haute école neuchâtel berne jura
arc gestion neuchâtel delémont

STATISTIQUE DESCRIPTIVE

Support de cours
Statistique 1

Sylvie Gonano

Version Septembre 2014

TABLE DES MATIERES DE LA STATISTIQUE DESCRIPTIVE

Chapitre 6 Les indicateurs de concentration

Chapitre 7 L'ajustement linéaire

Chapitre 8 Taux de croissance

Réponses aux exercices des chapitres 4, 5, 6, 7 et 8

CHAPITRE 1 INTRODUCTION

1.1 Définition

Le mot **statistique** a plusieurs sens:

- **les statistiques**: ensemble de données numériques publiées sous forme de tableaux, d'annuaires; informations chiffrées; notion du grand public;
  exemples: nombre de PME dans le canton de NE
  nombre de chômeurs en Suisse
  Ces données proviennent du **dénombrement**.

- **la statistique**: contenu de ce cours; ensemble de méthodes et d'outils qu'on utilise pour étudier et **analyser** des données numériques.

On distingue deux sortes de techniques appartenant à **la statistique**:

- la statistique **descriptive**: représentation de l'information d'une façon compréhensible et utilisable. Soit sous une forme graphique (histogrammes, diagramme circulaire,...), soit par des indicateurs qui résument les données de base (moyenne, écart-type,...).

- la statistique **inférentielle**: a pour but de généraliser l'information, c'est-à-dire d'étudier une partie de la population (échantillon) et d'en tirer des conclusions générales sur l'ensemble de la population.

1.2 Historique

Le mot **statistique** dérive du latin « status » qui signifie « état ». Il est donc tout naturel qu'au départ la statistique ait été liée à l'idée de dénombrement, d'inventaire et de recensement (état de la situation).
A l'époque des grandes civilisations de l'Antiquité déjà, on s'intéressait à la liste des hommes, des professions et des biens, car l'étendue et la complexité des empires nécessitaient une administration stricte et rigoureuse.
Les pratiques dans ce domaine ont bien entendu fortement évolué avec le temps, des méthodes nouvelles se sont développées et ainsi de la simple collecte de données, la statistique a peu à peu pris un sens nouveau.
A la notion de « comptage » ou de « mesure » est venu s'ajouter le problème de la **comparaison**. Or on ne peut pas comparer la masse entière des données, il faut obtenir des indicateurs qui résument l'ensemble des données. A ce moment, on peut comparer les différents indicateurs. C'est la naissance de la **statistique descriptive**.

1.3 Domaines d'utilisation

La statistique est utilisée dans de nombreux domaines, fort différents les uns des autres. Voici quelques exemples:

- **démographie**: étude de la croissance des populations humaines (taux de natalité, taux de mortalité, mouvement migratoire,...), étude de la structure de la population (caractéristiques personnelles, sociales et économiques,...).
- **biologie**: recherche fondamentale et expérimentations relatives aux principaux phénomènes présentés par les organismes vivants.
- **psychologie**: problèmes relatifs à l'évaluation de la capacité d'apprentissage, de l'intelligence, des caractéristiques personnelles et au comportement normal ou anormal de l'individu.
- **économie**: mesure du volume de production, des ressources, de l'emploi; analyse du comportement du consommateur, des réponses du marché aux changements de prix, de l'impact de la publicité.
- **santé**: coût des accidents, des soins médicaux, de l'hospitalisation.
- **médecine**: recherche fondamentale et expériences sur les causes, le diagnostic, le traitement et la prévention des maladies.

L'analyse statistique n'est pas une fin en soi. Qu'il s'agisse de schématiser ou d'expliquer, elle doit toujours oeuvrer dans le sens défini par une autre science. C'est pour cette raison que l'interprétation des résultats devrait toujours être faite par le spécialiste de la branche et non par le statisticien.

1.4 Vocabulaire

Population: ensemble statistique que l'on étudie.
Exemples: l'ensemble des personnes d'un pays mais aussi l'ensemble de la production d'une usine ou l'ensemble des prix d'articles de consommation,...

Individus: unités statistiques qui composent la population.
Exemples: chaque personne habitant un pays déterminé, chaque vélo produit par l'usine, chaque prix des différents articles,...

Echantillon: sous-ensemble d'une population sur lequel on effectue une étude statistique, en vue de tirer des conclusions relatives à la population.
Exemples: une partie des vélos produits que l'on teste afin de considérer toute la production de l'usine comme bonne ou non,...

Variable: caractéristique mesurable à laquelle on peut associer plusieurs valeurs (ou modalités) différentes. On distingue plusieurs types de variables:

- **variable qualitative catégorielle**: variable dont les valeurs sont des catégories. Sous forme de nomenclature.

- **variable quantitative**: variable dont les valeurs sont numériques. Il existe des variables **discontinues** (discrètes): nombres entiers (ex: nombre d'enfants par famille) et des variables **continues**: grand nombre de valeurs possibles dans un intervalle donné (ex: taille ou poids d'un individu, longueur d'un saut).

Catégorie: ensemble d'individus ayant une caractéristique commune.
Exemples: les catégories ouvriers, cadres, agriculteurs,... de la variable qualitative catégorielle « emploi ».

Variable dichotomique: variable qui ne peut prendre que deux valeurs.
Exemples: la variable qualitative catégorielle « sexe » qui ne peut prendre que les deux valeurs « féminin » et « masculin ». Notons que les variables quantitatives peuvent toujours être dichotomisées. Ainsi, la variable revenu, par exemple, peut être réduite en deux catégories: « bas revenu » et « haut revenu ».

Valeur: mesure quantitative ou qualitative associée à une variable.
Exemple de valeur quantitative: 0, 1, 2, 3,... enfants
Exemple de valeur qualitative: rouge, bleu, vert,...

1.5 Exercices

1. Pour chaque variable proposée, indiquer s'il s'agit d'une **variable qualitative**, d'une **variable quantitative discrète** ou d'une **variable quantitative continue**.

   a) la taille des individus pour une population donnée
   b) le nombre d'enfants par ménage pour une localité donnée
   c) le chiffre d'affaires mensuel d'une grande surface
   d) les notes des élèves d'une classe dans une branche donnée:
      d1) les notes sont arrondies au point (1, 2, 3,...,6)
      d2) les notes sont mises au 1/10 de point (1.0, 1.1, 1.2,...,5.9, 6.0)
   e) l'âge des personnes d'une population donnée
   f) le nombre de fois **face** obtenu lorsque l'on répète 100 fois l'expérience « jet d'une pièce de monnaie»
   g) la nationalité des résidents d'un pays à un moment donné
   h) le nombre d'articles défectueux dans une production journalière
   i) la couleur des voitures d'une marque donnée

1.6 Bibliographie

Bressoud Etienne, Statistique descriptive : applications avec Excel et calculatrices, Pearson, Paris, 2010.
Favre Jean-Pierre, Mathématiques de gestion, Digilex, Epalinges, 2012.
Gasquet Sylviane, Plus vite que son nombre : déchiffrer l'information, Ed. du Seuil, Paris, 1999.
Hahn Corinne, Méthodes statistiques appliquées au management, Pearson, Montreuil, 2012.
Morineau Alain, L'analyse statistique des données : apprendre, comprendre er réaliser avec Excel, Ellipses, 2005.
Pagès Jérôme, Statistiques générales pour utilisateurs, Presses universitaires de Rennes, 2005-2008.
Py Bernard, La statistique sans formule mathématique: avec 150 questions et exercices corrigés d'entraînement aux examens, Pearson, Paris, 2010.
Tribout Brigitte, Statistique pour économistes et gestionnaires, Pearson, Paris, 2013.
Walas Muriel, Outils statistiques pour le management : une approche par étapes, Pearson, Montreuil, 2013.

1.7 Les principales sources statistiques pour la Suisse

1. **Revue de politique économique**: Cette publication mensuelle du Département fédéral de l'économie publique fournit les principales données économiques et publie les résultats d'enquêtes annuelles, semestrielles ou trimestrielles.
   www.seco.admin.ch

2. **Les publications de la Banque Nationale Suisse**: bulletin mensuel, bulletin trimestriel et rapport de gestion.
   www.snb.ch

3. **Annuaire statistique de la Suisse**: publication annuelle de l'Office fédéral de la statistique.
   www.admin.ch/bfs

4. **Statistiques de l'OCDE**: rapport sur la Suisse (annuel), comptes nationaux (annuel), principaux indicateurs économiques (annuel).
   www.oecd.org

5. **Statistiques du canton de Neuchâtel :** tableaux et graphiques de l'Annuaire statistique du canton de Neuchâtel.
   www.ne.ch/stat

CHAPITRE 2 CLASSEMENT DES DONNEES ET TABLEAUX STATISTIQUES

2.1 Vocabulaire

Donnée / Observation: une donnée est le résultat d'une observation faite sur une population ou sur un échantillon. Les données sont liées à la variable étudiée. On dira ainsi que les données sont qualitatives, quantitatives discrètes ou continues si la variable associée est elle-même qualitative, quantitative discrète ou continue.

Effectif: correspond au nombre d'apparitions d'une valeur.
Exemple: nombre de ménages avec 2 enfants.
On trouve aussi le terme de **fréquence absolue**.

Fréquence: définie par le **rapport** du nombre d'apparitions au nombre total d'observations.
Exemple: soit les 8 observations suivantes (tailles en cm d'un groupe de personnes):
178 , 171 , 185 , 175 , 178 , 160 , 180 , 174
L'effectif de l'observation (ou de la valeur) **178** est égal à 2.
La fréquence est égale à 2/8 = 0.25.
On trouve aussi le terme de **fréquence relative**.

Pourcentage: défini par le **rapport** du nombre d'apparitions au nombre total d'observations **multiplié par 100.**
Exemple: soit les 8 observations identiques à celles de l'exemple ci-dessus, le pourcentage de la valeur **178** est égal à 25%.

Effectif cumulé: sert à déterminer le **nombre** d'observations dont la valeur est « inférieure ou égale » ou « supérieure » à une valeur donnée.
Exemple: dans l'exemple précédent, il y a **6 personnes** qui mesurent **moins de 178 cm ou 178 cm exactement** (par soustraction il y a donc 2 personnes qui mesurent plus de 178 cm).

Fréquence cumulée: sert à déterminer la **proportion** de valeurs « inférieures ou égales » ou « supérieures » à une valeur donnée.
On trouve aussi le terme de **fréquence relative cumulée.**
Exemple: dans l'exemple précédent, il y a 6 personnes sur 8 = **0.75** qui mesurent 178 cm ou moins. On peut transformer la fréquence cumulée en **pourcentage cumulé**. On dit alors qu'il y a 75 % des personnes qui mesurent 178 cm et moins (par soustraction il y en a 25 % qui mesurent plus de 178 cm).

Classes: **intervalles** sur les nombres réels qui s'excluent les uns des autres.

Bornes/extrémités de classe: valeur **minimale** et valeur **maximale** de la classe.

Amplitude de classe: c'est la taille de la classe, c'est-à-dire la **différence** entre la valeur maximale et la valeur minimale de la classe.

Intervalle [... ; ...[: intervalle **fermé à gauche** et **ouvert à droite**. On peut aussi écrire $a \leq x < b$.
Exemple: au lieu d'écrire une classe « de 40 à moins de 50 », on simplifie l'écriture et on note la classe: **[40 ; 50[**. De même, si on écrit **40 - 50**, la valeur 40 est comprise dans la classe, alors que la valeur 50 ne l'est pas.

2.2 Etablissement d'un tableau statistique

Les règles à respecter lorsque l'on établit un tableau statistique sont les suivantes:

a) **intituler** le tableau

b) **en-têtes** des lignes et des colonnes

c) mentionner les **unités**

d) citer la **source** du tableau

2.3 La fonction *FREQUENCE* d'Excel

La fonction matricielle *FREQUENCE* permet de trouver les **effectifs** correspondant aux différentes classes à partir de données brutes.

Marche à suivre

1) Dans une feuille de calcul, entrer les données brutes, puis les valeurs des bornes supérieures dans une autre colonne. Dans une colonne adjaçante, sélectionner une plage vide de la même grandeur que les bornes supérieures.

2) Insertion / Fonction / Statistiques / **Fréquence**.

3) Compléter la boîte de dialogue :

- tableau_données : sélectionner la plage des données brutes dans la feuille de calcul à l'aide de la souris ;
- matrice_intervalles : sélectionner la plage des bornes supérieures dans la feuille de calcul à l'aide de la souris ;

4) Pour valider la fonction, taper simultanément, après avoir positionné le curseur sur la barre de formule à la fin de la fonction, les trois touches **SHIFT(↑), Ctrl et Entrée (return)** puisqu'il s'agit d'une fonction matricielle.

Remarques :

1. Contrairemenr à l'usage retenu, la fonction *FREQUENCE* incorpore dans la classe l'observation dont la valeur est confondue avec la borne supérieure. Ainsi, l'individu de taille **170** appartient à la classe **de borne supérieure 170**, alors qu'il devrait appartenir à la classe suivante (c'est-à-dire **de borne supérieure 180**).

2. On peut auparavant déterminer la valeur minimale et la valeur maximale des données brutes, afin de savoir quelles classes former. On utilisera les fonctions ***MIN*** et ***MAX*** d'Excel.

2.4 Exercices

1. Une étude sur le poids des élèves de trois classes fournit les **observations** suivantes (en kg):

49	83	89	86	65	65	77	73	62	58
52	57	59	92	85	83	77	72	66	63
69	64	63	50	88	84	84	77	67	57
52	86	73	70	70	65	60	54	47	84
76	67	60	70	69	64	56	62	74.	73
75	62	87	69	62	68	56	61	83	76
61	69	67	76	73	70	69	71	84	73

   a) Classer les données en classes d'**amplitude 5** et calculer les **effectifs**.

   b) Compléter le tableau avec les colonnes **Fréquences, Pourcentages, Effectifs cumulés croissants, Effectifs cumulés décroissants, Pourcentages cumulés croissants et décroissants.**

   c) Refaire la démarche avec Excel.

2. Un contrôle de qualité relatif au diamètre des pièces usinées fournit les **mesures** suivantes exprimées en millimètres:

19.87	19.95	19.84	20.07	19.93	20.06	19.97	19.92	20.12	20.04
19.70	19.90	20.08	19.85	20.07	20.09	19.91	20.09	20.10	19.99
19.96	20.04	20.01	19.98	19.99	19.87	19.94	19.82	20.01	20.13
20.11	19.84	20.06	20.08	20.12	20.02	19.91	20.04	19.94	20.02
19.95	20.03	19.93	19.97	19.94	19.96	20.26	20.15	19.72	20.14
20.02	20.16	20.11	20.02	19.90	19.93	20.05	19.89	19.86	19.98
19.88	20.00	20.18	20.03	19.97	20.05	20.03	19.81	20.06	19.80
19.89	19.96	20.10	20.21	20.00	19.98	19.95	20.05	20.17	19.99

a) Présenter sous forme de tableau la distribution des effectifs en regroupant les valeurs du caractère en classes de 5 centièmes de millimètres, sauf la première et la dernière qui seront respectivement [19.70 ; 19.80[et [20.20 ; 20.30[.

b) Compléter le tableau avec les colonnes **Fréquences**, **Pourcentages**, **Effectifs cumulés croissants**, **Effectifs cumulés décroissants**, **Pourcentages cumulés croissants et décroissants.**

c) Refaire la démarche avec Excel.

CHAPITRE 3 LES REPRESENTATIONS GRAPHIQUES

3.1 Vocabulaire

Les types de graphiques les plus utilisés pour représenter des informations numériques provenant de l'étude d'une **variable qualitative catégorielle** sont :

- le **diagramme circulaire**
- le **diagramme en barres**
- le **diagramme figuratif**.

Diagramme circulaire: graphique constitué d'un cercle divisé en secteurs. Sert à représenter des données formant les différentes parties d'un tout.

Diagramme en barres: consiste en une suite de barres verticales ou horizontales de largeur identique et de longueur égale aux quantités à représenter.

Diagramme figuratif: construits à l'aide de figures (bouteilles, automobiles, personnes, ...) de dimension proportionnelle aux quantités à représenter ou bien à l'aide de figures de même taille que l'on reproduit un certain nombre de fois suivant les quantités à représenter.

Diagramme en bâtons: graphique utilisé pour représenter des distributions à **caractère discontinu**. Se présente sous la forme d'un graphe où l'on met en abscisse les valeurs de la variable et en ordonnée les effectifs (ou les fréquences).

Histogramme: graphique utilisé pour représenter des distributions à **caractère continu**. Consiste en un ensemble de rectangles. En abscisse, on a les différentes classes et en ordonnée les effectifs (ou fréquences).

Polygone des effectifs/ fréquences: droite segmentée reliant les points centraux des sommets de chaque rectangle d'un histogramme.

Courbe des effectifs/ fréquences: « polissage » d'un polygone des effectifs. On peut effectuer ce genre de polissage si le nombre d'observations devient très grand et si l'amplitude des classes devient très petite (très grand nombre de classes)

3.2 Excel et les représentations graphiques

Marche à suivre

1) Taper les données qui feront l'objet d'une représentation graphique dans une feuille de calcul sans oublier les intitulés de lignes et de colonnes.

2) Sélectionner la zone des données sur la feuille de calcul.

3) Insertion / Graphique.

4) Laissez-vous prendre en charge par l'assistant graphique. Il faut faire ensuite plusieurs essais afin d'obtenir la représentation graphique la plus convaincante possible.

3.3 Exercices

1. Soit les taux de chômage exprimés en pourcentage de la population active:

Pays	2010	2011	2012
Suisse	3.5	2.8	2.9
Allemagne	7.1	6.0	5.5
France	9.7	9.6	10.3
Italie	8.4	8.4	10.7
Grande-Bretagne	7.8	8.0	7.9
UE	10.1	10.1	11.4
Etats-Unis	9.6	9.0	8.1
Japon	5.1	4.6	4.4

Source: La Vie économique, Revue de politique économique, 7/8 – 2013

a) Présenter graphiquement ces données pour qu'elles représentent au mieux le phénomène étudié.

b) Reporter ces données dans une feuille de calcul Excel et construire la/les figure(s) appropriée(s).

2. Soit la population résidante en Suisse, selon l'âge en 2012:

Classes d'âges	En milliers
0 – 20	1641.3
20 – 40	2143.2
40 – 65	2853.4
65 et plus	1399.0

Source: La Vie économique, Revue de politique économique, 7/8 – 2013

Reporter ces données dans une feuille de calcul Excel et construire la/les figure(s) appropriée(s).

3. Soit le bilan analytique suivant:

ACTIF (Milliers de francs)		PASSIF (milliers de francs)	
Disponibilités	75	Dettes à court terme	148
Val. réalisables à court terme	171	Dettes à long terme	55
Valeurs d'exploitation	80	Fonds propres	461
Actif immobilisé	338		
Total	664		664

a) Présenter graphiquement ces données pour qu'elles représentent au mieux le phénomène observé.

b) Reporter ces données dans une feuille de calcul Excel et construire la/les figure(s) appropriée(s).

4. Le tableau ci-dessous indique les notes obtenues par les candidats à l'examen écrit de statistique. Les notes sont arrondies à l'unité.

Notes	Nombre de candidats
1	1
2	4
3	6
4	14
5	8
6	5
	Total: 38

a) De quel type de variable s'agit-il ?

b) Un candidat a-t-il obtenu la note 5,5 ?

c) Combien y a-t-il d'observations ?

d) Combien y a-t-il d'observations distinctes ?

e) Construire manuellement la représentation graphique appropriée.

5. Soit le taux de chômage en Suisse par classe d'âge en mai 2013:

Age	Taux de chômage (%)
15-20	1.6
20-25	3.6
25-30	3.9
30-35	3.8
35-40	3.4
40-45	2.8
45-50	2.7
50-55	2.8
55-60	2.6
60 et plus	2.4

Source: La Vie économique, Revue de politique économique, 7/8 – 2013

a) Présenter graphiquement ces données pour qu'elles représentent au mieux le phénomène étudié.

b) Reporter ces données dans une feuille de calcul Excel et construire la/les figure(s) appropriée(s).

6. Reprendre l'exercice 1 du chapitre précédent.

a) Présenter graphiquement les données du tableau que vous avez construit pour qu'elles représentent au mieux le phénomène observé.

b) Reporter ces données dans une feuille de calcul Excel et construire la/les figure(s) appropriée(s).

7. Idem pour l'exercice 2 du chapitre précédent.

8. Une enquête fiscale a permis d'établir le tableau suivant :

Classes du revenu net (en francs)	Contribuables
0 – 20'000	600
20'000 – 30'000	4'000
30'000 – 35'000	6'000
35'000 – 40'000	5'800
40'000 – 45'000	6'200
45'000 – 50'000	5'600
50'000 – 60'000	4'800
60'000 – 80'000	4'200
80'000 – 100'000	3'000
100'000 – 125'000	2'800
125'000 – 150'000	1'200
150'000 – 200'000	400

a) De quel type de variable s'agit-il ?

b) Quelle est la particularité de cette distribution statistique ?

c) Construire manuellement l'histogramme.

9. Voici l'évolution du nombre de danseurs entre 2006 et 2012.

Année	Nombre de danseurs
2006	50
2007	70
2008	130
2009	120
2010	100
2011	90
2012	80

A l'aide d'Excel, représenter les observations ci-dessus sous forme d'**histogramme**, puis remplacer les rectangles par une **image importée**.

10. A l'aide d'Excel, établir une **pyramide des âges** sur la base des données ci-dessous.

Population résidante permanente de la Suisse par groupe d'âges quinquennaux, ainsi que selon le sexe au 31 décembre 2004 (en milliers).

Age	*Hommes*	*Femmes*
0 - 5	187.9	177.5
5 - 10	206.9	195.7
10 - 15	224.6	212.9
15 - 20	222.0	211.7
20 - 25	222.3	219.1
25 - 30	230.0	233.6
30 - 35	265.0	267.9
35 - 40	309.1	309.0
40 - 45	313.1	308.2
45 - 50	277.8	271.5
50 - 55	246.2	245.4
55 - 60	238.4	238.7
60 - 65	199.4	207.0
65 - 70	152.2	171.5
70 - 75	127.2	158.4
75 - 80	96.6	140.4
80 - 85	66.0	113.6
85 - 90	30.0	62.5
90 et plus	14.1	42.0
Total	3628.7	3786.4

Source : Office fédéral de la Statistique – Section de l'évolution de la population

CHAPITRE 4 LES MESURES DE TENDANCE CENTRALE

4.1 Introduction

Une mesure de tendance centrale permet de résumer un ensemble de données relatives à une **variable quantitative**. Plus précisément, elle permet de déterminer une **valeur** fixe, appelée valeur **centrale**, autour de laquelle l'ensemble des données a tendance à se rassembler. Les principales mesures de tendance centrale sont:

- la **moyenne arithmétique**
- la **médiane**
- le **mode**.

4.2 Le mode

Le mode d'un ensemble d'observations est la **valeur** que l'on rencontre le plus fréquemment, c'est-à-dire celle qui a la plus grande **fréquence** (ou le plus grand **effectif**).

4.2.1 Cas des variables quantitatives discrètes (discontinues)

On peut utiliser la fonction ***MODE*** d'Excel.

Exercices

1. Indiquer le mode pour les deux séries d'observations suivantes:

a) 7 , 8 , 9 , 3 , 3 , 7 , 6 , 7 , 8 , 7 , 3 , 9 , 6 , 7

b) 4 , 4 , 7 , 6 , 8 , 12 , 6 , 4 , 8 , 7 , 8 , 13 , 4 , 8 , 6 , 5

c) utiliser la fonction *MODE* d'Excel pour vérifier les résultats. Quelles constatations peut-on faire ?

2. Une assurance a enregistré le nombre de sinistres par assuré en trois ans. Quel est le **mode** de cette distribution ?

Nombre d'accidents	Nombre d'assurés
0	30
1	100
2	180
3	145
4	105
5	84
6	25
7	3

4.2.2 Cas des variables quantitatives continues

Le mode est calculé à l'aide de la formule suivante :

$$\mathbf{Mo} = \mathbf{A} + \mathbf{a} \cdot \frac{k_1}{k_1 + k_2}$$

où A est la borne inférieure de la classe modale, a est l'amplitude de la classe modale, k_1 et k_2 sont les différences entre les **effectifs rectifiés** de la classe modale et des deux classes qui l'entourent.

Exercices

3. A partir de quelle **représentation graphique** détermine-t-on le mode

   a) pour les variables de type **discret** ou **discontinu** ?

   b) pour les variables de type **continu** ?

4. Déterminer **graphiquement** le mode pour les séries statistiques

   a) Poids des élèves (cf. chapitre 3, exercice 6)

   b) Contrôle de qualité (cf. chapitre 3, exercice 7)

5. Indiquer la **classe modale**, puis calculer le **mode** pour la série statistique suivante:

Classes (diamètres en mm)	Effectifs
20-25	9
25-30	27
30-35	36
35-40	45
40-45	18
45-50	9
50-55	3
55-60	3

6. Indiquer la **classe modale**, puis calculer le **mode** pour les séries statistiques

   a) Poids des élèves (cf. chapitre 2, exercice 1)

   b) Contrôle de qualité (cf. chapitre 2, exercice 2)

7. Indiquer la **classe modale**, puis calculer le **mode** pour la série statistique suivante:

Classes (diamètres en mm)	Effectifs
30,00-40,00	12
40,00-42,50	13
42,50-45,00	20
45,00-50,00	30
50,00-52,50	16
52,50-55,00	8

4.3 La médiane

La médiane est définie comme la valeur qui se trouve au **centre** d'un ensemble d'observations lorsque celles-ci sont rangées par ordre croissant ou décroissant. Nous trouvons donc **50 % des observations de chaque côté** de la médiane.

4.3.1 Cas des variables quantitatives discrètes (discontinues)

Lorsque les observations nous sont données de **façon individuelle**, le processus de calcul de la médiane est simple:

1. classer les n observations par ordre de grandeur

2. si le nombre d'observations est **impair**
   - repérer l'observation du milieu $\frac{n+1}{2}$
   - la médiane est égale à la **valeur de l'observation du milieu**

3. si le nombre d'observations est **pair**

   - repérer les deux observations du milieu $\frac{n}{2}$ *et* $\frac{n}{2}+1$
   - la médiane est égale à la **moyenne arithmétique** des **valeurs** de ces deux observations.

On peut utiliser la fonction ***MEDIANE*** d'Excel.

Exercices

1. Déterminer la **médiane** pour chacune des trois séries d'observations suivantes:

a) 3 , 3 , 4 , 4 , 4 , 6 , 6 , 7 , 8

b) 2 , 40 , 45 , 47 , 49 , 50 , 60 , 62 , 63 , 1000

c) 74 , 3 , 88 , 72 , 75 , 99 , 68 , 100 , 8 , 25

2. Utiliser la fonction *MEDIANE* d'Excel pour vérifier les résultats obtenus à l'exercice 1. Quelles constatations peut-on faire ?

3. Déterminer la médiane pour la série statistique suivante :

Nombre d'enfants par ménage	Effectifs	
0	712	
1	1701	
2	1422	
3	714	
4	180	
5	35	
Total		x x x x x

4.3.2 Cas des variables quantitatives continues

Lorsque les observations sont groupées en **classes**, la médiane se détermine de la façon suivante:

1. déterminer la **classe médiane** (celle qui comprend la $\frac{n}{2}$-ème observation ou $\frac{n+1}{2}$ si n est impair)
2. calculer la médiane par la formule suivante:

$$\text{médiane} = A + \left(\frac{n}{2} - ec_{\inf}\right) \cdot \frac{a}{e_{med}} \qquad \text{(remplacer } \frac{n}{2} \text{ par } \frac{n+1}{2} \text{ si n est impair)}$$

où A = borne inférieure de la classe médiane
n = nombre total d'observations
ec_{inf} = effectif cumulé croissant des classes inférieures à la classe médiane
a = amplitude de la classe médiane
e_{med} = effectif de la classe médiane

Exercices

4. Indiquer la **classe médiane**, puis calculer la **médiane** pour la série statistique suivante:

Classes (salaire horaire en francs)	**Effectifs**	
< 20	712	
20 - 30	1701	
30 - 40	1422	
40 – 50	714	
> 50	620	
Total		x x x x x

5. Indiquer la **classe médiane**, puis calculer la **médiane** pour les séries statistiques

a) Poids des élèves (cf. chapitre 2, exercice 1)

b) Contrôle de qualité (cf. chapitre 2, exercice 2)

6. Indiquer la **classe médiane**, puis calculer la **médiane** pour la série statistique suivante:

Classes (diamètres en mm)	Effectifs	
20-25	9	
25-30	27	
30-35	36	
35-40	45	
40-45	18	
45-50	9	
50-55	3	
55-60	3	
Total		x x x x x

7. Indiquer la **classe médiane**, puis calculer la **médiane** pour la série statistique suivante:

Classes (diamètres en mm)	Effectifs	
30,00-40,00	12	
40,00-42,50	13	
42,50-45,00	20	
45,00-50,00	30	
50,00-52,50	16	
52,50-55,00	8	
Total		x x x x x

4.4 La moyenne arithmétique

La moyenne arithmétique est la valeur centrale la plus utilisée. Sans mention spéciale, chaque fois que l'on parlera de moyenne, il s'agira d'une moyenne arithmétique.
Il y a deux sortes de moyenne arithmétique:

- **la moyenne arithmétique simple**
- **la moyenne arithmétique pondérée**

Exercices sur la moyenne arithmétique

1. **a)** Calculer la moyenne arithmétique simple pour la série de températures suivantes exprimées en degrés:

   5 7 9 15 16 17.5 16.5 14.5 8 4 3 -2 -6 -8 -20

   b) Utiliser la fonction ***MOYENNE*** d'Excel pour vérifier le résultat obtenu.

2. **a)** Comment détermine-t-on un **centre de classe** ?

   b) Quelle est la signification statistique du centre de classe ?

   c) Qu'est-ce qu'une **pondération** ?

   d) Donner quelques exemples de pondération.

3. Soit la série statistique suivante:

Prix de la place x_i	Nombre de places vendues n_i	$n_i x_i$
15	50	
19	30	
21	40	
25	100	
30	120	
35	20	
40	20	
	$\sum =$	$\sum =$

   a) Compléter le tableau ci-dessus.

   b) Calculer la moyenne arithmétique **pondérée**

4. Calculer de deux manières la **moyenne arithmétique pondérée** pour la série statistique suivante:

Classes (diamètres en mm)	x_i	**Effectifs n_i**	$n_i x_i$	f_i	$f_i x_i$
45 - 55		15			
55 - 65		35			
65 - 75		70			
75 - 85		55			
85 - 95		25			
		$\sum =$	$\sum =$	$\sum =$	$\sum =$

5. Soit la série statistique suivante:

Classes (salaire horaire en frs)	**x_i (a)**	**Effectifs n_i**	**$n_i x_i$ (a)**	**x_i (b)**	**$n_i x_i$ (b)**
< 20		2500			
20 - 30		1701			
30 - 40		1422			
40 - 50		714			
> 50		1500			
		$\sum =$	$\sum =$		$\sum =$

a) Calculer la moyenne arithmétique pondérée si l'on admet que le salaire horaire le plus bas est égal à frs. 8.- et le plus élevé à frs. 80.- .

b) Calculer la moyenne arithmétique pondérée si l'on admet que le salaire horaire le plus bas est égal à frs. 12.- et le plus élevé à frs. 160.-

c) Comparer les deux résultats; commenter.

6. Calculer de deux manières la **moyenne arithmétique pondérée** pour la série statistique suivante:

Classes (taille en cm)	**x_i**	**Effectifs n_i**	**$n_i x_i$**	**f_i**	**$f_i x_i$**
155 - 160		160			
160 - 165		188			
165 - 170		205			
170 - 175		386			
175 - 180		424			
180 - 185		298			
185 - 190		235			
190 - 195		87			
		$\sum =$	$\sum =$	$\sum =$	$\sum =$

7. Soit la série statistique suivante:

Indice i	Prix de la place x_i	Nombre de places vendues n_i	$n_i x_i$
1	15	50	
2	19	30	
3	21	40	
4	25	100	
5	30	120	
6	35	20	
7	40	20	
		$\sum =$	$\sum =$

a) Compléter le tableau ci-dessus.

b) Combien y a-t-il d'**observations au total** ? **N** =
Combien y a-t-il d'**observations distinctes** ? **p** =

c) Calculer sans chercher d'interprétation:

c1) $\sum_{i=2}^{5} n_i^2 \, (x_i + 4)$

c2) $\sum_{i=1}^{7} n_i \, x_i + 3 \sum_{i=2}^{3} n_i^3 \, x_i^2$

c3) $\frac{2n_4}{3x_2} + \frac{15n_2^2}{(2n_1)^4}$

c4) $\sum_{i=2}^{5} |x_i - 28|$

CHAPITRE 5 LES MESURES DE DISPERSION

5.1 Introduction

Une mesure de dispersion permet de décrire un ensemble de données en fournissant une indication sur la variabilité des valeurs au sein de l'ensemble des données.
La mesure de dispersion complète la description fournie par la mesure de tendance centrale d'une distribution. Deux populations ayant des valeurs centrales identiques peuvent être très dissemblables en raison de leur dispersion différente.
Nous pouvons classer les mesures de dispersion en deux groupes:

1) Mesures définies par la distance entre deux valeurs représentatives de la distribution:

   - l'**étendue de la série**
   - l'**intervalle interquantile**

2) Mesures calculées en fonction des déviations de chaque donnée par rapport à une valeur centrale:

   - l'**écart absolu moyen**
   - l'**écart-type**

5.2 L'étendue de la série

L'étendue de la série est la mesure de dispersion la plus facile à calculer. Elle est définie comme la différence entre la valeur observée la plus grande et la valeur observée la plus petite.

Exercice

Soit la série d'observations suivantes: relevé des températures durant le mois de février dans la ville Faitfrisquet:

-10 -9 -5 -12 -8 -5 -2 5 2 3 5 6 2 0 1 -1 -2 0 -6 -10 -11 -9 -6 -4
-3 0 0 1

a) Calculer l'étendue de la série

b) A l'aide d'Excel:

   - entrer les données
   - déterminer la valeur maximale ⇒ fonction :
   - déterminer la valeur minimale ⇒ fonction :
   - déterminer l'étendue ⇒ démarche:

5.3 Les quartiles. les déciles, les percentiles

La notion de quantile est une extension du concept de la médiane (qui divise une distribution en deux parties). Les quantiles les plus fréquemment utilisés sont:

- les **quartiles** qui divisent un ensemble d'observations en quatre parties
- les **déciles** qui divisent un ensemble d'observations en dix parties
- les **percentiles** qui divisent un ensemble d'observations en cent parties.

Dès lors, il convient tout d'abord de classer les valeurs du caractère étudié dans l'ordre croissant, puis de partager les effectifs en 4, 10 ou 100 pour obtenir respectivement le 1er quartile, le 1er décile ou le 1er percentile.

Exercice

Soit la série statistique suivante:

Classes	Effectifs	
0 - 500	5	
500 - 1000	17	
1000 - 1500	20	
1500 - 2500	11	
2500 - 3000	3	
3000 - 5000	2	
Total		x x x x

a) En vous inspirant du calcul de la médiane, déterminez la valeur du caractère correspondant au **1er quartile**.

b) Même question pour:

- le **2ème quartile**
- le **3ème quartile**
- le **9ème décile**
- le **38ème percentile**

5.4 Les intervalles interquantiles

L'intervalle interquantile est un intervalle qui comprend un certain pourcentage d'observations. Par exemple, l'intervalle recherché peut représenter la richesse des 25% des individus les plus riches d'un pays, celle des 50 % moyennement riches, ou celle des 10 % les moins riches, etc... Les deux intervalles les plus utilisés sont:

- l'**intervalle interquartile**
- l'**intervalle interdécile**

L'**intervalle interquartile** est égal à l'écart entre le 1er quartile et le 3ème quartile. Cet intervalle contient donc la moitié centrale des observations.
L'**intervalle interdécile** est égal à l'écart entre le 1er décile et le 9ème décile. Cet intervalle contient donc le 80 % au centre des observations.

Exercices

1. Soit la distribution statistique suivante:

Age des ouvriers d'une entreprise	Effectifs	
20 - 25	9	
25 - 30	27	
30 - 35	36	
35 - 40	45	
40 - 45	18	
45 - 50	9	
50 - 55	3	
55 - 65	3	
Total		x x x x

Déterminer:

- le **1er quartile**

- le **3ème quartile**

- l'**intervalle interquartile**

Interpréter le résultat obtenu:

2. Pour la série statistique **Contrôle de qualité** (cf exercice 2 du chap.2), calculer:

a) l'intervalle interquartile

b) le $7^{\text{ième}}$ décile

c) l'intervalle interdécile

5.5 L'écart absolu moyen

L'écart absolu moyen est la moyenne arithmétique des écarts, en valeur absolue, de chaque observation à la moyenne arithmétique de la série. A partir de cette définition, la formule

a) pour une **série simple** est:

b) pour une **série pondérée** est:

Exercices

1. Soit la série statistique simple suivante:

Production (en tonnes) par entreprise	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
Entreprise A: 41		
Entreprise B: 53		
Entreprise C: 59		
Entreprise D: 64		
Entreprise E: 71		
Entreprise F: 75		
Entreprise G: 78		
Total	x x x x	

Déterminer $\bar{x}$: et e_a :

Interpréter le résultat:

2. Utiliser la fonction *ECART.MOYEN* d'Excel pour vérifier le résultat de l'exercice 1.

3. Soit la série statistique pondérée suivante:

Classes	x_i	Effectifs	$n_i x_i$	$\lvert x_i - \bar{x} \rvert$	$n_i \lvert x_i - \bar{x} \rvert$
1450 - 1500		1			
1500 - 1550		8			
1550 - 1600		26			
1600 - 1650		90			
1650 - 1700		96			
1700 - 1750		58			
1750 - 1800		10			
1800 - 1850		6			
1850 - 1900		3			
1900 - 1950		2			
Total	x x x x			x x x x	

Déterminer l'**écart absolu moyen** et interpréter le résultat:

Ecart moyen relatif

Parallèlement à l'écart absolu moyen, on peut calculer une mesure de dispersion relative: l'**écart moyen relatif**, obtenu en divisant l'écart absolu moyen par la moyenne arithmétique de la série.

Exercices

4. Soit les deux séries statistiques suivantes:

Année	Véhicules vendus (en milliers)	$\|x_i - \bar{x}\|$
2007	100	
2008	115	
2009	80	
2010	85	
2011	106	
2012	120	
	$\sum =$	$\sum =$
	$\bar{x} =$	Ecart abs.moyen =

Année	Carburant vendu (en milliers de tonnes)	$\|x_i - \bar{x}\|$
2007	420	
2008	450	
2009	410	
2010	410	
2011	446	
2012	450	
	$\sum =$	$\sum =$
	$\bar{x} =$	Ecart abs.moyen =

Ecart moyen relatif:

Ecart moyen relatif:

Interprétation:

5.6 L'écart-type

De tous les critères de dispersion, l'**écart-type** est le plus utilisé. Il est défini comme étant égal à la racine carrée de la moyenne arithmétique des carrés des écarts de toutes les observations d'une série pris par rapport à la moyenne arithmétique.
L'écart-type, noté σ, est une mesure de dispersion qui possède la même unité que la série étudiée.
Nous avons les formules suivantes:

1) Pour une série **simple**:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}}$$

ou d'une manière équivalente: $\sigma=\sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}-\left(\bar{x}\right)^2}$

2) Pour une série **pondérée**:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{p} n_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^{p} n_i}}$$

ou d'une manière équivalente: $\sigma=\sqrt{\frac{\sum_{i=1}^{p} n_i \cdot x_i^2}{\sum_{i=1}^{p} n_i}-\left(\bar{x}\right)^2}$

Exercices

1. Soit la série statistique simple suivante:

Production (en tonnes) par entreprise	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2$
Entreprise A: 41		
Entreprise B: 53		
Entreprise C: 59		
Entreprise D: 64		
Entreprise E: 71		
Entreprise F: 75		
Entreprise G: 78		
Total	x x x x	

$\bar{x}$ = σ =

2. Soit la même série statistique simple :

Production (en tonnes) par entreprise	x_i^2
Entreprise A: 41	
Entreprise B: 53	
Entreprise C: 59	
Entreprise D: 64	
Entreprise E: 71	
Entreprise F: 75	
Entreprise G: 78	
Total	

$\bar{x}$ = $\left(\bar{x}\right)^2$ = σ =

Constatations:

3. Utiliser la fonction *STDEVPA* d'Excel pour vérifier le résultat de l'exercice 1.

4. Soit la série statistique pondérée suivante:

Age des ouvriers	Centre de classe x_i	Effectifs n_i	$n_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
20 - 25		9				
25 - 30		27				
30 - 35		36				
35 - 40		45				
40 - 45		18				
45 - 50		9				
50 - 55		3				
55 - 60		3				
Total	× × ×			× × × ×	× × × ×	

$\bar{x} =$ $\sigma =$

5. Soit la même série statistique pondérée:

Age des ouvriers	Centre de classe x_i	Effectifs n_i	x_i^2	$n_i x_i^2$
20 - 25		9		
25 - 30		27		
30 - 35		36		
35 - 40		45		
40 - 45		18		
45 - 50		9		
50 - 55		3		
55 - 60		3		
Total	× × × ×		× × × ×	

$\bar{x} =$ $(\bar{x})^2 =$ $\sigma =$

Constatations:

Le coefficient de variation

Le coefficient de variation permet de faire des comparaisons du degré de dispersion entre différentes séries statistiques. Il est obtenu en relativisant l'écart-type par rapport à la **moyenne arithmétique**.
Nous avons alors la formule suivante :

$$C = \frac{\sigma}{\bar{x}}$$

Exercice

Au cours du mois de février, sur le marché boursier de l'industrie électronique, la moyenne des prix quotidiens à la fermeture a été de Fr. 1'500.- avec un écart-type de Fr. 50.- pour la catégorie A d'actions, tandis que pour la catégorie B, durant la même période, la moyenne a été de Fr. 500.- et l'écart-type de Fr. 30.- .

Si on faisait une comparaison dans l'absolu, on pourrait déclarer que le prix de la catégorie A présente une variabilité plus grande que la catégorie B (50 comparé à 30) ; mais il est plus pertinent de faire une évaluation relative de la dispersion avant de passer à la comparaison.
On trouvera alors les coefficients de variation respectifs suivants :

- pour le titre A :

- pour le titre B :

- conclusion :

CHAPITRE 6 LES INDICATEURS DE CONCENTRATION

6.1 La médiale

La médiale est la **valeur** du caractère qui partage en deux la **masse globale** du caractère. Cette masse globale correspond à la somme des produits $n_i x_i$.
Le calcul de la **médiale** s'inspire de celui de la médiane. La comparaison des deux valeurs constitue une mesure de la **concentration**. Lorsque l'écart entre la médiane et la médiale est important par rapport à l'**étendue** de la série, la **concentration est forte**. Cela signifie qu'il y a une forte **inégalité** dans la répartition de la série.

Exemple

Soit la distribution statistique suivante :

Classes (salaires horaires en fr.)	Centres de classe	Nombre de salariés	Cumul du nombre de salariés	Masses salariales	Masses salariales cumulées
	x_i	n_i		$n_i x_i$	
18 - 22		4			
22 - 26		14			
26 - 30		26			
30 - 34		32			
34 - 38		14			
38 - 42		10			
		$\sum =$		$\sum =$	

Marche à suivre pour calculer la **médiale** :

1. partager la masse salariale en deux parties égales :

2. chercher la classe médiale :

3. en s'inspirant de la formule de la médiane, déterminer la médiale :

Calculer la **médiane** :

Déterminer l'**écart** entre la médiale et la médiane :

Interprétation de l'écart entre la médiale et la médiane

Nous allons, à l'aide de deux cas théoriques, interpréter l'écart entre la médiale et la médiane.

CAS 1 La médiale et la médiane ont la même valeur. L'écart est alors **nul**. Cela signifie que la moitié de l'effectif se répartit la moitié de la masse salariale. Il y a donc **égalité parfaite** de la distribution.

CAS 2 La médiale vaut 60.- Fr. et la médiane 25.- Fr. L'écart s'élève à 35.- Fr. Nous avons ici une répartition **inégalitaire**. Le 50% de l'effectif ne se partage pas le 50% de la masse salariale.

Indicateur de concentration relatif

Afin de pouvoir effectuer des comparaisons de concentration entre différentes séries statistiques, on **relativise** l'écart entre la médiale et la médiane en le divisant par l'**étendue** de la série. On obtient ainsi une mesure en pourcentage.

Exercice

Comparer le degré de concentration des salaires dans deux pays caractérisés par :

Pays A :	salaire horaire minimum	Fr. 13.-
	salaire horaire maximum	Fr. 115.-
	médiane	Fr. 28.-
	médiale	Fr. 45.-
Pays B :	salaire horaire minimum	$ 5
	salaire horaire maximum	$ 100
	médiane	$ 15
	médiale	$ 42

6.2 La courbe de Lorenz ou courbe de concentration

Dans le premier quadrant d'un graphe, on gradue de manière similaire l'abscisse et l'ordonnée. L'abscisse correspond au **pourcentage cumulé des effectifs** et l'ordonnée au **pourcentage cumulé de la masse globale** du caractère. Avant de reporter les points dans le graphe, on dessinera le carré ainsi que la **diagonale** partant de l'origine. La courbe de Lorenz s'obtient en reliant ces points.
Plus la courbe de Lorenz est proche de la diagonale du carré, plus la concentration est faible. Plus elle s'en éloigne, plus la concentration est forte.

Exercice

A partir des données ci-dessous, établir sur le même graphique une courbe de Lorenz pour les années 1955/56 et une autre courbe pour les années 1989/90. Quelles conclusions peut-on en tirer ?

Répartition des revenus imposables selon l'impôt fédéral direct. Source : J. DEISS, D. MEUWLY, Manuel d'économie politique, tome II, Ed. Fragnières, Fribourg 1994, p.93.

Années	0	20	40	60	80	100
1955/56	0	1,9	9,0	24,0	47,9	100
1965/66	0	1,9	8,8	24,6	47,6	100
1973/74	0	3,4	15,0	31,6	53,8	100
1978	0	3,7	15,1	32,4	55,6	100
1989/90	0	6,9	21,4	38,0	58,4	100

Le tableau se lit de la façon suivante :
1ère ligne : pourcentages cumulés des contribuables
à l'intérieur du tableau : pourcentages cumulés du revenu imposable total
exemple : en 1955/56, 40 % des contribuables se partagent 9,0 % du revenu imposable.

6.3 L'indice de Gini

Un indice de concentration peut être calculé à partir de la courbe de Lorenz. Il faut mettre en rapport deux **surfaces**, à savoir :

→ au numérateur : la surface comprise entre la courbe et la diagonale

→ au dénominateur : la surface du triangle

Ce rapport est appelé **indice de Gini**. Pour calculer les surfaces, le graphique doit être tracé sur du papier millimétré.

En supposant résolues toutes les difficultés relatives au calcul des surfaces

a) quelles valeurs peut prendre l'indice de Gini ?

b) que signifie un indice proche de 1 ?

c) que signifie un indice proche de 0 ?

6.4 Exercices

1. La distribution de la fortune dans une petite ville suisse est donnée dans le tableau suivant :

Classes de fortune (en milliers de francs)	Nombre de contribuables	
0 – 50	500	
50 -100	330	
100 – 500	120	
500 - 1000	60	

a) Compléter le tableau ci-dessus afin de pouvoir établir une courbe de Lorenz.

b) Etablir le graphe de la courbe de Lorenz.

c) Calculer l'écart entre la médiale et la médiane.

d) Déterminer l'indicateur de concentration relatif.

e) Calculer l'indice de Gini.

f) Que dire de la répartition de la fortune dans cette petite ville suisse ?

2. Soit la distribution des revenus imposables pour la période de taxation 2011 dans le canton de Neuchâtel :

Classes de revenu (en milliers de francs)	Nombre de contribuables	
Moins de 1	12663	
1 – 4	4502	
4 – 10	5134	
10 – 20	8990	
20 – 30	11142	
30 – 40	11774	
40 – 50	11169	
50 – 60	9035	
60 – 70	7064	
70 – 80	5598	
80 - 90	4595	
90 - 100	3640	
100 - 150	8006	
150 et plus	3597	
Total		

Source : Administration fédérale des contributions

Répondre aux questions suivantes en utilisant Excel :

a) Compléter le tableau ci-dessus afin de pouvoir établir une courbe de Lorenz.

b) Etablir le graphe de la courbe de Lorenz.

c) Déterminer l'indice de Gini en calculant les différentes surfaces (triangle, trapèzes) situées sous la courbe de Lorenz.

d) Que peut-on dire sur la concentration ?

CHAPITRE 7 L'AJUSTEMENT LINEAIRE

7.1 Introduction

Lorsque l'étude statistique d'une population porte simultanément sur deux caractères quantitatifs, on peut se poser la question d'une éventuelle **liaison** entre ces deux caractères.

Par exemple : existe-t-il une relation entre le poids et la taille d'un individu ?

peut-on établir une relation entre le coût unitaire de fabrication d'un objet et le nombre d'objets fabriqués ?

Déterminer une liaison, c'est précisément effectuer un **ajustement**. On parle aussi de **régression linéaire**.

Il existe trois sortes de liaisons :

- les **liaisons totales** ou **relations fonctionnelles**
- l'**absence** de liaison
- les **liaisons statistiques**

7.2 Les liaisons totales ou relations fonctionnelles

Dans le cas d'une liaison totale, lorsque l'on connaît les valeurs prises par l'un des caractères, on peut déterminer précisément les valeurs prises par l'autre caractère.

Exemples :
- la liaison entre la longueur du côté d'un carré et le pourtour de ce carré ;
- la liaison entre la consommation d'électricité et le montant de la facture.

Dans les deux exemples ci-dessus, nous avons une **relation linéaire** qui peut être représentée par une **droite** d'équation **y = ax + b**.

La liaison entre la longueur du côté d'un carré (**x**) et le pourtour de ce carré (**y**) nous donne l'équation suivante : **y = 4x**, où **y** est la **variable dépendante** et **x** la **variable indépendante**.

Si le prix du kwh est de Fr. 0.30 et que la taxe de base vaut Fr. 25.-, la relation entre la consommation d'électricité (**x**) et la facture d'électricité (**y**) peut être donnée par l'équation suivante : **y = 0.3x + 25**, où **y** est la variable dépendante et **x** la variable indépendante.

Exercices

1. Une voiture occasionne des frais s'élevant à Fr. 30.- pour 100 km. La taxe annuelle de base représente Fr. 350.- .

   a) Etablir l'expression mathématique annuelle de la relation des frais en fonction des kilomètres parcourus.

   b) Combien de kilomètres la voiture a-t-elle parcourus en une année lorsque les frais s'élèvent à Fr. 4085.- ?

2. Une entreprise fait fabriquer un produit et paie Fr. 162.50 la pièce. Elle se propose d'acheter une machine pour fabriquer elle-même ce produit. Ses frais fixes s'élèveraient alors à Fr. 25'000.- par année, alors que les frais de production variables seraient de Fr. 100.- par pièce.

   a) L'investissement est-il une bonne solution si

      a1) la production s'élève à 350 pièces (cas actuel) ?
      a2) la production était augmentée de 33,33% ?

   b) Pour quelle production est-il égal d'acheter la machine ou de faire fabriquer les pièces ?

7.3 L'absence de liaison

Nous avons vu que dans le cas d'une liaison totale, la relation pouvait être représentée par une droite. Si on reporte sur l'axe horizontal le premier caractère et sur l'axe vertical le deuxième caractère, **nous n'obtenons plus une droite s'il y a absence de liaison** mais un nuage de points répartis au hasard.

Exemple : il n'y a pas de relation entre la note de statistique obtenue par un étudiant et sa taille !

7.4 Les liaisons statistiques

On dit qu'il existe une liaison statistique lorsque les variations de l'un des caractères expliquent **en partie** les variations de l'autre caractère.

Exemple : relation entre la taille et le poids d'un individu.

Lorsqu'on représente graphiquement les points déterminés par les deux caractères, on obtient un **nuage qui permet de suggérer une relation linéaire**. On pourra donc procéder à un **ajustement linéaire** ou effectuer une **régression linéaire**. Cela consiste à rechercher la meilleure relation possible entre les deux caractères étudiés. Graphiquement, cela revient à dessiner la droite qui donne la meilleure représentation de l'ensemble des observations. En d'autres termes, nous allons dessiner la droite passant « au milieu » de l'ensemble des points.

Il faut trouver maintenant une méthode mathématique afin de déterminer l'équation de la droite d'ajustement. Toutes les observations ayant été placées sur un graphique, on va tracer à main levée la droite qui passe « au milieu » de tous ces points.

Cette droite a pour équation **y = ax + b**. On remarque que pour chaque point **x_i**, la valeur **y_i** définie par l'équation de la droite diffère de la valeur **observée** ; notons cette erreur ε_i. Il conviendra donc de déterminer l'équation de la droite en **minimisant l'ensemble des erreurs.** Il existe plusieurs méthodes mathématiques qui minimisent ces erreurs :

- min max $|\varepsilon_i|$

- min $\sum_{i=1}^{N} |\varepsilon_i|$

- min $\sum_{i=1}^{N} \varepsilon_i^2$

La méthode la plus fréquemment utilisée est la troisième ; c'est la **méthode des moindres carrés**. Elle a pour objet de minimiser la somme des carrés des erreurs.

- Chaque observation **Y_i** étant égale à : $Y_i = ax_i + b + \varepsilon_i$, on peut en déduire que $\varepsilon_i = Y_i - ax_i - b$. Minimiser $\sum_{i=1}^{N} \varepsilon_i^2$ revient donc à minimiser $\sum_{i=1}^{N} (Y_i - ax_i - b)^2$.

  Cette minimisation fait appel à la notion de dérivée partielle. Nous allons déterminer les paramètres a et b de la droite de régression qui minimisent les erreurs. Les formules sont les suivantes :

$$\boxed{a = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}}$$

$$\boxed{b = \bar{y} - a\bar{x}}$$

où $\bar{x}$ et $\bar{y}$ sont respectivement la moyenne des x_i et la moyenne des y_i.

Exercices

1. Utiliser la méthode des moindres carrés afin de déterminer la droite qui explique au mieux la relation qui existe entre la taille et le poids des joueurs d'une équipe de football.
Voici le tableau où l'on a relevé les différentes tailles et les différents poids des 12 joueurs.

Joueurs	Taille observée (cm) : x_i	Poids observé (kg) : y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	Poids calculé : $\hat{y}_i$
Joueur 1	165	66					
Joueur 2	197	107					
Joueur 3	155	59					
Joueur 4	178	81					
Joueur 5	183	88					
Joueur 6	172	77					
Joueur 7	175	78					
Joueur 8	205	114					
Joueur 9	188	94					
Joueur 10	185	100					
Joueur 11	160	57					
Joueur 12	180	85					
	$\sum=$	$\sum=$			$\sum=$	$\sum=$	

a) Reporter les couples d'observations sur un **graphe** approprié.

b) Quel est le caractère qui correspond à la **variable dépendante** ?
Quel est le caractère qui correspond à la **variable indépendante** ?

c) Donner la signification des symboles **a, b, N**.

d) Déterminer la valeur de **a** et celle de **b** en complétant le tableau ci-dessus et en utilisant les formules données.

e) Donner l'**expression mathématique** de la droite de régression.

f) Représenter sur le **graphe** du point a) la droite de régression trouvée au point e).

g) Un joueur arrive en cours d'année. Il connaît sa taille qui est de 182 cm, mais ignore son poids. **Estimer le poids** de ce nouveau joueur.

2. Un commerçant a évalué le total de ses charges, sans tenir compte du prix de revient de la marchandise. Pour les années 2003 à 2012, il a obtenu les chiffres reportés dans le tableau ci-dessous :

Année	Chiffre d'affaires en Fr.: x_i	Charges en Fr. : y_i				
2003	80'000	39'500				
2004	85'000	41'800				
2005	95'000	45'300				
2006	90'000	43'400				
2007	100'000	47'000				
2008	108'000	49'700				
2009	112'000	51'300				
2010	109'000	50'100				
2011	105'000	48'800				
2012	102'000	57'700				

a) Quelle est l'expression mathématique de la droite d'ajustement ?

b) Quel est le montant des frais fixes ?

c) Quel est le point mort, sachant que le prix de revient est égal à 45% du prix de vente ?

7.5 Le coefficient de corrélation

Le coefficient de corrélation est une mesure de **l'intensité de la relation linéaire** entre deux variables étudiées. Les valeurs possibles du coefficient se situent dans **l'intervalle [-1 ;1]**. Ces deux valeurs extrêmes représentent une relation linéaire parfaite entre les variables, « négative » dans le premier cas, « positive » dans l'autre. La valeur de zéro signifie l'absence de relation linéaire.
On note le coefficient de corrélation par **r** et il est défini de la façon suivante :

$$\boxed{r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}}}$$

On peut également mesurer le degré de corrélation entre deux variables par le **coefficient de détermination**. Il mesure la proportion de la variation dans la variable y qui est expliquée par la variation de la variable x. Il correspond au carré du coefficient de corrélation et peut donc varier entre 0 et 1.

Exercices

1. Déterminer le coefficient de corrélation et le coefficient de détermination pour l'exercice 1 du paragraphe 7.4 en utilisant le tableau ci-dessous et donner leur interprétation:

Joueurs	Taille observée (cm) : x_i	Poids observé (kg) : y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
Joueur 1	165	66					
Joueur 2	197	107					
Joueur 3	155	59					
Joueur 4	178	81					
Joueur 5	183	88					
Joueur 6	172	77					
Joueur 7	175	78					
Joueur 8	205	114					
Joueur 9	188	94					
Joueur 10	185	100					
Joueur 11	160	57					
Joueur 12	180	85					
	$\sum=$	$\sum=$			$\sum=$	$\sum=$	$\sum=$

2. Idem pour l'exercice 2 :

Année	Chiffre d'affaires en Fr. : x_i	Charges en Fr. : y_i					
2003	80'000	39'500					
2004	85'000	41'800					
2005	95'000	45'300					
2006	90'000	43'400					
2007	100'000	47'000					
2008	108'000	49'700					
2009	112'000	51'300					
2010	109'000	50'100					
2011	105'000	48'800					
2012	102'000	57'700					

7.6 L'ajustement à l'aide du solveur d'*EXCEL*

Le solveur va trouver la valeur des paramètres **a** et **b** pour lesquels la somme des écarts élevés au carré entre la valeur observée et la valeur calculée est **minimale**.

Marche à suivre

1) Dans une feuille de calcul, entrer les valeurs observées de la variable indépendante dans une colonne (par exemple colonne B), puis les valeurs observées de la variable dépendante dans une colonne adjacente (colonne C).

2) Dans une cellule (par exemple F4), poser une valeur initiale pour **a** (par exemple 1) et dans la cellule suivante (F5) une valeur initiale pour **b** (par exemple 1).

3) Dans la colonne D, dans la cellule en regard de la première observation (en D3), entrer la formule **=F4 * B3 + F5** puisqu'il s'agit d'une équation de la forme **y = ax + b**.
Faire ensuite une copie incrémentée en regard des autres observations.

4) En F7, par exemple, entrer la fonction **SOMME.XMY2(matrice_x ;matrice_y)** par insertion / Fonction / Math & Trigo /

   matrice_x : valeurs de Y observées ! ! ! ! (colonne C)
   matrice_y : valeurs de Y calculées (colonne D)

5) Appeler **Données / Solveur**

   Cellule cible : **F7 à minimiser**
   Cellules variables : **F4 ;F5**

Tester cette démarche avec la série de l'exercice 1 (tailles et poids).

La somme minimale des écarts devra être égale à **120.29795**.
Le paramètre **a** vaut **1.2172** et le paramètre **b** vaut **-133.5297**.

7.7 Représentation graphique de la droite d'ajustement à l'aide d'*EXCEL*

Marche à suivre :

a) Etablir le graphique des couples de points observés

1) Sélectionner - avec la souris - dans la feuille de calcul la colonne du premier caractère observé (variable indépendante **x**) et celle du second caractère observé (variable dépendante **y**).

2) Sélectionner Insertion / Graphique

3) Compléter la boîte de dialogue de l'assistant graphique :

- sélectionner **nuage de points** ;
- sélectionner le format graphique No 1 ;

b) Etablir l'ajustement linéaire

1) Cliquer avec la souris sur un des points du graphique pour **sélectionner les points représentés**.

2) Sélectionner avec la souris de droite : ajouter une courbe de tendance... pour obtenir la boîte de dialogue :

- sélectionner le type de régression / de courbe de tendance désiré ;
- compléter selon vos désirs, notamment *Prévision* pour prolonger l'ajustement ;
  sélectionner ou non : □ afficher l'équation sur le graphique
  □ afficher le coefficient de détermination

Tester cette démarche avec la série de l'exercice 1 (tailles et poids).

7.8 Les fonctions Pente et Ordonnée.origine d'*EXCEL*

Si l'on souhaite uniquement connaître la droite d'ajustement qui passe le mieux entre toutes les observations (au sens des moindres carrés), on dispose sur *EXCEL* de deux fonctions qui donne d'une part la **pente** de la droite et d'autre part l'**ordonnée à l'origine** de la droite.

7.8.1 La fonction Pente

Marche à suivre :

1) Dans une feuille de calcul, entrer les valeurs observées de la variable indépendante dans une colonne, puis les valeurs observées de la variable dépendante dans une colonne adjacente.

2) Insertion / Fonction / Statistiques / **Pente**.

3) Compléter la boîte de dialogue :

- y_connus ; sélectionner la plage correspondante (soit la variable dépendante observée) dans la feuille de calcul à l'aide de la souris ;
- x _connus ; sélectionner la plage correspondante (soit la variable indépendante observée) dans la feuille de calcul à l'aide de la souris ;

Tester cette démarche avec la série de l'exercice 1 (tailles et poids).

7.8.2 La fonction Ordonnée.origine

Marche à suivre :

1) Dans une feuille de calcul, entrer les valeurs observées de la variable indépendante dans une colonne, puis les valeurs observées de la variable dépendante dans une colonne adjacente.

2) Insertion / Fonction / Statistiques / **Ordonnée.origine**.

3) Compléter la boîte de dialogue :

   - y_connus ; sélectionner la plage correspondante (soit la variable dépendante observée) dans la feuille de calcul à l'aide de la souris ;
   - x _connus ; sélectionner la plage correspondante (soit la variable indépendante observée) dans la feuille de calcul à l'aide de la souris ;

Tester cette démarche avec la série de l'exercice 1 (tailles et poids).

7.9 La fonction Coefficient.corrélation d'*EXCEL*

Il existe sur *EXCEL* une fonction statistique qui donne le coefficient de corrélation entre deux séries de données. On peut ainsi déterminer le coefficient de corrélation d'un ajustement linéaire.

Marche à suivre :

1) Dans une feuille de calcul, entrer les valeurs observées de la variable indépendante dans une colonne, puis les valeurs observées de la variable dépendante dans une colonne adjacente.

2) Insertion / Fonction / Statistiques / **Coefficient.corrélation**.

3) Compléter la boîte de dialogue :

   - matrice1 ; sélectionner la plage correspondante (soit la variable dépendante observée) dans la feuille de calcul à l'aide de la souris ;
   - matrice2 ; sélectionner la plage correspondante (soit la variable indépendante observée) dans la feuille de calcul à l'aide de la souris ;

Tester cette démarche avec la série de l'exercice 1 (tailles et poids).

7.10 La fonction Coefficient.détermination d'*EXCEL*

Il existe sur *EXCEL* une fonction qui donne le coefficient de détermination d'un ajustement linéaire.

Marche à suivre :

1) Dans une feuille de calcul, entrer les valeurs observées de la variable indépendante dans une colonne, puis les valeurs observées de la variable dépendante dans une colonne adjacente.

2) Insertion / Fonction / Statistiques / **Coefficient.détermination**.

3) Compléter la boîte de dialogue :

- y_connus ; sélectionner la plage correspondante (soit la variable dépendante observée) dans la feuille de calcul à l'aide de la souris ;
- x _connus ; sélectionner la plage correspondante (soit la variable indépendante observée) dans la feuille de calcul à l'aide de la souris ;

Tester cette démarche avec la série de l'exercice 1 (tailles et poids).

7.11 Estimation des valeurs à l'aide d'*EXCEL*

Nous disposons avec *EXCEL* de fonctions permettant d'estimer des valeurs pour **y** à partir des observations. Il est ainsi possible de faire des **prévisions**.

On désire estimer le poids des individus dont la taille se situe **entre 150 cm et 210 cm ; le pas d'incrémentation sera égal à 4 cm**.

Marche à suivre :

1) Dans une feuille de calcul, entrer les valeurs observées de la variable indépendante dans une colonne, puis les valeurs observées de la variable dépendante dans une colonne adjacente.

2) Introduire dans une colonne la liste des valeurs de **x** pour lesquelles vous désirez obtenir une valeur calculée **y**.

3) En regard des **x** introduits (colonne adjacente à droite), sélectionner la plage des résultats ; il s'agira des cellules dans lesquelles vous désirez obtenir les **y** calculés.

4) Insertion / Fonction / Statistiques / **Tendance**.

5) Compléter la boîte de dialogue :

- y_connus ; sélectionner la plage correspondante (soit la variable dépendante observée) dans la feuille de calcul à l'aide de la souris ;

- x _connus ; sélectionner la plage correspondante (soit la variable indépendante observée) dans la feuille de calcul à l'aide de la souris ;
- x_nouveaux ; sélectionner la plage des **x** pour lesquels vous désirez une valeur calculée ; deux contraintes doivent être respectées :
    a) les cellules de x_nouveaux doivent contenir des valeurs
    b) le nombre de cellules de x_nouveaux doit être égal au nombre des cellules sélectionnées avant l'appel de la fonction **Tendance**.
- constante : taper **VRAI**.

6) Pour valider la fonction, taper simultanément, après avoir positionné le curseur sur la barre de formule à la fin de la fonction, les trois touches **SHIFT(↑), Ctrl et Entrée (return)** puisqu'il s'agit d'une fonction matricielle.

Tester cette démarche avec la série de l'exercice des tailles et des poids.

7.12 La fonction matricielle Droitereg d'*EXCEL*

Il existe sur *EXCEL* une fonction matricielle qui permet d'obtenir des indicateurs statistiques sur les ajustements linéaires. Il s'agit de 10 nombres disposés sous forme d'une matrice (5 lignes, 2 colonnes) de la façon suivante :

1	2
3	4
5	6
7	8
9	10

Voici la signification de ces 10 nombres :

1. **a** : pente de la droite **y = ax +b**
2. **b** : ordonnée à l'origine de la droite **y = ax + b**
3. **se(a)** : erreur-type pour la constante **a**
4. **se(b)** : erreur-type pour la constante **b**
5. $\mathbf{r^2}$: coefficient de détermination
6. **se(y calc)** : erreur-type des valeurs **y** calculées à partir de l'ajustement
7. **rapport F** : calcul de la valeur de Fisher
8. **ddl** : nombre de degrés de liberté
9. **ss(rég)** : variance expliquée : elle correspond à la somme des carrés des écarts entre les estimations de **y** et la moyenne arithmétique des valeurs **y** observées
10. **ss(rés)** : variance inexpliquée : elle correspond à la somme des carrés des écarts entre les valeurs **y** observées et les valeurs **y** calculées correspondantes.

Marche à suivre :

1) Sélectionner une plage vierge de 10 cellules (5 lignes, 2 colonnes) dans lesquelles aboutiront les résultats.

2) Insertion / Fonction / Statistiques / **Droitereg**.

3) Compléter la boîte de dialogue :

- y_connus ; sélectionner la plage correspondante (soit la variable dépendante observée) dans la feuille de calcul à l'aide de la souris ;
- x _connus ; sélectionner la plage correspondante (soit la variable indépendante observée) dans la feuille de calcul à l'aide de la souris ;
- constante : taper **VRAI** ;
- statistiques : taper **VRAI**.

4) Pour valider la fonction, taper simultanément, après avoir positionné le curseur sur la barre de formule à la fin de la fonction, les trois touches **SHIFT(↑), Ctrl et Entrée (return)** puisqu'il s'agit d'une fonction matricielle.

Tester cette démarche avec la série de l'exercice 1 (tailles et poids).

CHAPITRE 8 TAUX DE CROISSANCE

8.1 Introduction

Le taux de croissance d'une grandeur quelconque est déterminé en comparant les valeurs que prend ce caractère à deux époques. On peut donc définir un taux de croissance annuel, un taux de croissance mensuel, un taux de croissance trimestriel, etc…

8.2 Aspects mathématiques

Si l'on veut calculer un taux de croissance annuel, on va comparer les valeurs de deux années successives.
Soit V_0 la valeur à l'année 0 et V_1 la valeur l'année suivante, on prend la différence entre V_1 et V_0 et on rapporte cette différence à l'année 0 :

$$\frac{(V_1 - V_0)}{V_0}$$

Si on veut exprimer ce taux en pourcentage, il suffit de multiplier par 100 :

$$\frac{(V_1 - V_0)}{V_0} \times 100$$

La même démarche peut être effectuée pour des taux de croissance sur une autre période de temps.

8.3 Exemple

Le nombre de pièces fabriquées dans une usine a été de 12055 en 2010 et de 12784 en 2011.
Le taux de croissance annuel est donc :
(12784 – 12055) / 12055 x 100 = 6 %.

8.4 Taux de croissance négatif

Il peut arriver que l'on obtienne un taux de croissance négatif ; cela se produit lorsque la valeur V_1 est plus petite que la valeur V_0 . La différence est donc négative et le taux de croissance devient négatif. C'est en fait un taux de décroissance.
Par exemple, si nous reprenons l'exemple ci-dessus et changeons la valeur de V_1 en 11895 pièces, nous avons le nouveau taux annuel suivant :
(11895 – 12055) / 12055 x 100 = -1.3 %, c'est-à-dire une décroissance de 1.3 %.

8.5 Taux de croissance sur plusieurs années

Soit une valeur V qui croît à un taux annuel t. A l'année 0, on aura V_0 ; à l'année 1, on aura $V_1 = V_0 (1 + t)$; à l'année 2, on aura $V_2 = V_1 (1 + t)$; si on remplace V_1 dans cette dernière expression, on obtient $V_2 = V_0 (1 + t)^2$ et ainsi de suite.
Après n années, on obtient $V_n = V_0 (1 + t)^n$.
Si l'on veut connaître le taux de croissance moyen annuel, il faut isoler t :

$$t = \sqrt[n]{\frac{V_n}{V_0}} - 1.$$

8.6 Exemple

Dans ce tableau, on trouve la production de pièces pour les années 2007 à 2011 :

Années	Nombre de pièces
2007	9440
2008	8910
2009	10589
2010	12055
2011	12784

Le taux de croissance annuel moyen est égal à :

$t = \sqrt[4]{\frac{12784}{9440}} - 1 = 1.079 - 1 = 0.079$ ou 7.9%.

8.7 Exercices

1. La population résidant en Suisse de façon permanente s'élevait à fin 2009 à 7'785'806. Elle était de 7'870'134 à fin 2010. Quel est le taux de croissance annuel ?
2. Dans le canton de Neuchâtel, le nombre de chômeurs a passé de 4045 pour le mois d'avril 2012 à 3957 pour le mois de mai 2012. Quel est le taux de croissance mensuel ?
3. Dans le tableau suivant, vous trouvez le nombre de véhicules immatriculés dans le canton de Neuchâtel pour les années 2000 à 2010. Quel est le taux de croissance annuel moyen sur la période ?

Année	2000	2004	2006	2008	2010
Véhicules	84'563	87'796	86'894	88'117	89'980

4. Une valeur augmente de 18% la première année et baisse de 18% la deuxième année.
   a) De combien a-t-elle globalement varié au cours de ces deux années ?
   b) Si elle augmente de 18% la première année, de combien doit-elle varier la deuxième année pour retrouver son niveau initial ?

Réponses exercices chapitre 4

Ex.1 p.17: a) 7 b) 4 et 8
2 p.17 : 2
3 p.18: a) diagramme en bâtons b) histogramme
5 p.18 : 36.25
6 p.18: a) 67.5 b) 19.99
7 p.19 : 43.96
1 p.19: a) 4 b) 49-50 c) 73
3 p.20: 1
4 p.20 : 31.21
5 p.21: a) 69.23 b) 20.00
6 p.21 : 35.33
7 p.21 : 45.83
1 p.22 : 5.3
3 p.22 : 25.68
4 p.22 : 72
5 p.23 : a) 32.78 b) 41.08
6 p.23 : 175.05
7 p.23-24: b) N=380, p= 7 c1) 840'300 c2) 113'922'760 c3) 3.5089 c4) 21

Réponses exercices chapitre 5

Exercice p.25: 18
Exercice p.26: a) 779.41 b) 1175 1636.36 2427.27 1001
Ex.1 p.27: $Q_1 = 30.21$ $Q_3 = 39.5$ $I_Q = 9.29$
2 p.27: a) $Q_1 = 19.93$ $Q_3 = 20.08$ $I_Q = 0.15$ b) 20.06 c) $D_1 = 19.86$ $D_9 = 20.14$ $I_D = 0.28$
Ex.1 p.28: $\bar{x} = 63$ ea = 10.286 EMR = 16.3% → dispersion moyenne
3 p.28: $\bar{x} = 1666.33$ ea = 49.445 EMR = 3% → dispersion faible
4 p.29: Véhicules: $\bar{x} = 101$ ea = 12.67 emr = 12.54%
Carburant: $\bar{x} = 431$ ea = 17.67 emr = 4.10%
Ex.1 p.30 : $\bar{x} = 63$ $\sigma = 12.154$
4 p.31: $\bar{x} = 35.5$ $\sigma = 7.42$
Exercice p.32: Titre A: C = 3.33% Titre B: C = 6%

Réponses exercices chapitre 6

Exercice p.34 : Pays A : 16.67% Pays B : 28.42%
Ex.1 p.36 : c) 343.056 – 50.758 = 292.298 d) 29.23% e) 0.60 f) très mauvaise
2 p.37 : c) 0.82

Réponses exercices chapitre 7

Ex.1 p.39 : a) y = 0.3x + 350 b) 12450
Ex.2 p.39 : a1) 60'000>56'875 investissement pas rentable
a2) 71'700<75'887.50 investissement rentable b) 400
Ex.1 p.41 : b) variable dépendante: poids

variable indépendante: taille
c) a: pente ; b: ordonnée à l'origine ; N: nb d'observations
d) a = 1.22 b = -134.04
e) y = 1.22x - 134.04
g) 88 kg

Ex.2 p. 42 : a) y = 0.39 x + 9006 b) 9006 francs c) x = 0.39x + 0.45x +9006
donc x = 56287.50 francs

Ex.1 p.43 : r =0.98 corrélation forte positive
r2 = 0.96 96 % de la variation du poids est expliqué par la variation de la taille.

Ex.2 p.43 : r =0.8 corrélation moyenne positive
r2 = 0.64 64 % de la variation des charges est expliqué par la variation du chiffre d'affaires.

Réponses exercices chapitre 8

Ex.1 p.51 : 1.08%
Ex.2 p.51 : -2.18%
Ex.3 p.51 : 0.62%
Ex.4 p.51 : a) -3.24% b) -15.25%