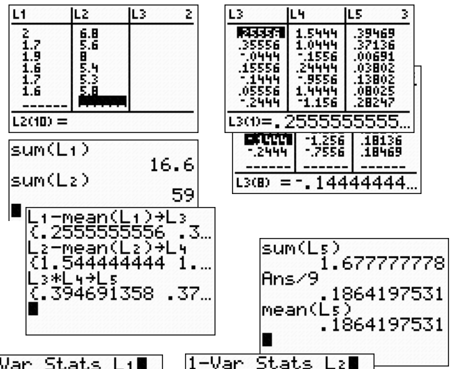
15a Voor een punt (X, Y) in (kwadrant) I geldt: $X > \overline{X}$ en $Y > \overline{Y} \Rightarrow X - \overline{X} > 0$ en $Y - \overline{Y} > 0.$

15b Voor een punt (X, Y) in (kwadrant) II geldt: $X - \overline{X} < 0$ en $Y - \overline{Y} > 0.$

15c Voor een punt (X, Y) in (kwadrant) III geldt: $(X - \overline{X}) < 0$ en $(Y - \overline{Y}) < 0 \Rightarrow (X - \overline{X})(Y - \overline{Y}) > 0.$

15d Bij een positieve correlatie liggen waarschijnlijk de meeste punten in (kwadrant) III en (kwadrant) I.

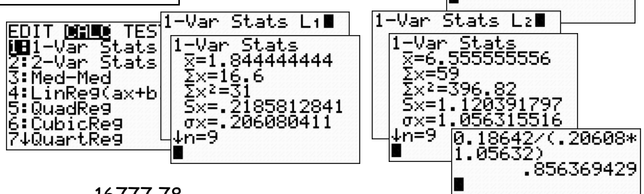
atleet	X	Y	$X - \overline{X}$	$Y - \overline{Y}$	$(X - \overline{X})(Y - \overline{Y})$
A	2,1	8,1	0,25556	1,5444	0,39469
B	2,2	7,6	0,35556	1,0444	0,37136
C	1,8	6,4	–0,0444	–0,1556	0,00691
D	2,0	6,8	0,15556	0,24444	0,03802
E	1,7	5,6	–0,1444	–0,9556	0,13802
F	1,9	8,0	0,05556	1,4444	0,08025
G	1,6	5,4	–0,2444	–1,156	0,28247
H	1,7	5,3	–0,1444	–1,256	0,18136
I	1,6	5,8	–0,2444	–0,7556	0,18469
	$\sum X = 16{,}6$ +	$\sum Y = 59{,}0$ +			1,67778 +



16a De covariantie $\sigma_{XY} = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{n} \approx 0{,}18642 \approx 0{,}19.$

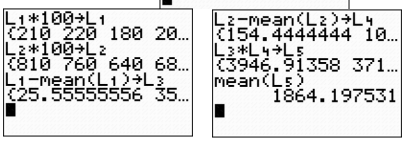
16b $\sigma_X \approx 0{,}20608;\ \sigma_Y \approx 1{,}05632$ en $\sigma_{XY} \approx 0{,}18642.$

$\frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \approx \frac{0{,}18642}{0{,}20608 \cdot 1{,}05632} \approx 0{,}86.$



16c $\overline{X} = \frac{1660}{9},\ \overline{Y} = \frac{5900}{9}$ en $\sum(X - \overline{X})(Y - \overline{Y}) \approx 16\,777{,}78 \Rightarrow \sigma_{XY} \approx \frac{16777{,}78}{9} \approx 1864{,}2.$

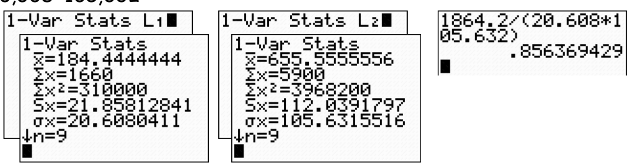
16d De covariantie σ_{XY} blijkt afhankelijk van de gebruikt eenheid.



16e $\sigma_X \approx 20{,}608;\ \sigma_Y \approx 105{,}632$ en $\sigma_{XY} \approx 1864{,}2 \Rightarrow \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \approx \frac{1864{,}2}{20{,}608 \cdot 105{,}632} \approx 0{,}86.$

De uitkomst is hetzelfde als bij 16b,

dus $\frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$ blijkt onafhankelijk van de gebruikte eenheid.



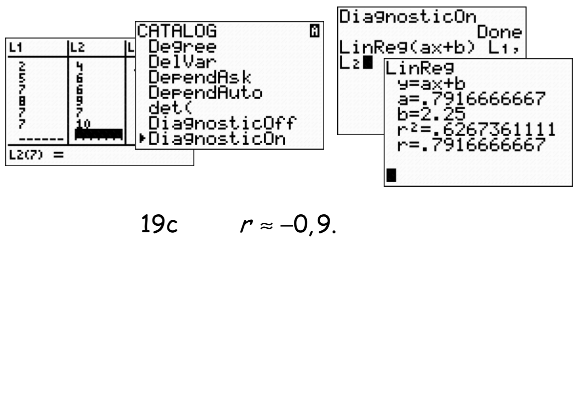
17 $\sigma_{XY} = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{n} = \frac{\sum(X \cdot Y - X \cdot \overline{Y} - \overline{X} \cdot Y + \overline{X} \cdot \overline{Y})}{n}$

$= \frac{\sum(X \cdot Y)}{n} - \frac{\sum(X \cdot \overline{Y})}{n} - \frac{\sum(\overline{X} \cdot Y)}{n} + \frac{\sum(\overline{X} \cdot \overline{Y})}{n} = \overline{X \cdot Y} - \frac{\sum(X)}{n} \cdot \overline{Y} - \overline{X} \cdot \frac{\sum(Y)}{n} + \frac{n \cdot \overline{X} \cdot \overline{Y}}{n}$

$= \overline{X \cdot Y} - \overline{X} \cdot \overline{Y} - \overline{X} \cdot \overline{Y} + \overline{X} \cdot \overline{Y} = \overline{XY} - \overline{X} \cdot \overline{Y}.$

*** **Neem GR - practicum 9C door.** (zie de schermen hiernaast)

18 Voer $L1 = \{2, 5, 7, 8, 7, 7\}$ en $L2 = \{4, 6, 6, 9, 7, 10\}$ in op de GR.
Kies DiagnosticOn in het CATALOG-menu.
LinReg(ax+b) L1, L2 geeft $\widehat{Y} \approx 0{,}86X + 0{,}51$ met pmcc $r \approx 0{,}79.$



19a $r \approx 0{,}9.$ 19b $r \approx 0{,}4.$ 19c $r \approx -0{,}9.$

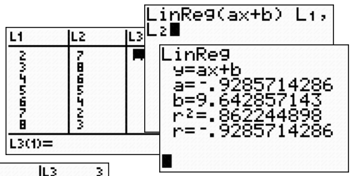
20a $r \approx 0{,}2.$ 20b $r \approx 0{,}9.$

21 $r = 1$. (volkomen positieve correlatie)

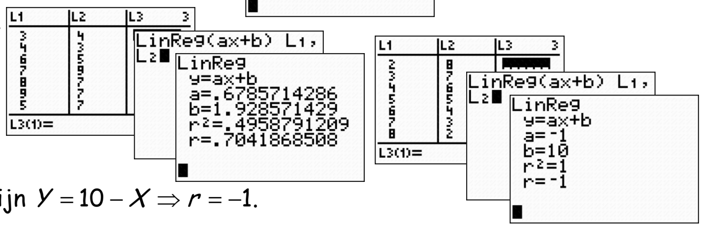
22 $Y = 25 - X \Rightarrow \overline{Y} = \overline{25 - X} = 25 - \overline{X} = 25 - 18{,}6 = 6{,}4$; $\sigma_Y = \sigma_X = 2{,}9$ en $r = 1$.

23a $r \approx 0{,}3$. 23b $r \approx 0{,}1$. 23c $r \approx 0{,}7$.

24a Tabel I: (DiagnosticOn is eerder gebeurd)
Voer $L1 = \{2, 3, 4, 5, 6, 7, 8\}$ en $L2 = \{7, 8, 6, 5, 4, 2, 3\}$ in.
LinReg(ax+b) geeft $\hat{Y} \approx ....$ met pmcc $r \approx -0{,}93$.



Tabel II:
$L1 = \{3, 4, 6, 7, 8, 9, 5\}$ en $L2 = \{4, 3, 5, 9, 7, 7, 7\}$.
LinReg(ax+b) geeft $\hat{Y} \approx ....$ met pmcc $r \approx 0{,}70$.
Tabel III:
$L1 = \{2, 3, 4, 5, 6, 7, 8\}$ en $L2 = \{8, 7, 6, 5, 4, 3, 2\}$.
LinReg(ax+b) geeft $\hat{Y} \approx ....$ met pmcc $r = -1$.

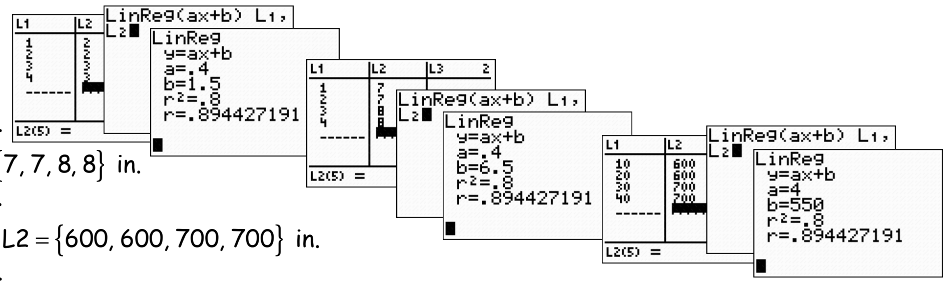


24b Alle punten (X, Y) liggen op de rechte (dalende) lijn $Y = 10 - X \Rightarrow r = -1$.

25a r verandert niet.
In de formule $r = \dfrac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sigma_X \cdot \sigma_Y}$ kunnen X en Y verwisseld worden zonder dat de uitkomst van r verandert.

25b r verandert niet.
In de formule $r = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$ verandert er niets als elke Y-waarde met 2 verminderd wordt.

25c r verandert niet.
In de formule $r = \dfrac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sigma_X \cdot \sigma_Y}$ worden $\overline{XY}$, $\overline{X}$ en σ_X 10 keer zo groot als elke X-waarde met 10 wordt vermeniguldigd. Dus de teller en noemer worden beide 10 keer zo groot en r blijft dus gelijk.

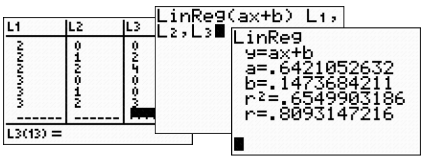


26a Voer $L1 = \{1, 2, 3, 4\}$
en $L2 = \{2, 2, 3, 3\}$ in.
LinReg(ax+b) geeft $r \approx 0{,}89$.

26b Voer $L1 = \{1, 2, 3, 4\}$ en $L2 = \{7, 7, 8, 8\}$ in.
LinReg(ax+b) geeft $r \approx 0{,}89$.

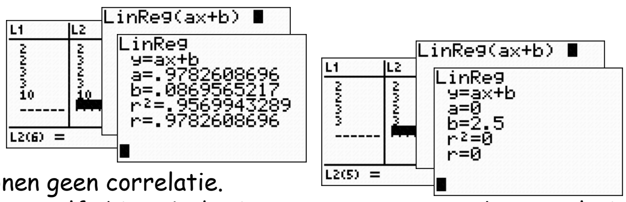
26c Voer $L1 = \{10, 20, 30, 40\}$ en $L2 = \{600, 600, 700, 700\}$ in.
LinReg(ax+b) geeft $r \approx 0{,}89$.

27 Voer $L1 = \{0, 0, 0, 1, 1, 1, 2, 2, 2, 3, 3, 3\}$,
$L2 = \{0, 1, 2, 0, 1, 2, 0, 1, 2, 0, 1, 2\}$ en
$L3 = \{5, 2, 0, 3, 1, 0, 0, 2, 4, 0, 0, 3\}$ in.
LinReg(ax+b) L1, L2, L3 geeft $r \approx 0{,}81$.



28a Voer $L1 = \{2, 2, 3, 3, 10\}$ en $L2 = \{2, 3, 2, 3, 10\}$ in.
LinReg(ax+b) geeft $r \approx 0{,}98$.

28b Laat de waarneming (10, 10) weg.
LinReg(ax+b) geeft $r = 0$.



28c De vier overgebleven waarnemingen in 28b vertonen geen correlatie.
Door toevoeging van de uitschieter (10, 10) wordt r zelfs bijna 1, dus is er opeens zeer sterke correlatie.

29a Voer $L1 = \{2, 3, 4, 5, 6, 7, 8\}$ en $L2 = \{4, 5, 6, 7, 6, 5, 4\}$ in.
LinReg(ax+b) geeft de pmcc $= r = 0$.

29b Zie een plot van het spreidingsdiagram hiernaast.

29c Ja, er is een sterke samenhang tussen X en Y.
Er is zelfs een formule voor Y als functie van X, namelijk $Y = 7 - |X - 5|$.

29d $r = 0$ suggereert geen samenhang, maar r zegt alleen iets over lineaire samenhang en die is er niet.

30a Heterogeniteit (II). 30c Niet lineair (I). 30e Selectie (IV).
30b Selectie (IV). 30d Interpretatie (V).

31a leeftijd 31b welvaart 31c weer 31d lengte

32a Er is niet per individu gemeten.
32b Er is geen oorzaak-gevolg zoals de slijter suggereert.
Misschien is er andersom wel causaliteit, maar dat is niet uit een berekende correlatie te halen.

33a $\frac{\sum(X-\overline{X})^2}{n}=\frac{\sum(X^2-2X\overline{X}-\overline{X}^2)}{n}=\frac{\sum X^2}{n}-2\cdot\frac{\sum X}{n}\cdot\overline{X}+\frac{\sum\overline{X}^2}{n}=\frac{\sum X^2}{n}-2\cdot\overline{X}\cdot\overline{X}+\frac{n\cdot\overline{X}^2}{n}=\frac{\sum X^2}{n}-2\cdot\overline{X}^2+\overline{X}^2=\frac{\sum X^2}{n}-\overline{X}^2.$

33b $r\cdot\frac{\sigma_Y}{\sigma_X}=\frac{\overline{XY}-\overline{X}\cdot\overline{Y}}{(\sigma_X)^2}=\frac{\frac{\sum XY}{n}-\frac{\sum X}{n}\cdot\frac{\sum Y}{n}}{\frac{\sum(X-\overline{X})^2}{n}}=\frac{n^2\cdot\left(\frac{\sum XY}{n}-\frac{\sum X}{n}\cdot\frac{\sum Y}{n}\right)}{n^2\cdot\frac{\sum(X-\overline{X})^2}{n}}$ (gebruik in de noemer de herleiding uit 33a)

$=\frac{n\sum XY-\sum X\cdot\sum Y}{n^2\cdot\left(\frac{\sum X^2}{n}-\overline{X}^2\right)}=\frac{n\sum XY-\sum X\cdot\sum Y}{n\sum X^2-n^2\cdot\overline{X}^2}=\frac{n\sum XY-\sum X\cdot\sum Y}{n\sum X^2-n^2\cdot\left(\frac{\sum X}{n}\right)^2}=\frac{n\sum XY-\sum X\cdot\sum Y}{n\sum X^2-(\sum X)^2}.$

34a Stel $\hat{Y}=aX+b$ met $a=r\cdot\frac{\sigma_Y}{\sigma_X}=0{,}4\cdot\frac{8000}{4}=800.$

```
0.4*8000/4
                800
```

$\left.\begin{array}{l}\hat{Y}=800X+b\\ Z(11,10\,000)\end{array}\right\}\Rightarrow 10\,000=800\cdot 11+b\Rightarrow b=1200.$ Dus $\hat{Y}=800X+1200.$

```
10000-800*11
               1200
```

$X=16$ geeft $\hat{Y}=800\cdot 16+1200=14\,000.$ Zijn inkomen wordt geschat op 14 000 dollar.

34b $X=5$ geeft $\hat{Y}=800\cdot 5+1200=5200.$ Zijn inkomen wordt geschat op 5 200 dollar.

```
800*16+1200
              14000
800*5+1200
               5200
```

35a Stel $\hat{Y}=aX+b$ met $a=r\cdot\frac{\sigma_Y}{\sigma_X}=0{,}8\cdot\frac{20}{20}=0{,}8\cdot 1=0{,}8.$

```
105-0.8*105
                 21
0.8*125+21
                121
```

$\left.\begin{array}{l}\hat{Y}=0{,}8X+b\\ Z(105,105)\end{array}\right\}\Rightarrow 105=0{,}8\cdot 105+b\Rightarrow b=21.$ Dus $\hat{Y}=0{,}8X+21$ en $X=125$ geeft $\hat{Y}=0{,}8\cdot 125+21=121.$

35b $X=95$ geeft $\hat{Y}=0{,}8\cdot 95+21=97.$

```
0.8*95+21
                 97
```

35c Zie Frits in 35a: $0{,}8\cdot 125+21=0{,}8\cdot 125+0{,}2\cdot 105<0{,}8\cdot 125+0{,}2\cdot 125=1\cdot 125.$
Zie Frans in 35b: $0{,}8\cdot 95+21=0{,}8\cdot 95+0{,}2\cdot 105>0{,}8\cdot 95+0{,}2\cdot 95=1\cdot 95.$

```
0.8*105
                 84
Ans+21
                105
```

36a $\left.\begin{array}{l}a_X=r\cdot\frac{\sigma_X}{\sigma_Y}\\ a_X=r\end{array}\right\}\Rightarrow\frac{\sigma_X}{\sigma_Y}=1\Rightarrow\sigma_X=\sigma_Y.$

36b $\left.\begin{array}{l}a_X\cdot a_Y=r^2\\ a_X=0{,}8\\ a_Y=2{,}1\end{array}\right\}\Rightarrow r^2=0{,}8\cdot 2{,}1=1{,}68.$

$r^2=1{,}68>1\Rightarrow$ er is een fout gemaakt ($-1\le r\le 1$).

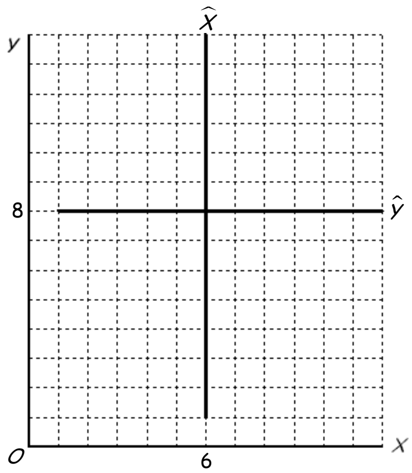
36c $\left.\begin{array}{l}a_X\cdot a_Y=r^2\\ a_X=3\end{array}\right\}\Rightarrow\left.\begin{array}{l}a_Y=\frac{1}{3}r^2\\ r^2\le 1\end{array}\right\}\Rightarrow a_Y$ is maximaal $\frac{1}{3}$.

36d $\left.\begin{array}{l}a_X\cdot a_Y=r^2\\ a_X=-0{,}7\\ a_Y=0{,}7\end{array}\right\}\Rightarrow r^2=-0{,}49<0$ kan niet.

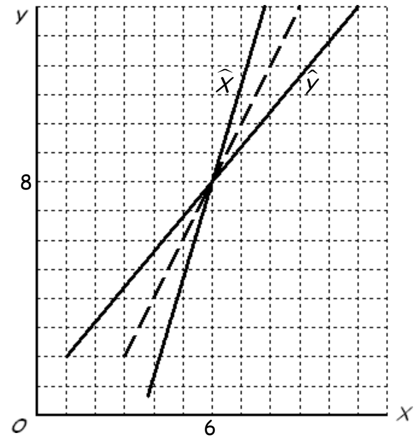
Dus a_Y kan niet 0,7 zijn.

36e $\left.\begin{array}{l}a_X\cdot a_Y=r^2\\ a_X=a_Y\end{array}\right\}\Rightarrow r^2={a_X}^2={a_Y}^2\Rightarrow r=a_X=a_Y.$

Volgens vraag 36a is dus $\sigma_X=\sigma_Y$.



37a $r=0\Rightarrow a_Y=0$ én $a_X=0.$
Dus een horizontale en verticale lijn door het zwaartepunt $Z(6,8)$. (zie de figuur hierboven)



37b $r = 0{,}6 \Rightarrow a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}6 \cdot \frac{2}{1} = 1{,}2$ én $a_X = r \cdot \frac{\sigma_X}{\sigma_Y} = 0{,}6 \cdot \frac{1}{2} = 0{,}3$.

$\hat{Y} = 1{,}2(X-6)+8$ of $\hat{Y} = 1{,}2X + 0{,}8$ door (6, 8) en (1, 2).

```
1.2*-6+8
                .8
```

$\widehat{X} = 0{,}3(Y-8)+6$ of $\widehat{X} = 0{,}3Y + 3{,}6$ of $Y = \frac{\widehat{X}-3{,}6}{0{,}3}$ door (6, 8) en (9, 18).

```
0.3*-8+6
               3.6
```

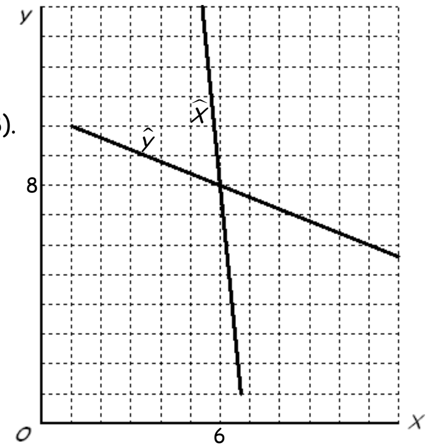
(zie de grafieken van $\hat{Y}$ en $\widehat{X}$ in de figuur hiernaast)

37c $r = 1 \Rightarrow a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 1 \cdot \frac{2}{1} = 2$ én $a_X = r \cdot \frac{\sigma_X}{\sigma_Y} = 1 \cdot \frac{1}{2} = 0{,}5$.

$\hat{Y} = 2(X-6)+8$ of $\hat{Y} = 2X - 4$ door (6, 8) en (2, 0).

$\widehat{X} = 0{,}5(Y-8)+6$ of $\widehat{X} = 0{,}5Y + 2$ of $Y = \frac{\widehat{X}-2}{0{,}5}$ door (6, 8) en (2, 0).

(zie de gestippelde regressielijn in de figuur hiernaast)



37d $r = -0{,}2 \Rightarrow a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = -0{,}2 \cdot \frac{2}{1} = -0{,}4$ én $a_X = r \cdot \frac{\sigma_X}{\sigma_Y} = -0{,}2 \cdot \frac{1}{2} = -0{,}1$.

$\hat{Y} = -0{,}4(X-6)+8$ of $\hat{Y} = -0{,}4X + 10{,}4$ door (6, 8) en (11, 6).

$\widehat{X} = -0{,}1(Y-8)+6$ of $\widehat{X} = -0{,}1Y + 6{,}8$ of $Y = \frac{-\widehat{X}+6{,}8}{0{,}1} = -10X + 68$ door (6, 8).

(zie de grafieken van $\hat{Y}$ en $\widehat{X}$ in de figuur hiernaast)

38a $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}7 \cdot \frac{8}{5} = \frac{5{,}6}{5} = \frac{11{,}2}{10} = 1{,}12$.

$\hat{Y} = 1{,}12(X-65)+75$ of $\hat{Y} = 1{,}12X + 2{,}2$.

```
0.7*8/5
              1.12
1.12*-65+75
               2.2
```

38b $a_X = r \cdot \frac{\sigma_X}{\sigma_Y} = 0{,}7 \cdot \frac{5}{8} = 0{,}4375$.

$\widehat{X} = 0{,}4375(Y-75)+65$ of $\widehat{X} = 0{,}4375Y + 32{,}1875$.

```
0.7*5/8
             .4375
Ans*-75+65
           32.1875
```

38c $X = 75$ geeft als voorspelling van de taalscore $\hat{Y} = 1{,}12 \cdot 75 + 2{,}2 \approx 86$.

38d $Y = 86$ geeft als voorspelling van de rekenscore $\widehat{X} = 0{,}4375(86-75)+65 \approx 70$.

```
1.12*75+2.2
              86.2
0.4375*(86-75)+6
5
           69.8125
```

39 $\hat{Y} = a_Y(X - 6{,}4) + 4{,}5$ én $X = 8{,}8$ geeft $\hat{Y} = 5{,}5$.

Dus $5{,}5 = a_Y(8{,}8-6{,}4)+4{,}5 \Rightarrow 1 = a_Y \cdot 2{,}4 \Rightarrow a_Y = \frac{1}{2{,}4}$.

$a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} \Rightarrow \text{pmcc} = r = a_Y \cdot \frac{\sigma_X}{\sigma_Y} = \frac{1}{2{,}4} \cdot \frac{0{,}8}{0{,}5} \approx 0{,}67$.

```
5.5-4.5
                 1
8.8-6.4
               2.4
1/2.4
      .4166666667
```

```
1/2.4*0.8/0.5
      .6666666667
```

40 $a_X \cdot a_Y = r^2 = 1 \Rightarrow a_X = \frac{1}{a_Y}$. Verder is $\hat{Y} = a_X(X - \overline{X}) + \overline{Y}$ en

$\widehat{X} = a_Y(Y - \overline{Y}) + \overline{X} \Rightarrow a_Y(Y-\overline{Y}) = \widehat{X} - \overline{X} \Rightarrow Y - \overline{Y} = \frac{1}{a_Y}(\widehat{X} - \overline{X}) \Rightarrow Y = \frac{1}{a_Y}(\widehat{X} - \overline{X}) + \overline{Y} = a_X(\widehat{X} - \overline{X}) + \overline{Y}$.

Beide lijnen gaan door $(\overline{X}, \overline{Y})$ en hebben rc $= a_X \Rightarrow$ beide regressielijnen vallen samen.

41 De voorspelling bij figuur 10.24a zal betrouwbaarder omdat daar de puntenwolk dichter bij de regressielijn ligt.

42a $\hat{Y}$ en d zijn onafhankelijk van elkaar, dus voor $Y = \hat{Y} + d$ geldt $\sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_d^2$.

42b Omdat bij $\hat{Y} = aX + b$ de variabelen a en b constant zijn, geldt: $\sigma_{\hat{Y}}^2 = \sigma_{aX}^2 = (a \cdot \sigma_X)^2 = a^2 \cdot \sigma_X^2$.

Hieruit volgt $\sigma_{\hat{Y}} = a\sigma_X$ of $\sigma_{\hat{Y}} = -a\sigma_X$.

Omdat $\sigma_{\hat{Y}}$ en σ_X beide postief zijn, geldt $\sigma_{\hat{Y}} = a\sigma_X$ voor $a > 0$ en $\sigma_{\hat{Y}} = -a\sigma_X$ voor $a < 0$.

42c $$\left.\begin{array}{l} \sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_d^2 \text{ (42a)} \\ \sigma_{\hat{Y}}^2 = a^2\sigma_X^2 \text{ (42b)} \end{array}\right\} \Rightarrow \sigma_Y^2 = a^2\sigma_X^2 + \sigma_d^2 \Rightarrow \sigma_d^2 = \sigma_Y^2 - a^2\sigma_X^2.$$

42d $$\left.\begin{array}{l} \sigma_d^2 = \sigma_Y^2 - a^2\sigma_X^2 \text{ (42c)} \\ a = r \cdot \frac{\sigma_Y}{\sigma_X} \end{array}\right\} \Rightarrow \sigma_d^2 = \sigma_Y^2 - r^2 \cdot \left(\frac{\sigma_Y}{\sigma_X}\right)^2 \cdot \sigma_X^2 = \sigma_Y^2 - r^2 \cdot \sigma_Y^2 = \sigma_Y^2(1 - r^2) \Rightarrow \sigma_d = \sigma_Y \cdot \sqrt{1 - r^2}.$$

43a $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}6 \cdot \frac{10}{5} = 0{,}6 \cdot 2 = 1{,}2 \Rightarrow \widehat{Y} = 1{,}2(X - 80) + 65$ of $\widehat{Y} = 1{,}2X - 31$.

```
1.2*-80+65
                -31
```

43b De standaardschattingsfout $\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} = 10 \cdot \sqrt{1 - 0{,}6^2} = 10 \cdot 0{,}8 = 8$.

```
√(1-0.6²)
                 .8
Ans*10
                  8
```

43c $x = 88$ geeft $\widehat{Y} = 1{,}2(88 - 80) + 65 = 74{,}6$.

```
74.6-8
               66.6
74.6+8
               82.6
```

43d Het 68%-interval is volgens vuistregel I van de normale verdeling het interval $\langle 74{,}6 - 8;\, 74{,}6 + 8 \rangle = \langle 66{,}6;\, 82{,}6 \rangle$.

43e $x = 70$ geeft $\widehat{Y} = 1{,}2(70 - 80) + 65 = 1{,}2 \cdot -10 + 65 = -12 + 65 = 53$.
Volgens vuistregel II ligt 95% van de waarnemingen in het interval $\langle 53 - 2 \cdot 8;\, 53 + 2 \cdot 8 \rangle = \langle 37,\, 69 \rangle$.

```
53-2*8
                 37
53+2*8
                 69
```

44a De standaardschattingsfout $\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} = 8 \cdot \sqrt{1 - 0{,}9^2} \approx 3{,}49$ (cm).

```
8*√(1-0.9²)
        3.487119155
```

44b $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}90 \cdot \frac{8}{8} = 0{,}90 \Rightarrow \widehat{Y} = 0{,}9(X - 178) + 178$ of $\widehat{Y} = 0{,}9X + 17{,}8$.
$X = 184 \Rightarrow \widehat{Y} = 0{,}9(184 - 178) + 178 = 183{,}4$. (de voorwaardelijke voorspelling van de lengte van de ander is 183,4 cm)

```
0.9*-178+178
               17.8
0.9*(184-178)+17
8
              183.4
```

44c 95% van de waarnemingen ligt in het interval $\langle 183{,}4 - 2 \cdot 3{,}49;\, 183{,}4 + 2 \cdot 3{,}49 \rangle = \langle 176{,}4;\, 190{,}4 \rangle$.

```
183.4-2*8*√(1-0.
9²)
        176.4257617
183.4+2*8*√(1-0.
9²)
        190.3742383
```

45a Bij volkomen correlatie is $r^2 = 1 \Rightarrow$ de standaardschattingsfout $\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} = \sigma_Y \cdot \sqrt{1 - 1} = \sigma_Y \cdot 0 = 0$.
Er is dus geen spreiding rond de modelwaarde $\widehat{Y} \Rightarrow$ alle punten liggen op de regressielijn.

45b Als er geen correlatie is, dan is $r = 0 \Rightarrow$ de standaardschattingsfout $\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} = \sigma_Y \cdot \sqrt{1 - 0} = \sigma_Y \cdot 1 = \sigma_Y$.

46 $P(X < 165) = \text{normalcdf}(-10^{99}, 165, 170, 10) \approx 0{,}309$.
$P(X > 185) = \text{normalcdf}(185, 10^{99}, 170, 10) \approx 0{,}067$.

```
normalcdf(-10^99
,165,170,10)
       .3085375322
normalcdf(185,10
^99,170,10)
       .0668072287
```

47a $P(Y < 160) = \text{normalcdf}(-10^{99}, 160, 170, 10) \approx 0{,}159$. Dus 15,9%.

```
normalcdf(-10^99
,160,170,10)
       .1586552596
Ans*100
```

47b $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}4 \cdot \frac{10}{12} = \frac{4}{12} = \frac{1}{3} \Rightarrow \widehat{Y} = \frac{1}{3}(X - 178) + 170$.
$X = 185 \Rightarrow \widehat{Y} = \frac{1}{3}(185 - 178) + 170 = 172\frac{1}{3}$.
De standaardschattingsfout $\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} = 10 \cdot \sqrt{1 - 0{,}4^2} \approx 9{,}165$ (cm).
$P(Y < 160) = \text{normalcdf}(-10^{99}, 160, 172\frac{1}{3}, 9.165...) \approx 0{,}089$. Dus 8,9%.

```
1/3(185-178)+170
        172.3333333
Ans▸Frac
              517/3
```

```
10*√(1-0.4²)
         9.16515139
normalcdf(-10^99
,160,172+1/3,Ans
)
       .0892034159
Ans*100
```

47c $X = 170 \Rightarrow \widehat{Y} = \frac{1}{3}(170 - 178) + 170 = 167\frac{1}{3}$.
$P(Y < 160) = \text{normalcdf}(-10^{99}, 160, 167\frac{1}{3}, 9.165...) \approx 0{,}212$. Dus 21,2%.

```
1/3(170-178)+170
        167.3333333
Ans▸Frac
              502/3
```

```
10*√(1-0.4²)
         9.16515139
normalcdf(-10^99
,160,167+1/3,Ans
)
       .2118170199
Ans*100
```

48a $\text{normalcdf}(-10^{99}, s, 70, 10) = 0{,}30$ (intersect) $\Rightarrow s \approx 64{,}76$.
of $0 = \text{normalcdf}(-10^{99}, s, 70, 10) - 0{,}30$ (SOLVE) $\Rightarrow s \approx 64{,}76$.
of $s = \text{invNorm}(0.30, 70, 10) \Rightarrow s \approx 64{,}76$.
Gehele scores $\Rightarrow$ bij de scores tot en met 64 wordt het proefwerk overgemaakt.

```
invNorm(0.3,70,1
0)
         64.7559949
```

```
MATH NUM CPX PRB
6↑fMin(
7:fMax(
8:nDeriv(
9:fnInt(
0:summatio
A:logBASE(
B:Solver…
```

```
EQUATION SOLVER
eqn:0=normalcdf(
L,R,M,S)-O
```

```
normalcdf(L,R…=0
  L=-1E99
▪ R=64.755995431…
  M=70
  S=10
  O=.3
  bound={-1E99,1…
▪ left-rt=0
```

48b $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}8 \cdot \frac{10}{20} = 0{,}8 \cdot \frac{1}{2} = 0{,}4 \Rightarrow \widehat{Y} = 0{,}4(X - 50) + 70$.
$X = 70 \Rightarrow \widehat{Y} = 0{,}4(70 - 50) + 70 = 78$.

```
0.4(70-50)+70
                 78
```

48c De standaardschattingsfout $\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} = 10 \cdot \sqrt{1 - 0{,}8^2} = 6$.
$\text{normalcdf}(-10^{99}, s, 78, 6) = 0{,}30$ (intersect) $\Rightarrow s \approx 74{,}85$.
of $0 = \text{normalcdf}(-10^{99}, s, 78, 6) - 0{,}30$ (SOLVE) $\Rightarrow s \approx 74{,}85$.
of $s = \text{invNorm}(0.30, 78, 6) \Rightarrow s \approx 74{,}85$.
Gehele scores $\Rightarrow$ bij de scores tot en met 74 behoort men tot de slechtste 30%.

```
10*√(1-0.8²)
                  6
invNorm(0.30,78,
6)
        74.85359694
```

```
EQUATION SOLVER
eqn:0=normalcdf(
L,R,M,S)-O
```

```
normalcdf(L,R…=0
  L=-1E99
▪ R=74.853597258…
  M=78
  S=6
  O=.3
  bound={-1E99,1…
▪ left-rt=0
```

48d $P(Y \le 64) = \text{normalcdf}(-10^{99}, 64.5, 78, 6) \approx 0{,}012$.
Dus ongeveer 1,2%.

49a Ratioschaal.
49b Nominale schaal.
49c Ratioschaal.
49d Nominale schaal.
49e Nominale schaal.
49f Nominale schaal.
49g Ratioschaal.
49h Nominale schaal.
49i Ratioschaal.
49j Intervalschaal.

50a Intervalschaal.
50b Ratioschaal.
50c Nominale schaal.
50d Ordinale schaal.

51 Bij intervalschaal en bij ratioschaal.

52a Groter en kleiner hebben bij nominaal niveau geen betekenis.
52b Bij ordinaal niveau kun je wel spreken van negatieve of positieve correlatie.
52c Ook bij intervalniveau kun je wel spreken van negatieve of positieve correlatie.

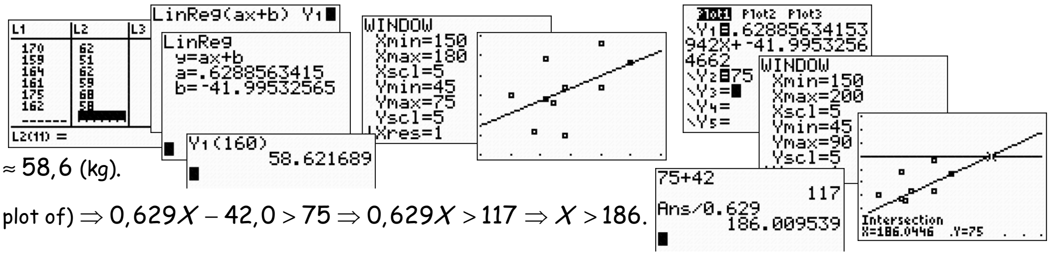
53ab ∗
53c $\hat{Y} = 0,0119X + 633,9$.
53d $\hat{Y} = 0,002699X + 604,705$.
$X = 20\,000$ geeft $\hat{Y} = 0,002699 \cdot 20\,000 + 604,705 = 1144,5348$.
$r^2 \approx 0,735$ en $\sigma_Y \approx 567,80 \Rightarrow \sigma_d = \sigma_Y \cdot \sqrt{1-r^2} \approx 567,8 \cdot \sqrt{1-0,735^2} \approx 292,29$.
Het 95%-betrouwbaarheidsinterval is $\langle 1144,5348 - 2 \cdot 292,29; 1144,5348 + 2 \cdot 292,29 \rangle \approx \langle 560; 1729 \rangle$.
53e $r \approx 0,8339$.
Positieve correlatie, dus bij groter gewicht hoort een hoger verbruik.
Een zware auto verbruikt dus meer brandstof dan een lichte auto.
53f $r \approx -0,8309$.
Negatieve correlatie, dus minder kilometers per liter bij een hogere prijs.
Hieruit volgt dat een dure auto meer brandstof verbruikt dan een goedkopere.
53g De correlatie is het sterkst tussen het verbruik bij 90 km/u en het verbruik bij 120 km/u.
Bij deze twee variabelen wordt vrijwel hetzelfde gemeten, dus een grote r is niet verbazingwekkend.

54a ∗
54b $\text{cs} = 1,07801 \cdot \text{so} - 0,9357 \Rightarrow \text{so} = 6,0$ geeft $\text{cs} = 1,07801 \cdot 6,0 - 0,9357 \approx 5,53$.
$r^2 \approx 0,7764$ en $\sigma_{cs} \approx 1,7715$.
$\sigma_d = \sigma_{cs} \cdot \sqrt{1-r^2} = 1,7715 \cdot \sqrt{1-0,7764} \approx 0,8377$.
invNorm(0.10,5.53,0.8377) $\approx 4,456$ en invNorm(0.90,5.53,0.8377) $\approx 6,604$.
Het 80%-betrouwbaarheidsinterval is $\langle 4,5; 6,6 \rangle$.
54c Het cs-cijfer heeft de sterkste correlatie met de tijd die aan het huiswerk is besteed.
54d Ongesplitst is $r = 0,8812$.
Splitsen in j/m geeft voor de meisjes $r = 0,9182$ en voor de jongens $r = 0,8716$.
Het maakt dus vrijwel geen verschil.
54e Ongesplitst is $r = 0,8379$.
Splitsen in wel/niet een bijbaan geeft voor niet een bijbaan $r = 0,725$ en voor wel een bijbaan $r = 0,9227$.
Er is een kleine invloed van splitsen in wel/niet een bijbaan.

55a ∗
55b $g_m = 0,66645l - 55,3746$ en $g_v = 0,441257l - 18,1834$.
$g_m = g_v \Rightarrow 0,66645l - 55,3746 = 0,441257l - 18,1834 \Rightarrow 0,225193l = 37,1912 \Rightarrow l = 165,15$.
Dus bij een lengte van 165 cm zijn beide voorspellingen gelijk.
55c Voor mannen met bril geldt $l_m = 0,4920325g + 140,732$
en voor mannen zonder bril geldt $l_z = 0,4814902g + 141,51$.
$g = 100$ geeft $l_m = 189,325$ en $l_z = 189,659$.
Er is dus vrijwel geen verschil.

Diagnostische toets

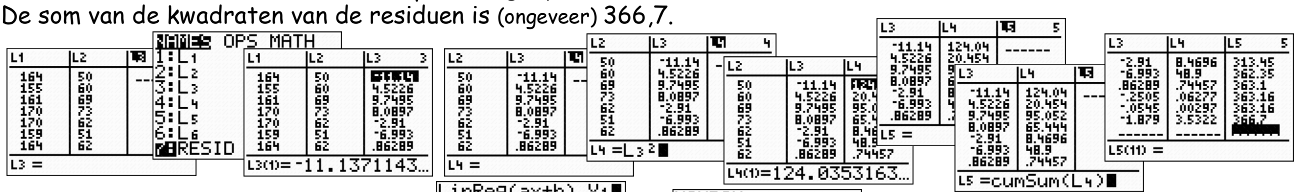
D1a Zie de plot hiernaast.
$\hat{Y} \approx 0{,}629X - 42{,}0$.



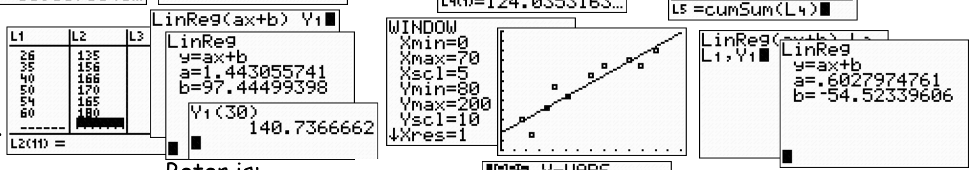
D1b $X = 160 \Rightarrow \hat{Y} = \hat{Y}(160) \approx 58{,}6$ (kg).

D1c $\hat{Y} > 75$ (intersect en een plot of) $\Rightarrow 0{,}629X - 42{,}0 > 75 \Rightarrow 0{,}629X > 117 \Rightarrow X > 186$.
Dus vanaf 186 cm.

D1d Voer de residuen in in L3. (ga op L3 staan en dan [2nd][STAT][7][ENTER])
Zet de kwadraten van L3 (de kwadraten van de residuen) in L4. (ga op de L4 staan en toets L3^2 in met [2nd][3][x^2][ENTER])
Tel de kwadraten uit L4 cumulatief op in L5. (ga op L5 staan en toets cumSum(L4) in met [2nd][STAT] [▸] [6][2nd][4][)][ENTER])
De som van de kwadraten van de residuen is (ongeveer) 366,7.

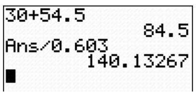


D2a Zie de plot hiernaast.
$\hat{Y} \approx 1{,}44X + 97{,}4$.

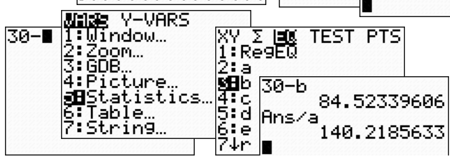


D2b $X = 30 \Rightarrow \hat{Y} = \hat{Y}(30) \approx 141$.

D2c $\hat{X} \approx 0{,}603Y - 54{,}5$.
$\hat{X} > 30$
$0{,}603Y - 54{,}5 > 30$
$0{,}603Y > 84{,}5$
$Y > 140{,}1$.

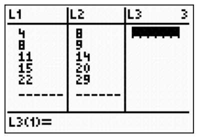


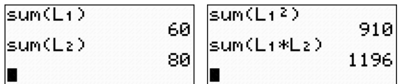
Beter is:
$\hat{X} \approx 0{,}602...Y - 54{,}5...$
$\hat{X} > 30 \Rightarrow Y > 140{,}2$.



D3

X	Y	X^2	XY
4	8	4^2	$4 \cdot 8$
8	9	8^2	$8 \cdot 9$
11	14	11^2	$11 \cdot 14$
15	20	15^2	$15 \cdot 20$
22 +	29 +	22^2 +	$22 \cdot 29$ +
$\sum X = 60$	$\sum Y = 80$	$\sum X^2 = 910$	$\sum XY = 1196$



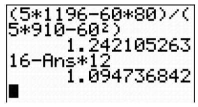


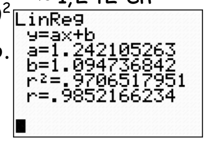
$n = 5$, $\overline{X} = \frac{\sum X}{n} = \frac{60}{5} = 12$ en $\overline{Y} = \frac{\sum Y}{n} = \frac{80}{5} = 16$;

$a_Y = \frac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5 \cdot 1196 - 60 \cdot 80}{5 \cdot 910 - 60^2} \approx 1{,}242$ en

$b = \overline{Y} - a\overline{X} \approx 16 - 1{,}242 \cdot 12 \approx 1{,}096$.

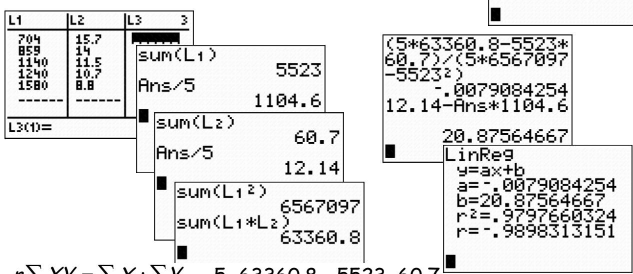
Dus $\hat{Y} \approx -1{,}24X + 1{,}10$.





D4a

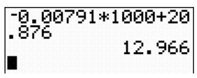
X	Y	X^2	XY
704	15,7	704^2	$704 \cdot 15{,}7$
859	14,0	859^2	$859 \cdot 14{,}0$
1104	11,5	1104^2	$1104 \cdot 11{,}5$
1240	10,7	1240^2	$1204 \cdot 10{,}7$
1580 +	8,8 +	1580^2 +	$1580 \cdot 8{,}8$ +
$\sum X = 5\,523$	$\sum Y = 60{,}7$	$\sum X^2 = 6\,567\,097$	$\sum XY = 63\,360{,}8$



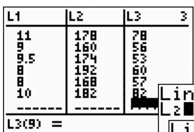
$n = 5$, $\overline{X} = \frac{\sum X}{n} = \frac{5523}{5} = 1104{,}6$ en $\overline{Y} = \frac{\sum Y}{n} = \frac{60{,}5}{5} = 12{,}14$; $a_Y = \frac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5 \cdot 63360{,}8 - 5523 \cdot 60{,}7}{5 \cdot 6567097 - 5523^2} \approx -0{,}00791$ en

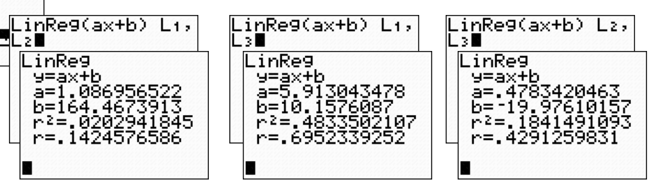
$b = \overline{Y} - a\overline{X} \approx 12{,}14 + 0{,}00791 \cdot 1104{,}6 \approx 20{,}876$. Dus $\hat{Y} \approx -0{,}00791X + 20{,}876$.

D4b $X = 1000$ geeft $\hat{Y} \approx -0{,}00791 \cdot 1000 + 20{,}876 \approx 12{,}97$. Je kunt bijna 13 km rijden.



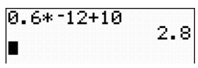
D5 Voer L1 = $\{7, 9.5, 11, 9, 9.5, 8, 8, 10\}$,
L2 = $\{169, 171, 178, 160, 174, 192, 168, 182\}$
en L3 = $\{56, 65, 78, 56, 53, 60, 57, 82\}$ in.
LinReg(ax+b) L1, L2 geeft bij X en Y $r \approx 0{,}142$.
LinReg(ax+b) L1, L3 geeft bij X en Z $r \approx 0{,}695$.
LinReg(ax+b) L2, L3 geeft bij Y en Z $r \approx 0{,}429$.
Dus X en Z vertonen de sterkste (lineaire) correlatie.

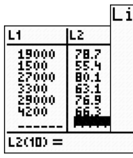


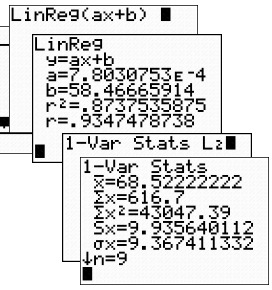


D6a $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}8 \cdot \frac{1}{2} = 0{,}4 \Rightarrow \hat{Y} = 0{,}4(X - 10) + 12$. Dus $X = 13$ geeft $\hat{Y} = 0{,}4(13 - 10) + 12 = 13{,}2$.

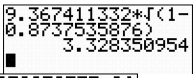
D6b $\hat{X} = 0{,}6Y + 2{,}8$ ofwel $\hat{X} = 0{,}6(Y - 12) + 10 \Rightarrow a_X = 0{,}6 = r \cdot \frac{\sigma_X}{\sigma_Y} = r \cdot \frac{2}{1} \Rightarrow 2r = 0{,}6 \Rightarrow r = 0{,}3$.



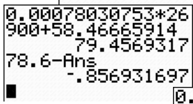




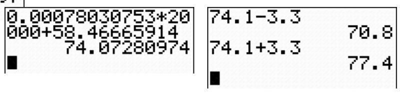
D7a Voer L1 $=\{4\,700, 25\,300, 1\,980, 19\,000, 1\,500, 27\,000, 3\,300, 29\,000, 4\,200\}$
en L2 $=\{61.1, 78, 57.1, 78.7, 55.4, 80.1, 63.1, 76.9, 66.3\}$ in.
LinReg(ax+b) L1, L2 geeft $\widehat{Y} \approx 0{,}000780X + 58{,}5$.



D7b $r^2 \approx 0{,}8737535876$ en 1-Var Stats L2 geeft $\sigma_Y \approx 9{,}367411332$.
$\sigma_d = \sigma_Y \cdot \sqrt{1-r^2} \approx 9{,}367411332 \cdot \sqrt{1-0{,}8737535876} \approx 3{,}33$.

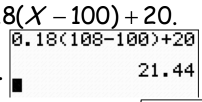


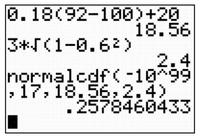
D7c $X = 26\,900$ geeft $\widehat{Y} \approx 0{,}00078... \cdot 26\,900 + 58{,}466... \approx 79{,}46$.
Het residu is $Y - \widehat{Y} \approx 78{,}6 - 79{,}5 = -0{,}9$.



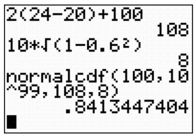
D7d $X = 20\,000$ geeft $\widehat{Y} \approx 0{,}00078... \cdot 20\,000 + 58{,}466... \approx 74{,}1$.
Het 68%-betrouwbaarheidsinterval is $\langle 74{,}1 - 3{,}3; 74{,}1 + 3{,}3 \rangle = \langle 70{,}8; 77{,}4 \rangle$.

D8a $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = 0{,}6 \cdot \frac{3}{10} = 0{,}6 \cdot 0{,}3 = 0{,}18 \Rightarrow \widehat{Y} = 0{,}18(X - 100) + 20$.
$X = 108$ geeft $\widehat{Y} = 0{,}18(108 - 100) + 20 = 21{,}44$.





D8b $X = 92$ geeft $\widehat{Y} = 0{,}18(92 - 100) + 20 = 18{,}56$ en $\sigma_d = \sigma_Y \cdot \sqrt{1-r^2} = 3 \cdot \sqrt{1-0{,}6^2} = 2{,}4$.
$P(Y < 17) = \text{normalcdf}(-10^{99}, 17, 18.56, 2.4) \approx 0{,}258$.

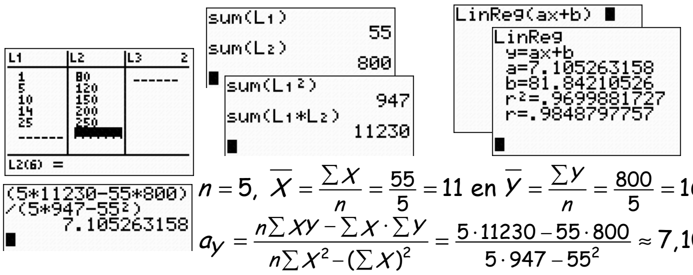


D8c $a_X = r \cdot \frac{\sigma_X}{\sigma_Y} = 0{,}6 \cdot \frac{10}{3} = \frac{6}{3} = 2 \Rightarrow \widehat{X} = 2(Y - 20) + 100$.
$Y = 24$ geeft $\widehat{X} = 2(24 - 20) + 100 = 108$ en $\sigma_d = \sigma_X \cdot \sqrt{1-r^2} = 10 \cdot \sqrt{1-0{,}6^2} = 8$.
$P(X > 100) = \text{normalcdf}(100, 10^{99}, 108, 8) \approx 0{,}841$.

Gemengde opgaven 10. Correlatie en regressie

G12

X	Y	X^2	XY
1	80	1	80
5	120	25	600
10	150	100	1500
14	200	196	2800
25 +	250 +	625 +	6250 +
$\sum X = 55$	$\sum Y = 800$	$\sum X^2 = 947$	$\sum XY = 11230$

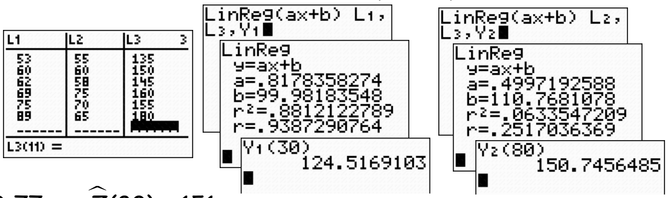


$n = 5$, $\overline{X} = \frac{\sum X}{n} = \frac{55}{5} = 11$ en $\overline{Y} = \frac{\sum Y}{n} = \frac{800}{5} = 160$;

$a_Y = \frac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5 \cdot 11230 - 55 \cdot 800}{5 \cdot 947 - 55^2} \approx 7{,}105.$

Dus $\widehat{Y} \approx 7{,}11(X - 11) + 160$.

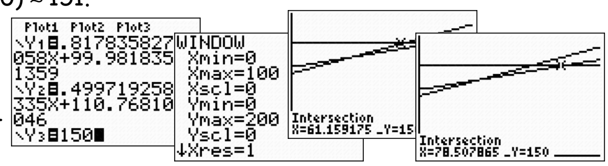
G13a Zet de waarden van de leeftijd X in lijst L1.
Zet de waarden van het gewicht Y in lijst L2.
Zet de waarden van de bloeddruk Z in lijst L3.
De optie LinReg(ax+b) L1,L3,Y1 geeft
$\widehat{Z} \approx 0{,}818X + 99{,}98$ en $\widehat{Z}(30) \approx 125$.



G13b LinReg(ax+b) L2,L3,Y1 geeft $\widehat{Z} \approx 0{,}4997Y + 110{,}77$ en $\widehat{Z}(80) \approx 151$.

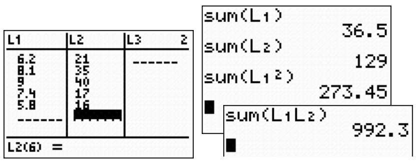
G13c $\widehat{Z} \approx 0{,}818X + 99{,}98 > 150$ algebraïsch of intersect geeft
$X > 61{,}16$. Dus vanaf 62 jaar.

G13d $\widehat{Z} \approx 0{,}4997Y + 110{,}77 > 150$ algebraïsch of intersect geeft
$Y > 78{,}5$. Dus vanaf 78,5 kg.



G14a

X	Y
6,2	21
8,1	35
9,0	40
7,4	17
5,8 +	16 +
$\sum X = 36{,}5$	$\sum Y = 129$

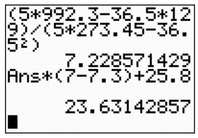


$n = 5$, $\overline{X} = \frac{\sum X}{n} = \frac{36{,}5}{5} = 7{,}3$ en $\overline{Y} = \frac{\sum Y}{n} = \frac{129}{5} = 25{,}8$;

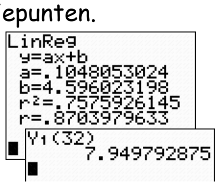
$a_Y = \frac{n\sum XY - \sum X \cdot \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5 \cdot 992{,}3 - 36{,}5 \cdot 129}{5 \cdot 273{,}45 - 36{,}5^2} \approx 7{,}23.$

Dus $\widehat{Y} \approx 7{,}23(X - 7{,}3) + 25{,}8$.

G14b $X = 7{,}0$ geeft $\widehat{Y} \approx 7{,}23(7{,}0 - 7{,}3) + 25{,}8 \approx 23{,}6$.
Dus de voorspelling is ongeveer 24 studiepunten.



G14c Zet de waarden van het aantal studiepunten Y in lijst L1.
Zet de waarden van het gemiddelde eindcijfer X in lijst L2.
De optie LinReg(ax+b) L1,L2,Y1 geeft $\widehat{X} \approx 0{,}105Y + 4{,}60$ en $\widehat{X}(32) \approx 7{,}9$.
Dus het gemiddelde van Marc is naar verwachting 7,9.



G15a $r^2 = a_X \cdot a_Y = 0{,}675 \cdot 1{,}2 = 0{,}81 \Rightarrow r = 0{,}9$.

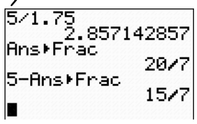
$a_Y = r \cdot \frac{\sigma_Y}{\sigma_X}$ geeft $\sigma_X = r \cdot \frac{\sigma_Y}{a_Y} = 0{,}9 \cdot \frac{\sigma_Y}{1{,}2} = 0{,}75\sigma_Y$.

σ_X en σ_Y zijn nog niet te berekenen.
Er is nog een gegeven nodig.

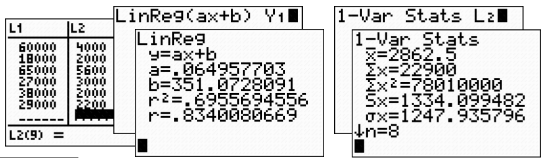
G15b $\left.\begin{array}{l}\sigma_X = 0{,}75\sigma_Y \text{ én} \\ \sigma_X + \sigma_Y = 5\end{array}\right\} \Rightarrow 0{,}75\sigma_Y + \sigma_Y = 5 \Rightarrow$

$1{,}75\sigma_Y = 5 \Rightarrow \sigma_Y = \frac{5}{1{,}75} = \frac{20}{7}$ en

$\sigma_X = 5 - \sigma_Y = 5 - \frac{20}{7} = \frac{15}{7}$.

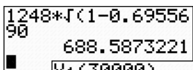


G16a Zet de waarden van het jaarinkomen X in lijst L1.
Zet de waarden van het bedrag per jaar aan kleding Y in lijst L2.
LinReg(ax+b) L1,L2,Y1 geeft $\widehat{Y} \approx 0{,}0650X + 351$.

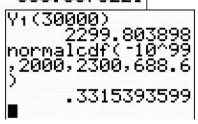


G16b $r^2 \approx 0{,}695569$ en $\sigma_Y \approx 1248$ (zie de schermen naast G16a).

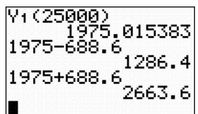
$\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} \approx 1248 \cdot \sqrt{1 - 0{,}695569} \approx 688{,}6$.



G16c $X = 30\,000$ geeft $\widehat{Y}(30\,000) \approx 2\,300$.
$P(Y < 2000) = \text{normalcdf}(-10^{99}, 2\,000, 2\,300, 688.6) \approx 0{,}332$.

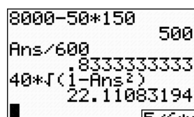


G16d $X = 25\,000$ geeft $\widehat{Y}(25\,000) \approx 1975$.
Het 68%-betrouwbaarheidsinterval van Y is $\langle \mu - \sigma, \mu + \sigma \rangle \approx \langle 1975 - 688{,}6; 1975 + 688{,}6 \rangle \approx \langle 1286, 2\,663 \rangle$.

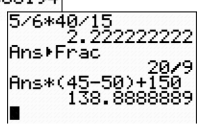


G17a pmcc $r = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sigma_X \cdot \sigma_Y} = \frac{\frac{\sum XY}{n} - \frac{\sum X}{n} \cdot \frac{\sum Y}{n}}{\sigma_X \cdot \sigma_Y} = \frac{\frac{\sum XY}{10} - \frac{\sum X}{10} \cdot \frac{\sum Y}{10}}{15 \cdot 40} = \frac{\frac{80\,000}{10} - \frac{500}{10} \cdot \frac{1\,500}{10}}{600} = \frac{500}{600} = \frac{5}{6} \approx 0{,}833$.

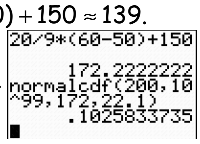
$\sigma_d = \sigma_Y \cdot \sqrt{1 - r^2} \approx 40 \cdot \sqrt{1 - 0{,}833^2} \approx 22{,}1$.



G17b $a_Y = r \cdot \frac{\sigma_Y}{\sigma_X} = \frac{5}{6} \cdot \frac{40}{15} = \frac{20}{9} \approx 2{,}22$. Dus $\widehat{Y} = \frac{20}{9}(X - 50) + 150$ en $\widehat{Y}(45) = \frac{20}{9}(45 - 50) + 150 \approx 139$.



G17c $X = 60$ geeft $\widehat{Y}(60) \approx 172 \Rightarrow P(Y > 200) = \text{normalcdf}(200, 10^{99}, 172, 22.1) \approx 0{,}103$.



G18a Zet de waarden van de bloeddruk X in lijst L1.
Zet de waarden van de bloeddruk Y in lijst L2.
LinReg(ax+b) L1,L2,Y1 geeft $\widehat{Y} \approx 1{,}09X + 2{,}51$ en pmcc $r \approx 0{,}87$.

```
L1   L2
136  140
180  213
133  160
150  167
141  158
166  166
131  149
```

```
LinReg(ax+b) L1,L2,Y1
LinReg
 y=ax+b
 a=1.094920899
 b=2.509575354
 r²=.7563409118
 r=.8696786256
Y1(150)
        166.7477102
```

```
LinReg(ax+b) L2,L1,Y2
LinReg
 y=ax+b
 a=.690772194
 b=34.36280861
 r²=.7563409118
 r=.8696786256
Y2(180)
        158.7018035
```

G18b $X = 150$ geeft $\widehat{Y}(150) \approx 166{,}75 \approx 167$.

G18c LinReg(ax+b) L2,L1,Y2 geeft $\widehat{X} \approx 0{,}69Y + 34{,}4$ en $\widehat{X}(180) \approx 158{,}70 \approx 159$.

G18d $r^2 \approx 0{,}756$ en $\sigma_Y \approx 21{,}59 \Rightarrow \sigma_d = \sigma_Y \cdot \sqrt{1-r^2} \approx 21{,}59 \cdot \sqrt{1-0{,}756} \approx 10{,}66$.

```
1-Var Stats L2
1-Var Stats
 x̄=164.7142857
 Σx=1153
 Σx²=193179
 Sx=23.32176584
 σx=21.59175958
 ↓n=7
```

```
21.59*√(1-0.756)
        10.66467329
```

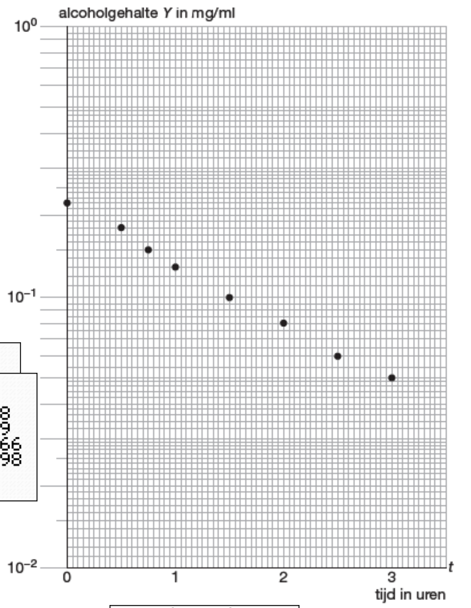
G18e 68%-betrouwbaarheidsinterval is $\langle 166{,}75 - 10{,}66;\ 166{,}75 + 10{,}66 \rangle \approx \langle 156, 177 \rangle$

G18f $X = 140$ geeft $\widehat{Y}(140) \approx 155{,}8$.
$P(Y > 180) = \text{normalcdf}(180, 10^{99}, 155.8, 10.66) \approx 0{,}012$.

```
Y1(140)
        155.7985012
normalcdf(180,10
^99,155.8,10.66)
        .0115986347
```

G19a op $[0; 0{,}5]$ is $\Delta y = 0{,}18 - 0{,}22 = -0{,}04$
op $[0{,}5; 1]$ is $\Delta y = 0{,}13 - 0{,}18 = -0{,}05$
$\}$ niet gelijk $\Rightarrow$ geen lineair verband.

G19b De punten liggen bij benadering op een rechte lijn. (zie hiernaast)
Dus er is een exponentieel verband.



G19c Zet de waarden van t in lijst L1.
Zet de waarden van Y in lijst L2.
Maak lijst L3 = log(L2).
LinReg(ax+b) L1,L3 geeft $\widehat{Y^*} \approx -0{,}221X - 0{,}656$.
ExpReg L1,L2 geeft $Y \approx 0{,}221 \cdot 0{,}601^X$.

```
L1    L2    L3
0     .22   -.6576
.5    .18   -.7447
.75   .15   -.8239
1     .13   -.8861
1.5   .1    -1
2     .08   -1.097
2.5   .06   -1.222
L3(1)=-.657577319...
```

```
LinReg(ax+b) L1,L3
LinReg
 y=ax+b
 a=-.2208535123
 b=-.6559321185
 r²=.9966699566
 r=-.9983335898
```

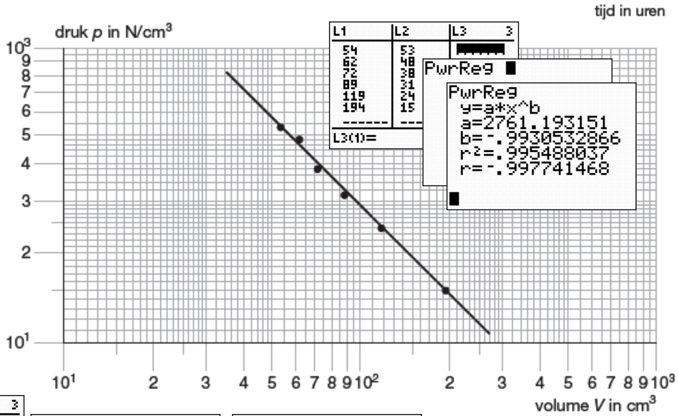
```
ExpReg L1,L2
ExpReg
 y=a*b^x
 a=.2208349878
 b=.6013765479
 r²=.9966699566
 r=-.9983335898
```

G19d $Y^* = \log(Y) = -0{,}221X - 0{,}656$.
$Y = 10^{-0{,}221X - 0{,}656}$

```
10^-0.221
        .6011737375
10^-0.656
        .2208004733
```

$= \left(10^{-0{,}221}\right)^X \cdot 10^{-0{,}656}$
$\approx 0{,}601^X \cdot 0{,}221 = 0{,}221 \cdot 0{,}601^X$. (of, zie ook in G19c, met ExpReg)

G20a Punten liggen bij benadering op een rechte lijn
(zie hiernaast) $\Rightarrow$ er is een machtsverband.



G20b Zet de waarden van V in lijst L1.
Zet de waarden van p in lijst L2.
Maak L3 = log(L1) en L4 = log(L2).
LinReg(ax+b) L3,L4 $\Rightarrow \widehat{p^*} \approx -0{,}993V^* + 3{,}441$. (Ga na!!!)
PwrReg L1,L2 $\Rightarrow Y \approx 0{,}221 \cdot 0{,}601^X$.

```
L1    L2
54    53
62    48
72    38
89    31
119   24
194   15
L3(1)=
```

```
PwrReg
PwrReg
 y=a*x^b
 a=2761.193151
 b=-.9930532866
 r²=.995488037
 r=-.997741468
```

G20c $\log(p) \approx -0{,}993\log(V) + 3{,}441$
$p \approx 10^{-0{,}993\log(V) + 3{,}441}$
$p \approx 10^{\log(V^{-0{,}993})} \cdot 10^{3{,}441}$
$p \approx 2761 \cdot V^{-0{,}993}$.

```
10^3.441
        2760.577856
```

G21a Zet de waarden van X in lijst L1,
die van Y in lijst L2 en die van Z in L3.
$\sum X = 526;\ \sum Y = 645;\ \sum Z = 1\,430;$

```
L1    L2    L3
53    55    135
60    60    150
62    58    145
69    75    160
75    70    155
89    65    180
L3(11) =
```

```
sum(L1)        526
sum(L2)        645
sum(L3)       1430
sum(L1L2)    34011
sum(L1L3)    79000
sum(L2L3)    92680
sum(L1²)     32292
sum(L2²)     42493
```

$\sum(XY) = 34\,011;\ \sum(XZ) = 79\,000;\ \sum(YZ) = 92\,680;\ \sum\left(X^2\right) = 32\,292$ en $\sum\left(Y^2\right) = 42\,493$.

Dit geeft $\begin{cases} 1\,430 = 526 \cdot a + 645 \cdot b + 10 \cdot c \\ 79\,000 = 32\,292 \cdot a + 34\,011 \cdot b + 526 \cdot c \\ 92\,680 = 34\,011 \cdot a + 42\,493 \cdot b + 645 \cdot c \end{cases}$

```
MATRIX[A] 3 ×4
[ 526    645    10    1430  ]
[ 32292  34011  526   79000 ]
[ 34011  42493  645   92680 ]
```

```
rref([A])
[[1 0 0 .81014680661]
 [0 1 0 .4232988975 ]
 [0 0 1 73.08349908]]
```

G21b $526 \cdot 0{,}81015 + 645 \cdot 0{,}423989 + 10 \cdot 73{,}0835 = 1\,430{,}446805 \approx 1\,430$.
$32\,292 \cdot 0{,}81015 + 34\,011 \cdot 0{,}423989 + 526 \cdot 73{,}0835 \approx 79\,023{,}... \approx 79\,000$.
$34\,011 \cdot 0{,}81015 + 42\,493 \cdot 0{,}423989 + 645 \cdot 73{,}0835 = 92\,709{,}... \approx 92\,680$.

```
526*0.81015+645*
0.423989+10*73.0
835
        1430.446805
32292*0.81015+34
011*0.423989+526
*73.0835
        79023.57468
34011*0.81015+42
493*0.423989+645
*73.0835
        92709.43373
```

G21c $X = 70$ en $Y = 80 \Rightarrow Z \approx 70 \cdot 0{,}81015 + 80 \cdot 0{,}423989 + 1 \cdot 73{,}0835 \approx 164$.

```
70*0.81015+80*0.
423989+1*73.0835
        163.71312
```

G21d Van X en Z is $r \approx 0{,}94$ en van Y en Z is $r \approx 0{,}25$ (zie G13ab).
Er lijkt dus een sterkere (lineaire) samenhang te bestaan tussen X en Z dan tussen Y en Z.
Dit komt overeen met de formule van $\widehat{Z}$, waarvoor geldt dat de invloed van X groter is dan die van Y.
Immers $a \approx 0{,}81015$ en $b \approx 0{,}423989$.

TI-84 9. Regressiemodellen

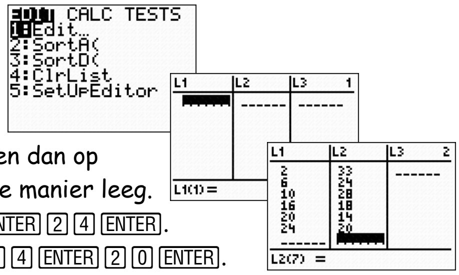
In dit practicum werk je met de tabel hiernaast.

x	2	6	10	16	20	24
y	33	24	28	18	14	20

A Het invoeren van een tabel

De tabel voer je als volgt in op de GR.

- Kies [STAT]. Je komt in het STAT-EDIT-menu.
- Kies de optie **1: Edit**… met [1] of [ENTER].
- Je komt in het **lijsten-invoerscherm**.
- Je ziet de lijsten L1, L2 en L3. Zie de schermen hiernaast.
- Maak lijst L1 zo nodig leeg door de cursor op L1 te zetten en dan op [CLEAR] [ENTER] te drukken. Andere lijsten maak je op dezelfde manier leeg.
- Voer bij L1 in [2] [ENTER] [6] [ENTER] [1] [0] [ENTER] [1] [6] [ENTER] [2] [0] [ENTER] [2] [4] [ENTER].
  Voer bij L2 in [3] [3] [ENTER] [2] [4] [ENTER] [2] [8] [ENTER] [1] [8] [ENTER] [1] [4] [ENTER] [2] [0] [ENTER].



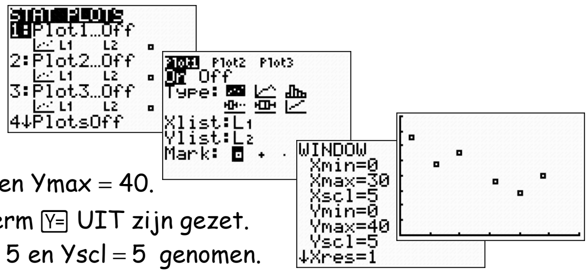
Je hebt de tabel ingevoerd op de GR. We zeggen dat de lijsten $L1 = \{2, 6, 10, 16, 20, 24\}$ en $L2 = \{33, 24, 28, 18, 14, 20\}$ zijn ingevoerd.

Als de lijsten L1 en L2 niet op het scherm staan, ga dan in het STAT-EDIT-menu naar 5: SetUpEditor.

Het plotten van de punten van een tabel

Met $L1 = \{2, 6, 10, 16, 20, 24\}$ en $L2 = \{33, 24, 28, 18, 14, 20\}$ krijg je als volgt de **puntengrafiek** van de tabel.

- Kies [STAT PLOT] ([2nd] [Y=]).
- Kies de optie **Plot1** met [1] of [ENTER].
- Zorg voor het scherm hiernaast.
- Ga naar [WINDOW] en kies $Xmin = 0$, $Xmax = 30$, $Ymin = 0$ en $Ymax = 40$.
- Zorg ervoor dat de formules op het formule-invoerscherm [Y=] UIT zijn gezet.
- Na [GRAPH] krijg je de figuur hiernaast. Hierbij is $Xscl = 5$ en $Yscl = 5$ genomen.

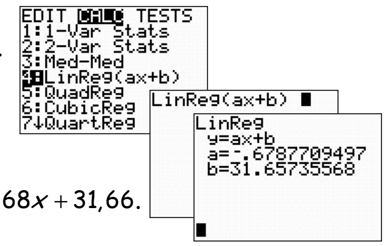


Het opstellen van de formule van een regressielijn

De punten van de tabel liggen niet op een rechte lijn.
De lijn die zo goed mogelijk bij de punten past heet de **regressielijn**.
De GR bezit een optie om de formule van de regressielijn op te stellen.
Uitgaande van de lijsten L1 en L2 gaat dat als volgt.

- Kies [STAT] en ga met [▸] naar het STAT-CALC-menu.
- Kies de optie **4: LinReg(ax+b)** met [4] of [▾][▾][▾][ENTER] en [ENTER].

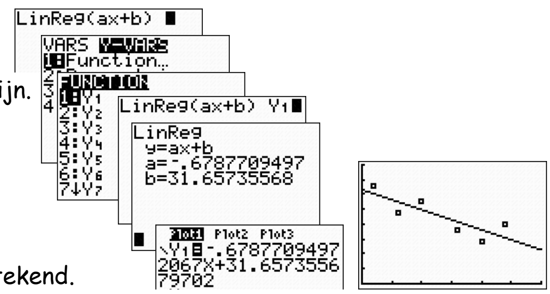
Afgerond op twee decimalen is de formule van de regressielijn $y = -0{,}68x + 31{,}66$.



Het plotten van een regressielijn

Uitgaande van de lijsten L1 en L2 plot je als volgt de regressielijn.

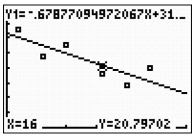
- Kies de optie **4: LinReg(ax+b)** uit het STAT-CALC-menu.
- Zorg voor LinReg(ax+b) Y1 op het basisscherm.
  Je krijgt Y1 met [VARS][▸][ENTER][ENTER].
- Druk op [ENTER] en kies [GRAPH].



Je ziet dat de best passende lijn door de gegeven punten is getekend.
Ga na dat de formule van de regressielijn is ingevoerd bij Y1 op het formule-invoerscherm.

Een lijst met residuen

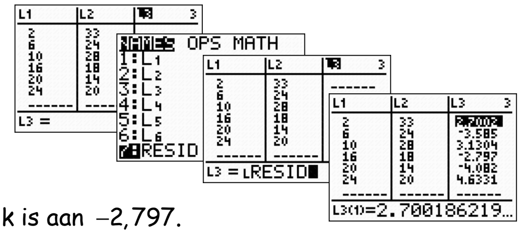
Uit de tabel volgt: bij $x = 16$ hoort $y = 18$.
Uit de formule volgt: bij $x = 16$ hoort $y = 20,797$.
Het verschil $18 - 20,797$ heet het **residu** bij $x = 16$.



Je kunt $y = 20,797$ controleren met [TRACE].
Zet de trace-cursor op de lijn met [▼] en tik in [1] [6] [ENTER].

Je krijgt als volgt een lijst met residuen.

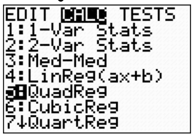
- Zorg voor LinReg(ax+b) Y1 op het basisscherm en [ENTER].
- Ga naar het lijsten-invoerscherm en zet de cursor op L3.
- Kies de optie **7: Resid** uit het LIST-NAMES-menu.
- Na [ENTER] krijg je het onderste scherm hiernaast.

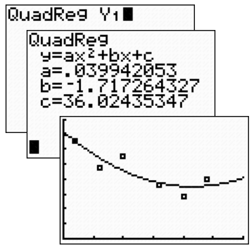


In lijst L3 staan de residuen. Je ziet dat het residu bij $x = 16$ gelijk is aan $-2,797$.

Andere regressiemodellen

Hiervoor heb je de best-passende lijn gevonden bij de gegeven tabel.
Je werkt dan met een **lineair model**.
Het is ook mogelijk op zoek te gaan naar de best-passende parabool die bij de tabel hoort.
Je gebruikt dan de optie **5: QuadReg** uit het STAT-CALC-menu.
Met **6: CubicReg** krijg je de best-passende derdemachtsformule en met **7: QuartReg** krijg je de best-passende vierdemachtsformule.
In het STAT-CALC-menu staan nog meer regressiemodellen.

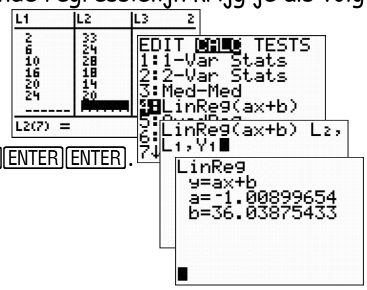




B Regressie van X op Y

De formule van de regressielijn van X op Y en de grafiek van de bijbehorende regressielijn krijg je als volgt.

- Zorg ervoor dat de lijsten $L1 = \{2, 6, 10, 16, 20, 24\}$ en $L2 = \{33, 24, 28, 18, 14, 20\}$ zijn ingevoerd.
- Kies de optie LinReg(ax+b) uit het STAT-CALC-menu met [4] en zorg op het basisscherm voor het scherm hiernaast met [2nd][1][,][2nd][2][,][VARS][▶][ENTER][ENTER].
- Na [ENTER] krijg je het scherm hiernaast.
  Afgerond op twee decimalen nauwkeurig is de formule van de regressielijn van X op Y gelijk aan $\widehat{X} = -1,01Y + 36,04$.



Merk op dat met de optie LinReg(ax+b) uit het STAT-CALC-menu standaard de regressie LinReg(ax+b) L1, L2, Y1 dus de regressielijn $\widehat{Y} = -0,68X + 31,66$ (regressie van L2 op L1) wordt berekend. Wil je hiervan afwijken, dan moet je na de optie LinReg(ax+b) de door jou gewenste lijsten op het basisscherm intikken en dan pas op [ENTER] drukken.

C Diagnostic on/off

Je kunt de GR zo instellen dat er bij elke regressieberekening de waarde van de pmcc r en de waarde van r^2 vermeld worden. Je kiest dan DiagnosticOn.
Je vindt deze opdracht met [CATALOG] ([2nd] [0]) in het CATALOG -menu.
Wil je weer regressieberekeningen zonder vermelding van r en r^2 dan moet je in CATALOG -menu kiezen voor DiagnosticOff.