Définition de modèle

- Dans le cadre de l'économétrie, un modèle consiste en une *présentation formalisée* d'un phénomène *sous forme d'équations* dont les *variables* sont des *grandeurs économiques*
- Représenter les traits les plus marquants d'une réalité à styliser
- Le modèle est donc l'outil utilisé pour comprendre et expliquer des phénomènes

Définition de modèle

- Pour ce faire, c'est nécessaire émettre des hypothèses et expliciter des relations
- Le modèle est donc une présentation schématique d'une réalité plus complexe.
- La difficulté principale est l'identification de la représentation intéressantes pour le problème à expliciter
- La même réalité peut ainsi être formalisée de diverses manières en fonction des objectifs

La construction des modèles

- En économie, les phénomènes étudiés concernent le plus souvent des comportements afin de mieux comprendre la nature et le fonctionnement des systèmes économiques.
- L'objectif du modélisateur est, au traverse d'une mesure statistique, de permettre aux agents économiques (l'Etat ou l'entrepreneur) d'intervenir de manière plus efficace.
- La construction d'un modèle est donc très importante pour obtenir une représentation correcte de la réalité.

La construction des modèles

- La construction d'un modèle peut articulé en différentes étapes
- En case de faiblesse d'un des 'maillons', le modèle peut se trouver invalide
- On peut considérer l'exemple de la construction d'un modèle a partir du modèle keynésien simplifié

La construction des modèles

- 1er étape: Référence à une théorie
  - Une théorie s'exprime au travers d'hypothèses auxquelles le modèle fait référence. Dans la théorie keynésienne:

    a. la consommation et le revenu sont liés;

    b. le niveau d'investissement privé et le taux d'intérêt sont liés

    c. il existe un investissement autonome public

    d. Le produit national est égal à la consommation plus l'investissement privé et public

La construction des modèles

- 2ème étape: Formalisation des relations et choix de la forme des fonctions:
  - Consommation et revenu: $C=f(Y)$, $f'>0$
  - Investissement et taux d'intérêt: $I=g(r)$, $g'<0$
  - Investissement autonome public: G
  - Enfin, le produit national $Y \equiv C+I+G$
- Nous n'avons postulé aucun forme particulière en ce qui concerne *f* et *g*.

La construction des modèles

- Des considérations théoriques nous renseignent sur le signes des dérivées.
- De tout façon, il y a des fonction de formes très différentes qui donnent des signes de dérivées identiques.
- $C = a_0+a_1Y$; $C=a_0Y^{a1}$
- Ces deux relations ne reflètent pas le même comportement.

La construction des modèles

- On appelle formes fonctionnelles ce choix de spécification précise du modèle
  - $C = a_0 + a_1 Y;\quad a_0 > 0$ et $0 < a_1 < 1$
    - a_1 = propension marginale à consommer
    - a_0 = consommation incompressible
  - $I = b_0 + b_1 r;\quad b_0 > 0$ et $b_1 < 0$
  - $Y \equiv C + I + G$
- Les deux premières équations reflètent des *relation* de *comportements* alors que la troisième est une *identité*

La construction des modèles

- 3ème étape: Sélection et mesure des variables
- Le modèle étant spécifié, il convient de collecter les variables représentatives des phénomènes.
- Ce choix n'est pas neutre:
  - Unité de mesure (euros constants ou courants)
  - Données bruts ou CVS
  - Quel taux d'intérêt

La construction des modèles

- Plusieurs types de données selon que le modèle est spécifié en:
  - Série temporelle: le cas plus fréquent en économétrie.
  - Coupe instantanée: données observées au même instant.
  - Panel: échantillon d'individus à intervalles réguliers.
  - Cohorte: constance de l'échantillon

La construction des modèles

- 4ème étape: Décalages temporels et modèle récursif
  - Dans le cadre des modèles spécifiés en séries temporelles, les relations ne sont pas toujours synchrones, mais peuvent être décalées dans le temps.
  - $C_t = f(Y_{t-1})$
  - Y_{t-1} est appelée variable endogène retardée.
  - Variable exogène: valeurs sont prédéterminées.
  - Variable endogène: valeurs dépendent des variables exogènes.
  - Modèles récursifs.

La construction des modèles

- 5ème étape: Validation du modèle
  - Le dernière étape de la construction du modèle est sa validation:
  - Les relations spécifiées sont-elles valides?
  - Peut-on estimer les coefficients avec suffisament de précision?
  - Le modèle est-il vérifié sur la totalité de la période?
  - Les coefficients sont-ils stables? Etc
- Les techniques économétriques s'efforcent d'apporter des réponses a ces questions.

Le rôle de l'économétrie

- Outil à la disposition de l'économiste qui lui permet d'infirmer de confirmer les théories qu'il construit.
- L'application de méthodes économétriques fournit des estimations sur la valeur des coefficients ainsi que la précision attendue.
- Etablir clairement les interrelations.
- Eviter d'omettre de liaisons importantes.
- Mesure de la confiance que l'économiste peut avoir en les estimations.

Le rôle de l'économétrie

- Outil d'analyse et investigation. Il est un aide à la modélisation
  - Mise en évidence de relations entre des variables économiques qui n'étaient pas évidentes.
  - L'induction statistique consiste à inférer à partir des caractéristiques d'un échantillon les caractéristiques d'une population.
  - Mesure de l'impact de la modification d'une variable sur une autre.
  - Prévision, utilisée afin d'anticiper et éventuellement réagir à l'environnement économique.

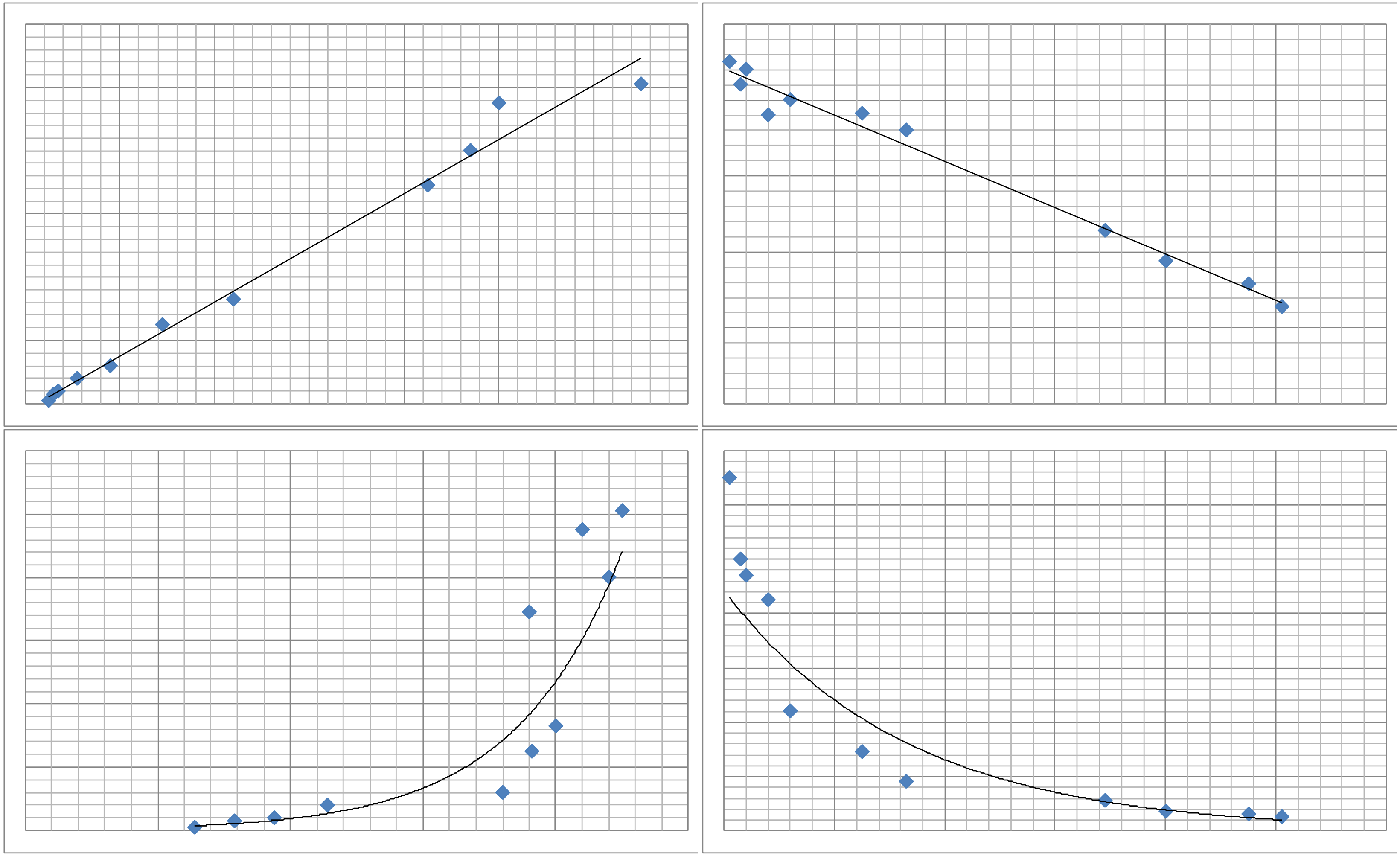
Théorie de la corrélation

- Lorsque deux phénomènes ont une évolution commune, nous disons qu'ils sont corrélés.
- Mesure du degré de liaison existant entre deux phénomènes représentés par deux variables.
- Corrélation **linéaire**: tous les points du couple de valeurs (x, y) semblent alignés sur une droite.
- Corrélation **non linéaire**: le couple de valeurs se trouve sur une même courbe d'allure quelconque.

Théorie de la corrélation

- Deux variables peuvent être:
- En corrélation positive: on constate une augmentation simultanée des valeurs des deux variables.
- Corrélation négative: lorsque le valeurs de l'une augmentent, les valeurs de l'autre diminuent.
- Non corrélées: il n'y a aucune relation entre les variations de l'une des variables et le valeurs de l'autre.

Théorie de la corrélation



Théorie de la corrélation

- Le coefficient de corrélation linéaire.
- La représentation graphique ne donne qu'une impression.
- On peut calculer le coefficient de corrélation linéaire simple:

$$r_{x,y} = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Théorie de la corrélation

- En développant la formule précédente, il vient:

$$r_{x,y} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

- On peut démontrer que, par construction ce coefficient reste compris entre -1 et 1:
  - proche de 1, le variables sont corrélées positivement
  - proche de -1, le variables sont corrélées positivement
  - Proche de 0 le variables ne sont pas corrélées

Théorie de la corrélation

- Le coefficient est rarement très proche a ces valeurs, et il est donc difficile de proposer une interprétation fiable.
- En plus, il n'est calculé qu'à partir d'un échantillon d'observations et non pas sur l'ensemble des valeurs.
- On appelle $\rho_{x,y}$ ce coefficient empirique, qui est une estimation du coefficient vrais $r_{x,y}$.

Théorie de la corrélation

- La théorie statistique nous permet de lever cette indétermination
- Soit l'hypothèse H0: $r_{x,y} = 0$ contre l'hypothèse H1: $r_{x,y} \neq 0$
- Sous l'hypothèse H0 nous pouvons démontrer que: $\frac{\rho_{x,y}}{\sqrt{\frac{(1-\rho_{x,y}^2)}{n-2}}}$

suit une loi de Student (n-2 degrés de liberté).

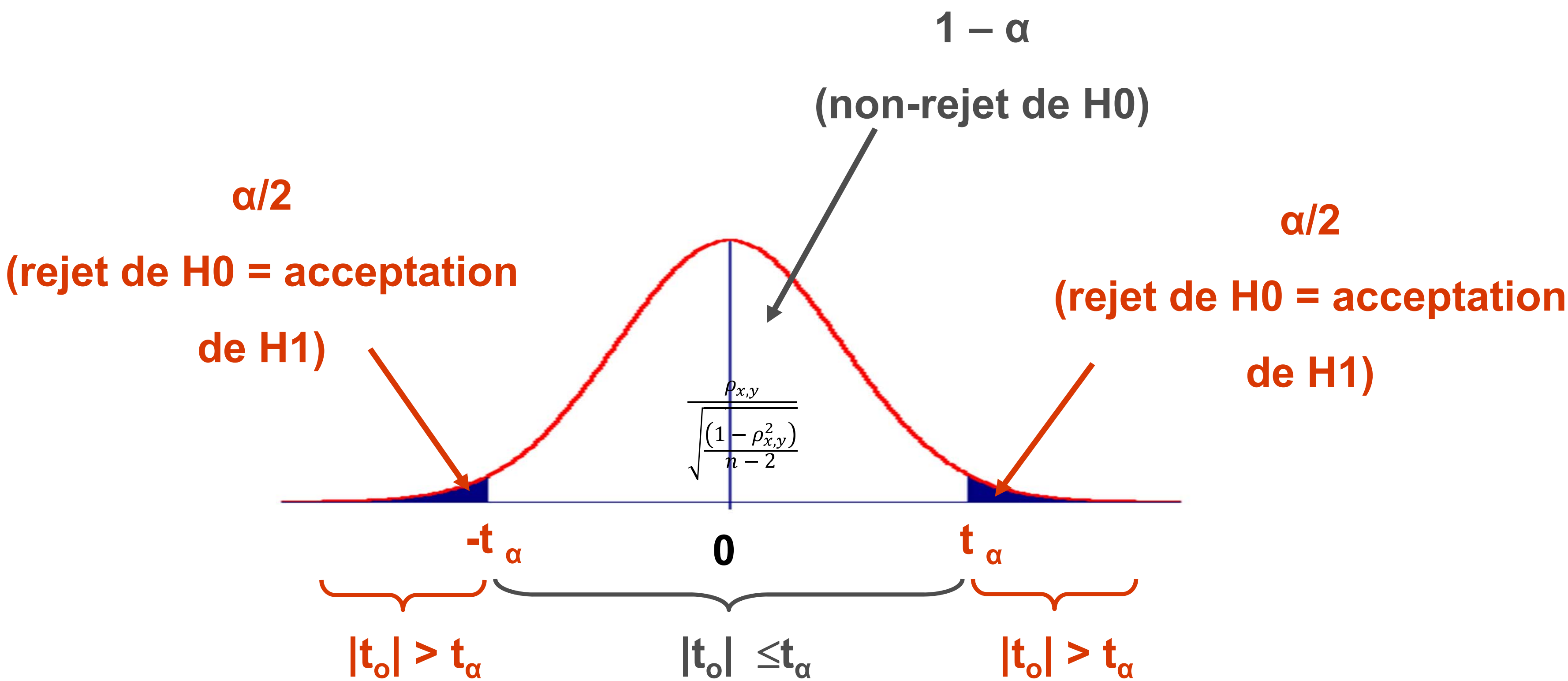
Théorie de la corrélation

$$t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{(1-\rho_{x,y}^2)}{n-2}}}$$

Si $t^* > t_{n-2}^{\alpha/2}$ valeur lue dans une table de Student[2] au seuil $\alpha = 0{,}05$ (5 %) à $n-2$ degrés de liberté[3], nous rejetons l'hypothèse H0, le coefficient de corrélation est donc significativement différent de 0 ; dans le cas contraire, l'hypothèse d'un coefficient de corrélation nul est acceptée. La loi de Student étant symétrique, nous calculons la valeur absolue du t empirique et nous procédons au test par comparaison avec la valeur lue directement dans la table.

L'application de la formule ne permet de déterminer que des corrélations linéaires.

Théorie de la corrélation



Abscisses : valeurs possibles de t sous H0 (ρ = 0)

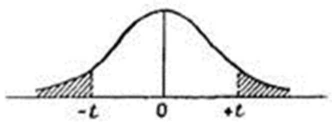
t_o : valeur observée/calculée de t sur l'échantillon

Méthodes Econométriques

Détermination du degré de signification associé à t_o (*P*-value)

Table de *t* (*).

La table donne la probabilité α pour que *t* égale ou dépasse, en valeur absolue, une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



d.d.l. \ α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	1,000	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,765	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,727	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,718	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,706	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,703	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,700	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,697	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,695	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,694	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,690	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,688	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,688	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,687	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,686	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,686	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,685	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,685	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,684	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,684	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,684	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,683	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,683	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,683	1,055	1,310	1,697	2,042	2,457	2,750	3,646
∞	0,126	0,674	1,036	1,282	1,645	1,960	2,326	2,576	3,291

(n-2) = 18

X

Rappel : P-value = probabilité d'observer une valeur plus grande que t_o sous l'hypothèse nulle H0

Théorie de la corrélation

- Pour pallier cette limite, il convient éventuellement de transformer les variables, préalablement au calcul du coefficient de corrélation
- Le fait d'avoir un coefficient de corrélation élevé ne signifie pas qu'il existe un autre lien que statistique.
- Corrélation ne signifie pas nécessairement une liaison d'ordre économique (corrélation n'est pas causalité)

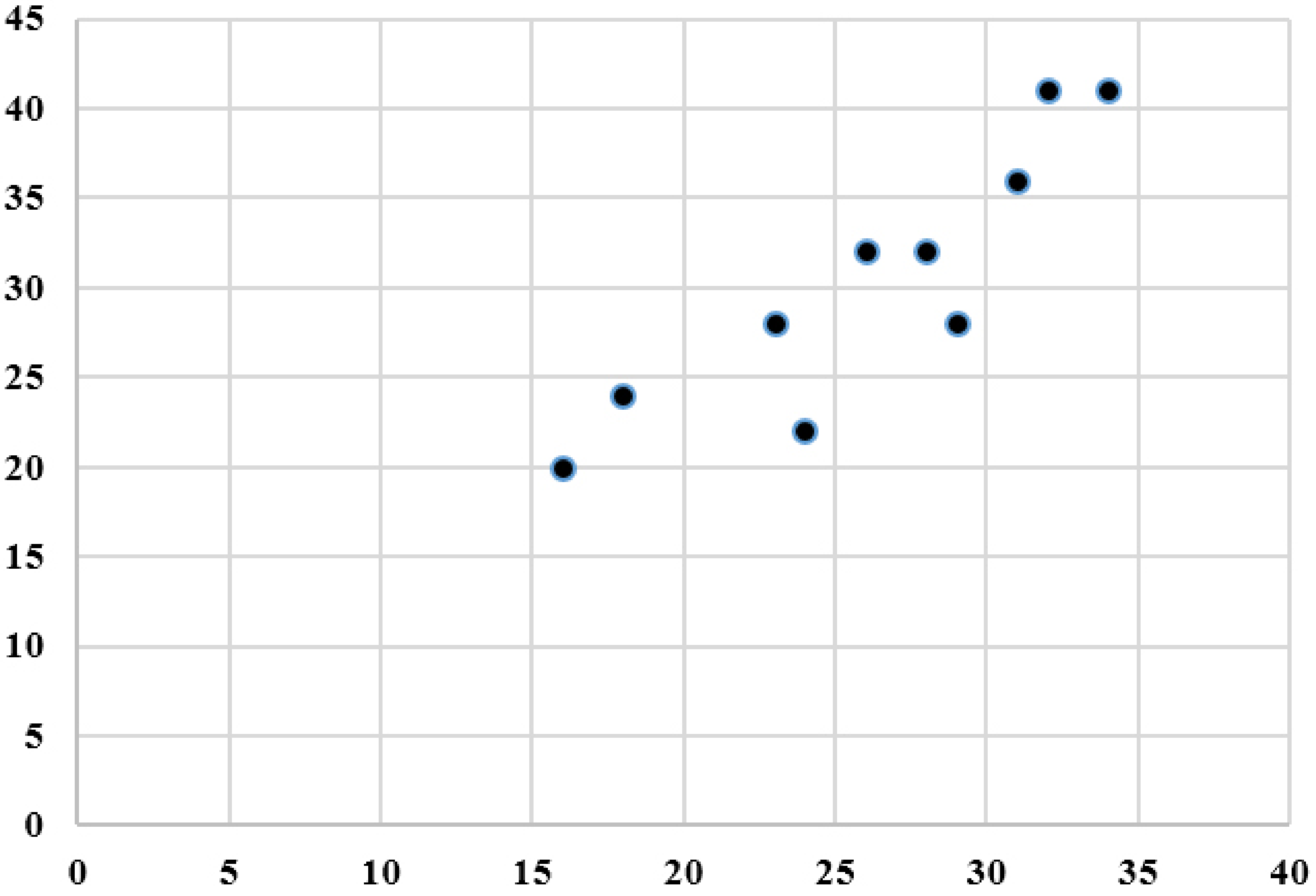
Exercices

Exercices 3

Un ingénieur s'intéresse à la liaison pouvant exister entre la productivité d'une machine (en tonne) et la quantité de matière première (en tonne). Il relève 10 couples de données consignés dans le tableau ci-dessous,

Rendement (x)	16	18	23	24	28	29	26	31	32	34
Matière première (y)	20	24	28	22	32	28	32	36	41	41

1. *Tracer le nuage de points et le commenter.*
2. *Calculer le coefficient de corrélation simple et tester sa signification par rapport à 0 pour un seuil* $\alpha = 0{,}05$, $\left(t_8^{0,025} = 2{,}306\right)$


45
40
35
30
25
20
15
10
5
0
0
5
10
15
20
25
30
35
40

y	x	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
20	16	-10.1	-10.4	102.01	108.16	105.04
24	18	-8.1	-6.4	65.61	40.96	51.84
28	23	-3.1	-2.4	9.61	5.76	7.44
22	24	-2.1	-8.4	4.41	70.56	17.64
32	28	1.9	1.6	3.61	2.56	3.04
28	29	2.9	-2.4	8.41	5.76	-6.96
32	26	-0.1	1.6	0.01	2.56	-0.16
36	31	4.9	5.6	24.01	31.36	27.44
41	32	5.9	10.6	34.81	112.36	62.54
41	34	7.9	10.6	62.41	112.36	83.74
				314.9	492.4	351.6

$$\bar{x} = 26.1$$
$$\bar{y} = 30.4$$

$$r_{x,y} = \frac{351{,}6}{\sqrt{314{,}9}\sqrt{492{,}4}} = 0{,}8929$$

89,29 %