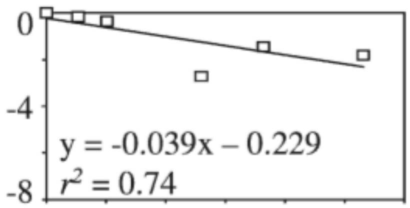
**Exercise 7** – Linear fit. Correlation coefficient. Closure effects.

1. Using t-test, find the probability that the points in the graph below are not correlated at all.

From our last lecture, the value of Student's t can be calculated as: $t = r\sqrt{\dfrac{N-2}{1-r^2}}$




2. *Closure* occurs when a set of data is normalized or forced to a common sum. By far the most common occurrence of closure is when the data is expressed in %. For example, compositions of metal alloys, as well as rocks and minerals, are commonly reported as weight (or molar) percentages, with totals summing to 100%. To determine the severity of the impact on correlation coefficients we will perform a simple simulation.

1. Generate 10 random sample compositions for four major components (let's call them A,B,C,D) with relative concentration ratios 10, 5, 2, 1. Then normalize these compositions to 100%.

```
A=10*normrnd(10,2,10,1)
B=5*normrnd(10,2,10,1)
C=2*normrnd(10,2,10,1)
D=1*normrnd(10,2,10,1)

sum=A+B+C+D
```

```
% now express these concentrations as percentages of the total
    Ap=A./sum
    Bp=B./sum
    Cp=C./sum
    Dp=D./sum
```

Calculate cross-correlation coefficients for both representations

```
corrcoef([A B C D])
corrcoef([Ap Bp Cp Dp])
```

Make scatter plots of the data to see what these correlation coefficients correspond to. For example: plot(Ap,Bp,'o'), plot(A,B,'o').

For values of the correlation coefficient that appear large, are they significantly different from r=0? Use t-test.

Check if increasing the number of points to 100 would have any effect.