ជំពូកទី១ តារាជាប់ពាក់ព័ន្ធ

Correlation

រៀបរៀងនិងបង្រៀនដោយបណ្ឌិត ង៉ាន់ ស៊ុនដេត

សាស្ត្រាចារ្យនៃសាកលវិទ្យាល័យភូមិន្ទនីតិសាស្ត្រនិងវិទ្យាសាស្ត្រសេដ្ឋកិច្ច

សេដ្ឋកិច្ចមាត្រសាស្ត្រ

Econometrics

What is Econometrics?

- Econometrics may be defined as the quantitative analysis of actual economic phenomena (A. H. Studenmund: *Using econometrics*. 2nd. 1992.).
- Econometrics has tree major uses:
  i. the description of economics reality,
  ii. the testing of hypotheses about economics theory,
  iii. the forecasting of future economic activity.

ភាពជាប់ពាក់ព័ន្ធ

Correlation

រៀបរៀងនិងបង្រៀន

បណ្ឌិត ង៉ាន់ ស៊ុនដេត សាស្ត្រាចារ្យ

នៃសាកលវិទ្យាល័យភូមិន្ទនីតិសាស្ត្រនិងវិទ្យាសាស្ត្រសេដ្ឋកិច្ច

1. Definition

- If X and Y are random variables than their ***correlation*** is

$$R = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}}$$

- As with the covariance, the correlation R between two random variables measures the degree of ***linear*** association them.

(R. Carter Hill, William E. Griffith, George G. Judge: Undergraduate Econometrics, 2001. 2ed.)

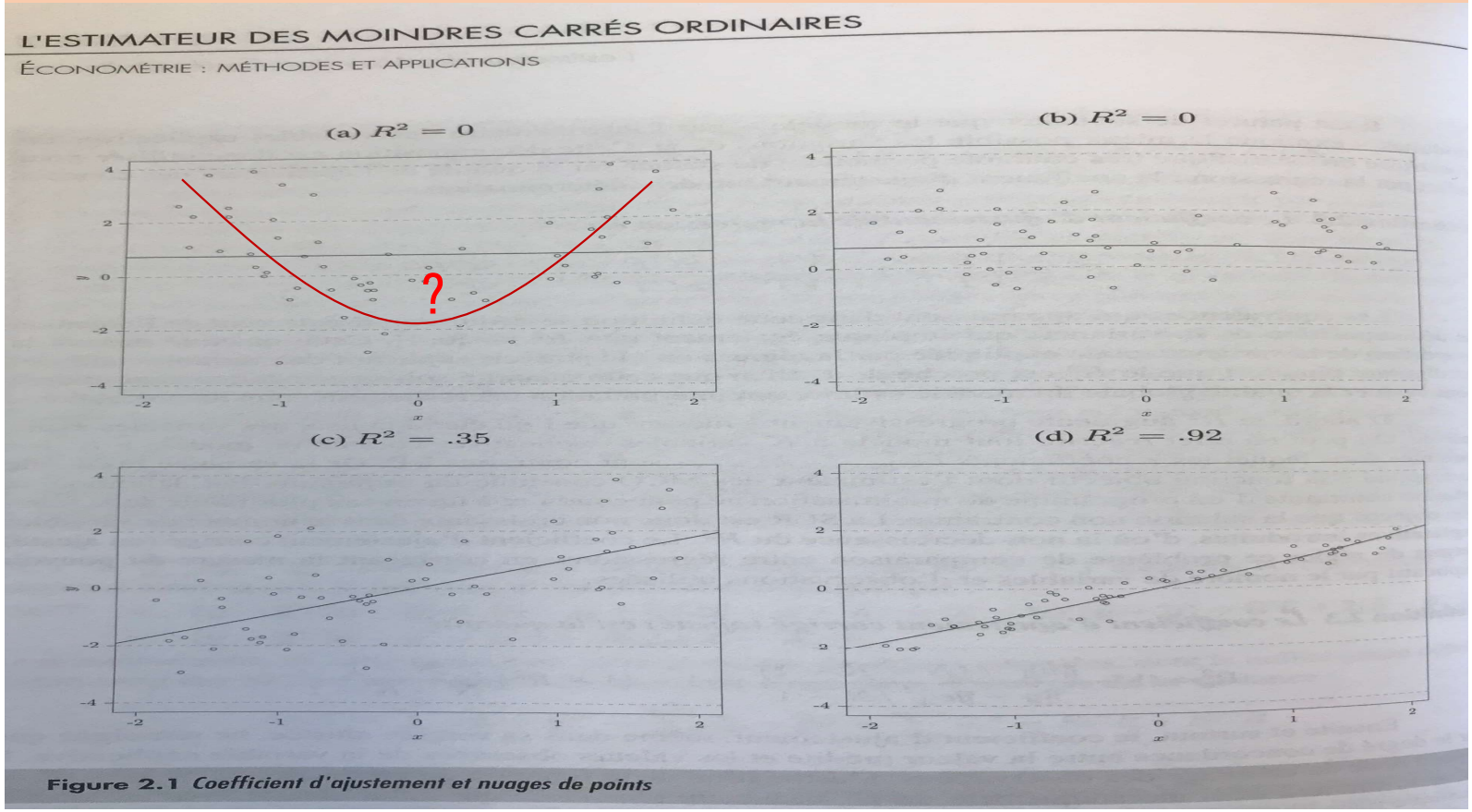
- The correlation must lie between -1 and +1.
- Thus, the correlation between X and Y is nearly +1 or -1, if X is a perfect positive or negative linear function of Y. If Cov(X,Y)=0 then R=0, X and Y can shout be not correlation linear.

2. The level of correlation

- If $R\ more\ than\ 0.6$, then X and Y is strength correlation.
- If $0.2\ \leq R\ \leq 0.6$, then X and Y is good correlation.
- If R less than 0.2, then X and Y is low correlation.

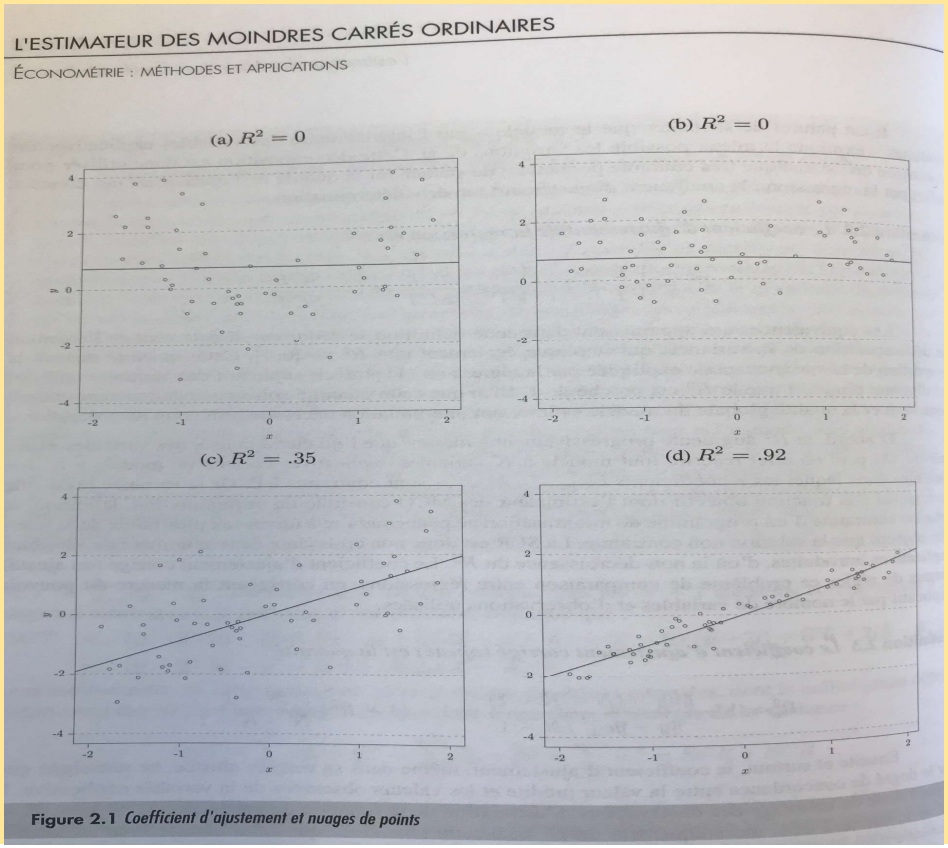
In practice case, this coefficient is rarely very nearly one of this tree closed boxes.

Compare the value of the coefficient correlation



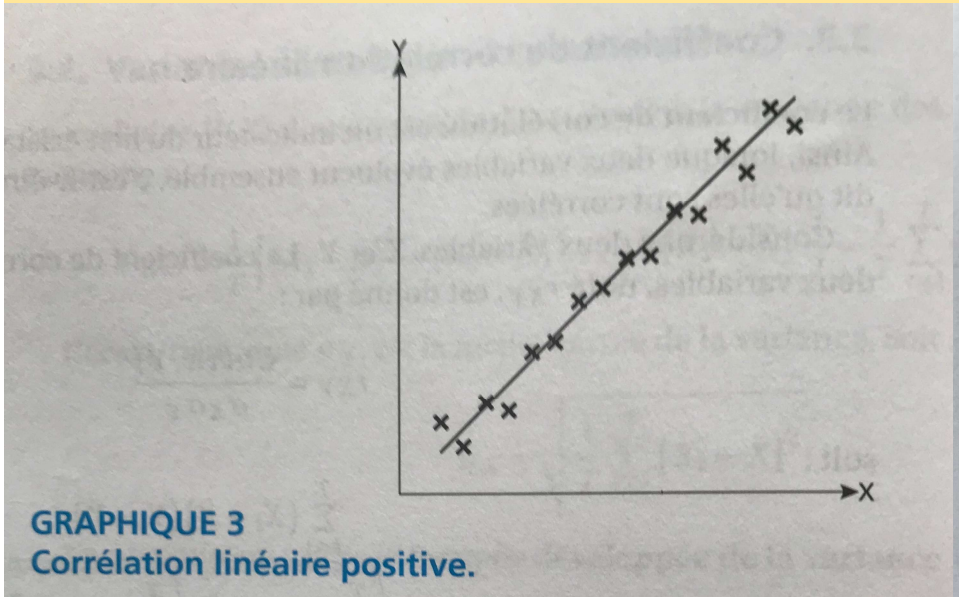
2. The level of correlation (cont.)

- Figure by Bruno Crépon and Nicolas Jacquemet,2018: Économétrie méthodes et applications.

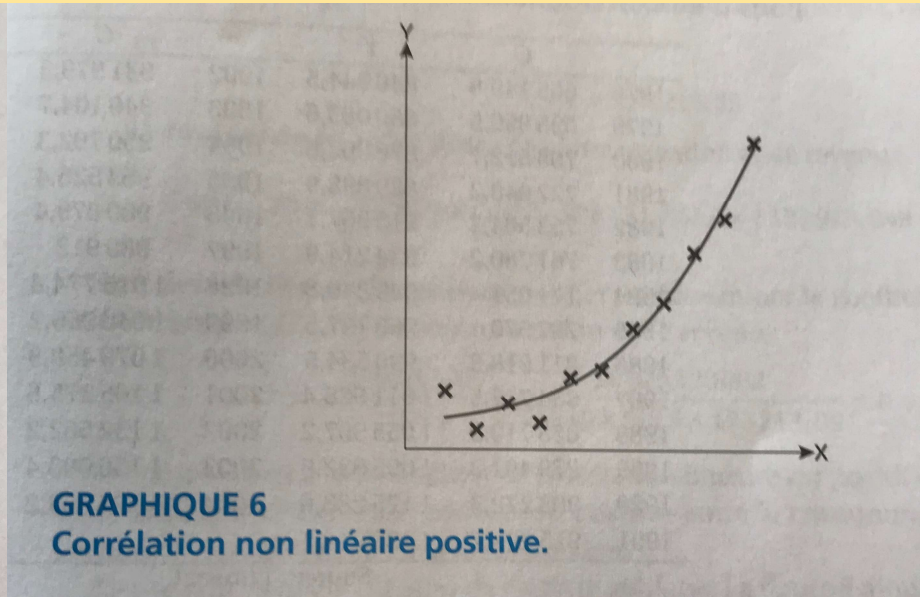


3. Representation type of correlation

Positive linear correlation



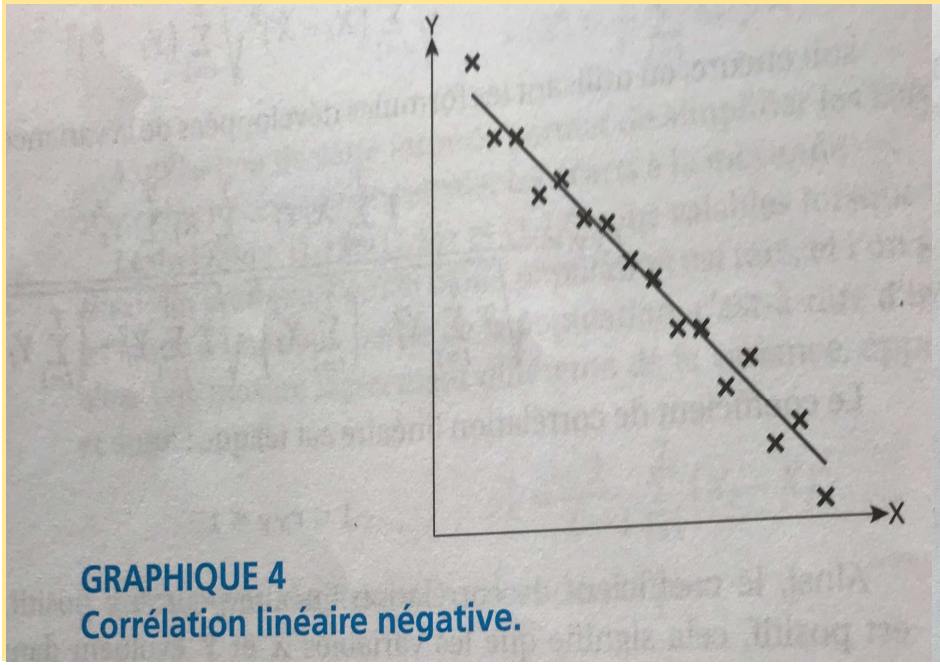
GRAPHIQUE 3
Corrélation linéaire positive.



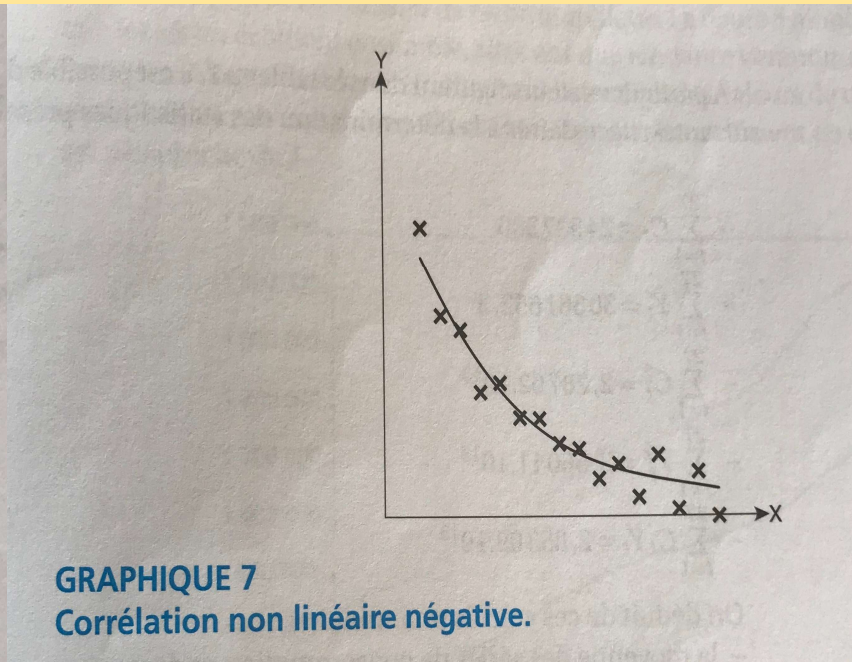
GRAPHIQUE 6
Corrélation non linéaire positive.

3. Representation type of correlation (cont.)

Negative linear correlation



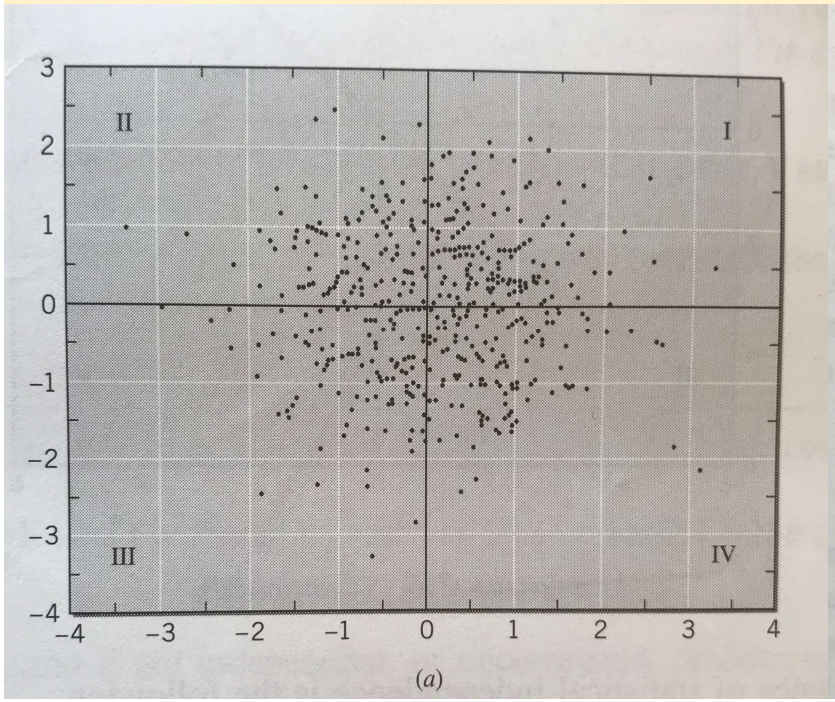
GRAPHIQUE 4
Corrélation linéaire négative.

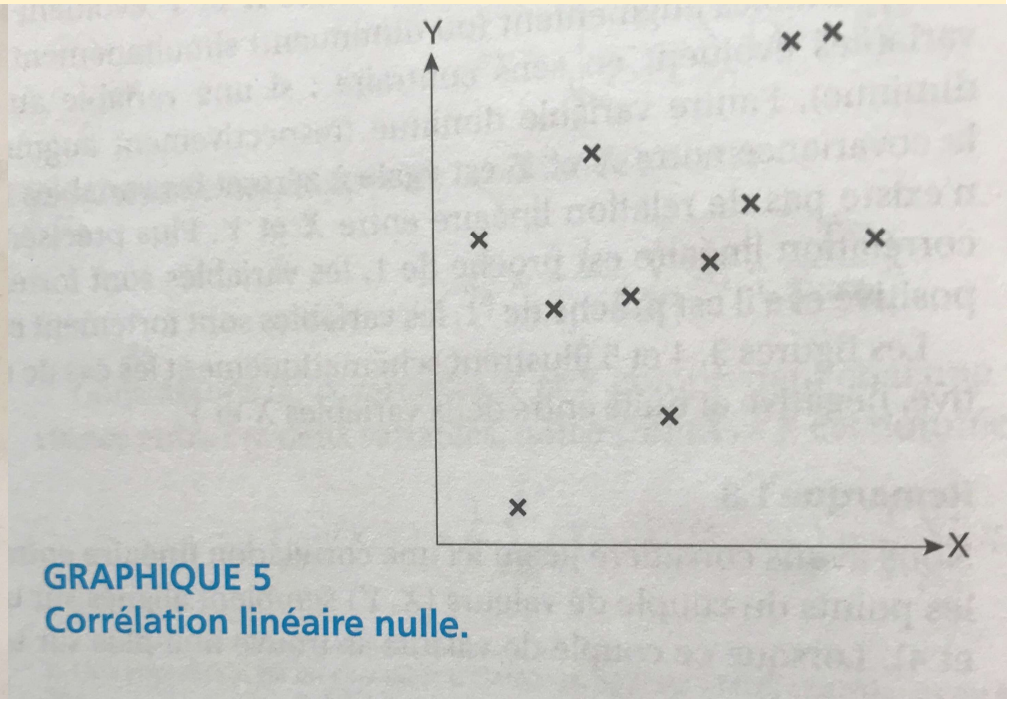


GRAPHIQUE 7
Corrélation non linéaire négative.

3. Representation type of correlation (cont.)

Non correlation between X and Y





GRAPHIQUE 5
Corrélation linéaire nulle.

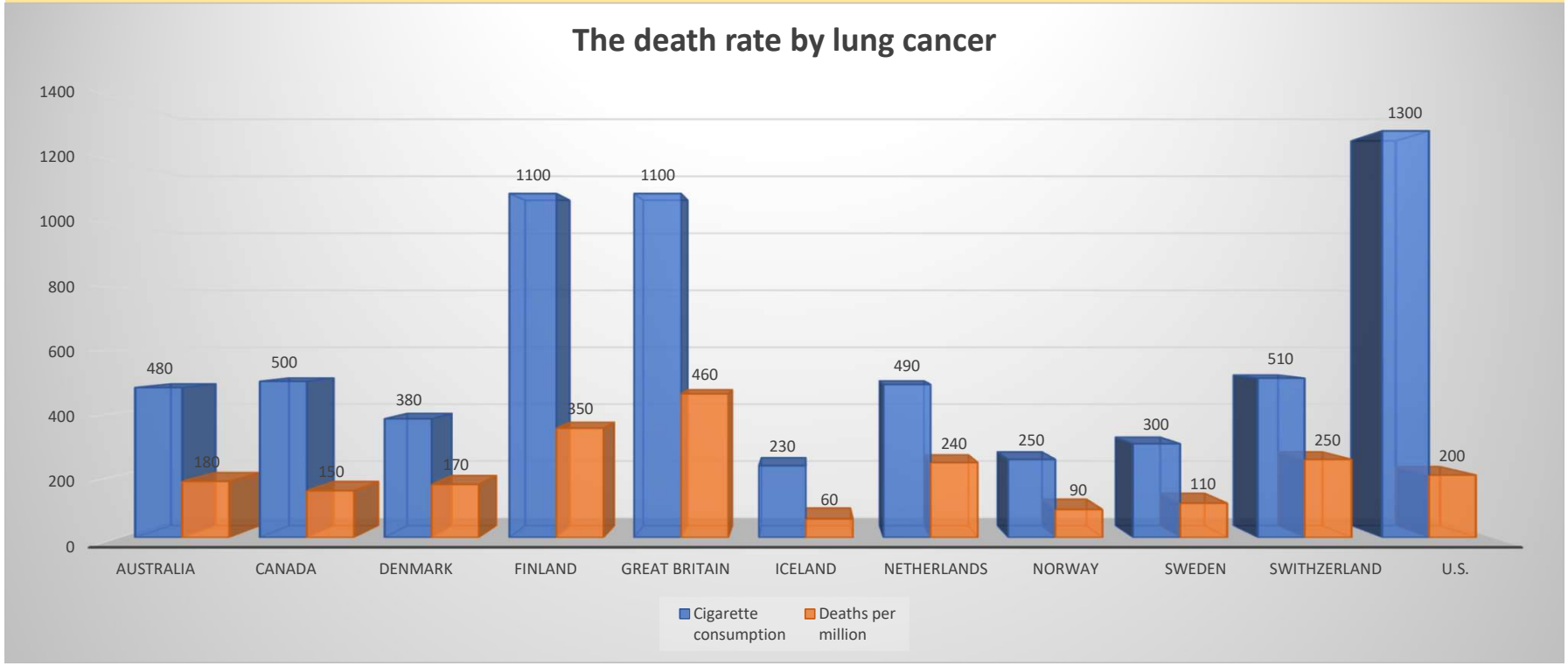
4. Numeric application

Example. True or fail, consumer Used their income for cigarettes for a long time since 1934 until 1950 deaths by lung cancer. Presentation by scatter graph and Calculate a R coefficient correlation.

No	Country	Cigarette consumption	Deaths per million
1	Australia	480	180
2	Canada	500	150
3	Denmark	380	170
4	Finland	1100	350
5	Great Britain	1100	460
6	Iceland	230	60
7	Netherlands	490	240
8	Norway	250	90
9	Sweden	300	110
10	Swithzerland	510	250
11	U.S.	1300	200

4. Numeric application(cont.)

The correlation scatter by graph, be low



4. Numeric application(cont.)

Calculate R coefficient

	Cigarette consumption = X	Deaths per million = Y	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})\,(Y_i - \bar{Y})$	$(\boldsymbol{X_i} - \bar{\boldsymbol{X}})^2$	$(\boldsymbol{Y_i} - \bar{\boldsymbol{Y}})^2$
	480	180	-123.6363	-25.4545	3147.1002	15285.93	647.9316
	500	150	-103.6363	-55.4545	5747.0992	10740.48	3075.202
	380	170	-223.6363	-35.4545	7928.9132	50013.19	1257.022
	1100	350	496.3637	144.5455	71747.139	246376.9	20893.4
	1100	460	496.3637	254.5455	126347.15	246376.9	64793.41
	230	60	-373.6363	-145.4545	54347.081	139604.1	21157.01
	490	240	-113.6363	34.5455	-3925.6228	12913.21	1193.392
	250	90	-353.6363	-115.4545	40828.902	125058.6	13329.74
	300	110	-303.6363	-95.4545	28983.451	92195	9111.562
	510	250	-93.6363	44.5455	-4171.0758	8767.757	1984.302
	1300	200	696.3637	-5.4545	-3798.3158	484922.4	29.75157
mean=	603.6363636	205.4545455					
		Somme total=	0.0007	0.0005	327181.82	1432255	137472.7

4. Numeric application(cont.)

Using formula,

$$R = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{327181.82}{\sqrt{1432255}\,\sqrt{137472.7}} = 0.7373451$$

The R coefficient value R = 0.7373451

The conclude that a deaths lung cancer by consumption cigarettes for long time since 1934 until 1954 is a correlation linear positive.

Whit is a level of confidence for this R coefficient correlation?

5. Test confident level of R coefficient correlation

- The coefficient correlation R is estimated by formula $R = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$
- According theory of test, permitted to test statistic determined
  i. Null hypothesis H0 : $R = 0$
  ii. Alternative hypothesis H1: $R \neq 0$
- Student score at degree of freedom n-2, we calculated t-statistic, it called t-student score

$$t^* = \frac{|R|}{\sqrt{\frac{1-R^2}{n-2}}}$$ (Régis Bourbonnais: Économétrie, 2015. 9th ed)

- If $t^* > t_{n-2}^{\alpha\%/2}$: reject H0, R is significant statistic at risk $\propto$ % .

5. Test confident level of R coefficient correlation (cont.)

- For our example precedence,
- Define hypothesis: $H0:\ R = 0, and\ \ H1: R \neq 0$
- We calculate the R coefficient correlation, using formula

$$t^* = \frac{|R|}{\sqrt{\frac{1-R^2}{n-2}}} = \frac{|0.7373451|}{\sqrt{\frac{1-0.7373451^2}{11-2}}} = 3.274585$$

- Use right-tail table critical values for the t-distribution
- $t^* = 3.274585 > t_{11-2}^{5\%/2} = 2.262$ reject H0, R coefficient correlation is significant at risk 5% and different zero. Thus, the deaths lung cancer by using cigarettes for a long time is correlated in confident level at risk 5%.

Example: presentation data by curve and calculate R coefficient correlation

Total Consumption and Gross Income for household In France(in million EURO),					
Year	C	INC	**Year**	C	INC
1978	665149,6	840944,5	1992	931979,3	1161063,2
1979	695990,5	860083,6	1993	940104,7	1149570,9
1980	709512,7	876002,6	1994	950702,3	1173503,5
1981	727046,2	889838,9	1995	963526,4	1201128,0
1982	753584,2	915957,1	1996	980879,4	1214274,2
1983	761780,2	934214,9	1997	989912,0	1243432,7
1984	771051,0	949840,8	1998	1015774,4	1287787,5
1985	787570,0	968197,5	1999	1046365,2	1330917,8
1986	811918,9	990534,5	2000	1079458,9	1385073,3
1987	834712,5	1011528,4	2001	1105375,6	1413487,4
1988	858710,9	1055507,2	2002	1132562,2	1430752,1
1989	879401,3	1095932,8	2003	1150003,4	1442220,8
1990	903272,3	1125288,6	2004	1175804,2	1475653,4
1991	915231,7	1138916,1			