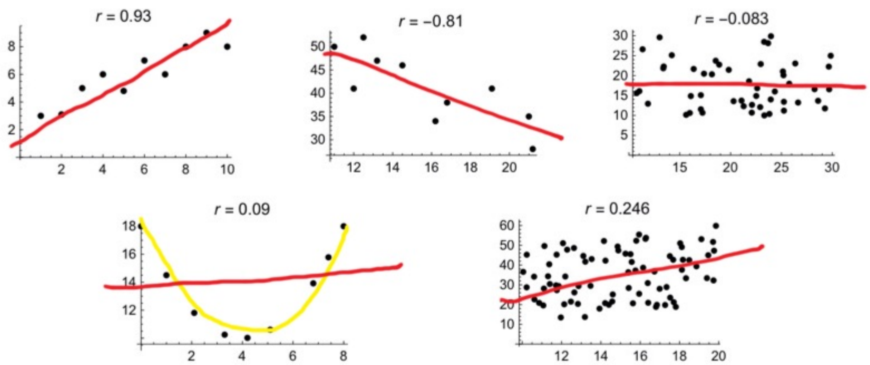
10.1 - Correlation

linear correlation

Correlation is measured using the **coefficient of correlation**, *r*, and can have values between −1 and 1 inclusive.

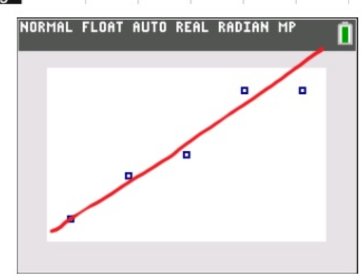


$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \cdot \sqrt{n(\sum y^2) - (\sum y)^2}}$$ (Our calculator will do this for us.)

EXAMPLE 1 Enter the data below into L_1 and L_2 and calculate **2-Var Stats** on the TI84. Write down: n, $\sum x$, $\sum x^2$, $\sum y$, $\sum y^2$, $\sum xy$. Use the above formula to calculate the correlation coefficient. Also, verify that the data show a linear relation. Is there strong correlation between x and y?

x	1	2	3	4	5
y	2	4	5	8	8

	A	B	C	D	E	F
1						
2		x	y	x^2	y^2	x y
3		1	2	1	4	2
4		2	4	4	16	8
5		3	5	9	25	15
6		4	8	16	64	32
7		5	8	25	64	40
8	sums=	15	27	55	173	97
9						



$$r = \frac{5(97) - (15)(27)}{\sqrt{5(55) - 15^2} \cdot \sqrt{5(173) - 27^2}}$$

$$= \frac{80}{\sqrt{50} \cdot \sqrt{136}}$$

$$= \frac{80}{\sqrt{6800}}$$

$$\boxed{r \approx 0.97}$$

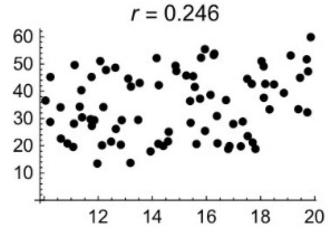
TABLE A-6 Critical Values of the Pearson Correlation Coefficient *r*

n	α = .05	α = .01
4	.950	.990
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834

The critical r value is ±0.878. Since 0.97 > 0.878 we do have signifant linear correlation at the α=0.05 level.

EXAMPLE 2 Determine if there is significant correlation for *Example 1*, and for the fifth scatterplot above where $n = 80$, using Table A-6.

$r = 0.246$



$n = 80$
$r = 0.246$

$r_c = 0.220$

since $r > r_c$ we do have linear correlation.

TABLE A-6 Critical Values of the Pearson Correlation Coefficient r

n	$\alpha = .05$	$\alpha = .01$
4	.950	.990
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

NOTE: To test $H_0: \rho = 0$ (no correlation) against $H_1: \rho \neq 0$ (correlation), reject H_0 if the absolute value of r is greater than or equal to the critical value in the table.

LinRegTTest

$H_0: \rho = 0$ (no correlation)

$H_1: \rho \neq 0$ (there is correlation)

```
NORMAL FLOAT AUTO REAL RADIAN MP
LinRegTTest
Xlist:L1
Ylist:L2
Freq:1
β & ρ:≠0 <0 >0
RegEQ:
Calculate
```

```
NORMAL FLOAT AUTO REAL RADIAN MP
LinRegTTest
y=a+bx
β≠0 and ρ≠0
t=6.92820323
p=0.0061653731
df=3
a=0.6
b=1.6
↓s=0.7302967433
r²=0.9411764706
r=0.9701425001
```

Correlation and Regression

	Var1	Var2
1	1	2
2	2	4
3	3	5
4	4	8
5	5	8
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

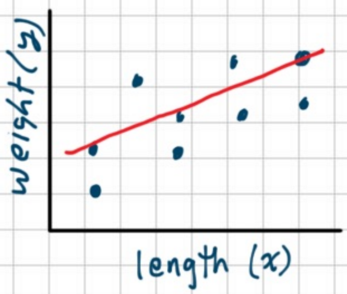
Correlation r =	0.9701
r^2 =	0.9412
Regression: $y = a + b x$	
a =	0.6000
b =	1.6000
degrees of freedom =	3
α =	0.1
critical r_{cr} =	± 0.8054
t =	6.9282
P-value =	0.0062

Since $p < 0.05$ we reject H_0 and conclude $\rho \neq 0$ meaning there is linear correlation.

Interpretation of r^2.

r^2 is the proportion of variation in y that can be attributed to the variation in x.

Example: vehicle length (x) -vs- weight (y)



$r = 0.241$

$r^2 = 0.058$

About 6% of a vehicle's weight can be explained by the length of the vehicle.

- suv or sedan
- type (truck, van, etc.)
- Quality
- height and width
- metal thickness