

Statistique descriptive

<http://math.unice.fr/~dehon/Ens/L1aes>

<http://math.unice.fr/~dehon/Ens/L1aes>

Cours : 10 séances de 1H30

TD : 5 séances de 1H30 en alternance avec les TD d'analyse
(1 semaine sur 2)

Plan du cours

[page web du cours L1aes TQA stat](#)

Bibliographie et glossaire

Bibliographie : 0 élément(s)

Az Glossaire : 0 élément(s)

Annonces

Aucune annonce n'est accessible dans ce cours.

★ Actualités

21 janvier 2013 [page web du cours L1aes TQA Stat](#) Est disponible.

[→ Consulter toutes les actualités](#)

 Légende
 Lien web

Auteur : Francois-Xavier Dehon

L1 AES ([ISEM](#)) - semestre 2 - Cours & TD Statistiques descriptives (½TQA) - 2012-13,

Calendrier : [Calendrier universitaire](#), [emploi du temps](#)

10 séances de cours de 1.5H (le mardi entre 9h et 12h, [SJA 1](#), amphi 2 et 3), 5 séances de TD de 1.5 H. Premier cours le 22 janvier 2013, premier TD la semaine du 4 février.

Rq : cours d'analyse par Jérôme Vétois le mardi 9h-12h amphi 2 et 3.

TD par Jérôme Vétois et moi-même.

Contrôle continu : une interrogation d'analyse (1h) pendant le cours du 5 mars, une interro. de stat (1h) pendant le cours du 12 mars

¿Tutorat pour les deux cours de TQA (réponses aux questions sur le cours et les td) ?

[Présentation du cours](#)

Progression du cours :

1. (22 jan) Présentation du cours.

Quelques objectifs de la Statistique descriptive (résumer les données brutes, distinguer "ce qu'il y a à voir", comparer des populations, modéliser la population).

Rappel des notions vues au 1er semestre (population étudiée par un caractère) : synthèse des données brutes, résumé en un ou quelques mots.

Notions nouvelles pour ce semestre : population étudiée à travers plusieurs caractères, étude de l'indépendance ou des liaisons entre les caractères.

Document projeté lors du 1er cours : [page web du cours en 2011-12](#), etc.

un petit échantillon ? Utilisation d'un échantillon pour l'estimation d'une fréquence (institut de sondage) ou d'une mesure (coopératives agricoles).
[Document projeté.](#)

10. Réponses aux questions.

Documents de cours :

[Documents du 1er cours](#) (25jan11, format pour impression).

[Documents du 2ème cours](#) (30jan11).

[Feuille de TD no1](#) (3fev11). Corrigé de la question 3b [ici](#).

[Documents du 3ème cours](#)

[Feuille de TD no2](#) (14mar11). Corrigé des exercices 1 et 2 [ici](#), corrigé de l'exercice 3 [là](#) (question 2 dans le corrigé, les commentaires se rapportent au [sujet B](#) de l'examen de 2009-10)

[Feuille de TD no3](#) (11avr11)

[Interrogation de mars 2011](#), corrigé de l'exercice [là](#) (question 1 dans le corrigé)

[Examen 1ère session](#), [corrigé de la partie statistique avec barème de notation](#).

[Examen 2ème session](#), [corrigé de la partie statistique avec barème](#).

[Archives du cours en 2009-10](#)

Lectures :

[1] A. Hamon & N. Jégou, *Statistique descriptive*, Presse Univ. Rennes 2008. Disponible à la [BU St Jean d'Angely](#).

Statistique descriptive en L1AES - semestre 2

TQA, suite du cours du premier semestre

Objectifs du cours :

1. Savoir interpréter les objets mathématiques (tableaux, graphiques ou plus simplement nombres) des rapports publics des observatoires de la vie économique et sociale tels que l'[INSEE](#) (voir des [exemples plus bas](#)).

Pour un tel objet avec sa légende, on répondra aux questions telles que : quelle est la population étudiée, par quels caractères est elle étudiée, quelle est la nature des nombres écrits ou représentés graphiquement. Cf l'exercice 2 de la [feuille de TD2 de 2009-10](#).

2.a. Décrire conjointement plusieurs caractères d'une même population, ou un même caractère dans plusieurs sous-populations. Observer par les fréquences conditionnelles et leurs représentations graphiques l'indépendance ou au contraire la dépendance entre deux caractères ou d'un caractère en la sous-population choisie.

b. Plus abstraitement : notion d'évènements, fréquence d'un évènement, conditionnement, observation de l'indépendance ou d'une corrélation entre évènements.

c. Analyse de la dépendance entre deux caractères quantitatifs par la régression linéaire.

3. Echantillons et modèles statistiques : Echantillons, population entière ou population idéale comme modèle. Représentativité de la fréquence d'un évènement dans un échantillon suivant la taille de l'échantillon, intervalle de confiance, conséquences sur l'observation de l'indépendance entre deux évènements.

Institut national de la statistique
et des études économiques

Chercher sur le site

Ok

Google

L'Insee et la
statistique
publique

Thèmes

Bases de données

Publications et
services

Régions

Définitions et
méthodesAccès par
public

Trouver rapidement

- Indice des prix et pouvoir d'achat : tous les résultats et produits
- Revaloriser une pension alimentaire : module de calcul et conseils pratiques
- Indice de référence des loyers
- Réviser un bail professionnel (ICC, ILC, ILAT)
- Le Smic
- Indices et séries chronologiques
- Immatriculer une entreprise au Répertoire SIRENE
- Obtenir un avis de situation au répertoire SIRENE
- Métiers et concours

Recensement de la population

- La population légale de ma commune

Nom de la commune

Ok

RECENSEMENT DE LA POPULATION CHACUN DE NOUS COMPTE

Le recensement de la population commence
Votre agent recenseur viendra
vous remettre vos questionnaires

Actualités

21 janvier 2013

Tableaux de comparaisons régionales et départementales

Taux de chômage régionaux et départementaux (3e trimestre 2012) et estimations d'emploi salarié régionaux (2e trimestre 2012).

18 janvier 2013

En novembre 2012, les commandes sont stables dans l'industrie (-0,1 %)

En novembre 2012, les commandes sont quasi stables (-0,1 %) dans l'industrie hors matériels de transport (mais y compris automobile). Elles étaient en hausse en octobre (+1,4 %). Les commandes à l'exp...

17 janvier 2013

Entre 2011 et 2012, les prix ont augmenté de 2,0 % en moyenne

Ce document présente l'évolution des moyennes annuelles des prix observés par l'Insee tout au long de l'année 2012 dans le cadre de l'indice des prix à la consommation.

16 janvier 2013

Les consommations d'énergie dans l'industrie en 2011 y compris IAA et scieries

Indicateurs clés

Population

65 800 000 en 1^{er} janvier 2013

Croissance économique

+ 0,1 % au 3e trimestre de 2012

Inflation

+ 0,3 % en décembre 2012

Chômage (y c. DOM)

10,3 % au 3e trimestre de 2012

Consommation des ménages

+ 0,2 % en novembre 2012

Production industrielle

+ 0,2 % en novembre 2012

Climat des affaires

87 en décembre 2012 (moy. 100)

Dernière parution Historique

Tous les indicateurs de conjoncture

France, portrait social

Édition
2012

nuage de points, droite de Régression, nuage de points des restes.

Tirage d'échantillons dans une population, histogramme des fréquences observées.

Quelques extraits de rapports publics :

- [Pyramide des ages par pays et par années](#) (histogramme des effectifs par intervalles d'age, comparaison d'histogrammes).
- [Rapport sur les inégalités de niveaux de vie en 2008](#) (quantiles)
- [Etudiants des universités par discipline et par cursus](#) (effectifs conjoints, fréquences conditionnelles)
- [Taux d'activité selon le sexe et la configuration familiale](#) (fréquences conditionnelles)
- [Rapport 2007-2008 de l'Observatoire National de la Pauvreté et de l'Exclusion Sociale, extrait](#) (interprétation probabiliste d'une fréquence p117, relation de dépendance entre deux événements par comparaison de probabilités conditionnelles ("augmentation du risque") p118, comparaison d'histogrammes p170).
- [Rendement des haricots suivant leur précocité](#) (droite de régression).
- Héritéité chez F. Galton : droite de régression (régression vers la moyenne)

Bibliographie pour le cours

A. Hamon & N. Jégou, *Statistique descriptive*, Presse Univ. Rennes 2008.

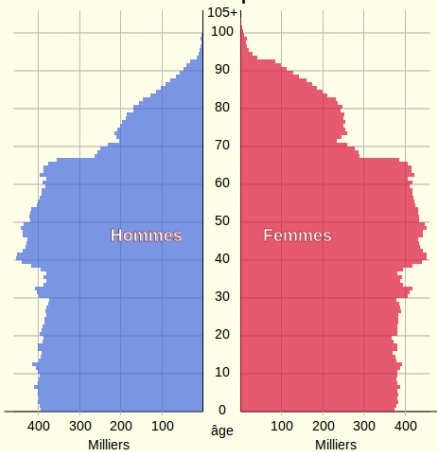
Lectures :

(livres de médecins, pages web)

[Page du cours en 2009-2010](#)



Pyramide des âges au 1er janvier 2013 France métropolitaine


 Afficher l'excédent d'hommes/de femmes

 Animation

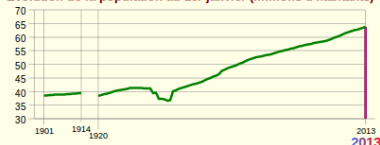
 Aide

Champ : France métropolitaine, territoire courant

Source : Insee, estimations de population (résultats arrêtés fin 2012)

Bilan démographique 2012

Évolution de la population au 1er janvier (millions d'habitants)



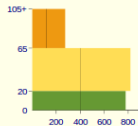
Les estimations de population sont provisoires pour 2011, 2012 et 2013. Les pyramides de 1915 à 1919 ne sont pas disponibles. Les pyramides de 1901 à 1914, 1939 à 1943 et 1945 se rapportent à 87 départements et celle de 1944 à 86 départements.

[Voir les pyramides de la France](#)

[Télécharger les données](#)

Groupes d'âges (2013)

âge	millions	%	% femmes
65+	11,28	17,7	58
20 - 64	36,84	57,8	50,7
<20	15,58	24,5	48,9
Total	63,7	100	51,5

 Modifier les groupes d'âges


Vocabulaire (connu)

Population, individus, taille, données [brutes, agrégées], caractères [qualitatif, quantitatif], valeurs ou modalités du caractère, étendue du caractère, médiane, quantiles, box-plot ou boîte à moustaches, moyenne, variance - écart type, intervalles, classes, effectif, fréquence, histogramme.

Objectifs de la statistique descriptive

1. résumé plus ou moins détaillé des caractéristiques d'une population. Qu'y a-t-il à voir ?

Objectifs de la statistique descriptive

1. résumé plus ou moins détaillé des caractéristiques d'une population. Qu'y a-t-il à voir ?
 - agrégation des données brutes par classes

Objectifs de la statistique descriptive

1. résumé plus ou moins détaillé des caractéristiques d'une population. Qu'y a-t-il à voir ?
 - agrégation des données brutes par classes
 - valeur centrale et dispersion

Objectifs de la statistique descriptive

1. résumé plus ou moins détaillé des caractéristiques d'une population. Qu'y a-t-il à voir ?

- agrégation des données brutes par classes
- valeur centrale et dispersion
- histogramme : on en retient sa forme

Objectifs de la statistique descriptive

1. résumé plus ou moins détaillé des caractéristiques d'une population. Qu'y a-t-il à voir ?

- agrégation des données brutes par classes
- valeur centrale et dispersion
- histogramme : on en retient sa forme

2. Comparaison de deux populations par comparaison des résumés. Que peut-on comparer ?

Objectifs de la statistique descriptive

- 1. résumé plus ou moins détaillé des caractéristiques d'une population. Qu'y a t-il à voir ?**
 - agrégation des données brutes par classes
 - valeur centrale et dispersion
 - histogramme : on en retient sa forme
- 2. Comparaison de deux populations par comparaison des résumés. Que peut on comparer ?**
- 3. Plusieurs caractères → liaison ou indépendance entre les caractères**

Objectifs de la statistique inférentielle

1. Approximer le résumé d'une population à partir du résumé d'un échantillon.

Objectifs de la statistique inférentielle

1. Approximer le résumé d'une population à partir du résumé d'un échantillon.
 - Dépendance du résumé en le choix de l'échantillon

Objectifs de la statistique inférentielle

1. Approximer le résumé d'une population à partir du résumé d'un échantillon.
 - Dépendance du résumé en le choix de l'échantillon
 - Intervalle de confiance pour l'estimation d'une moyenne, d'une fréquence, etc.
 - Analyse critique des objectifs 1 et 2 de la statistique descriptive

Objectifs de la statistique inférentielle

1. Approximer le résumé d'une population à partir du résumé d'un échantillon.

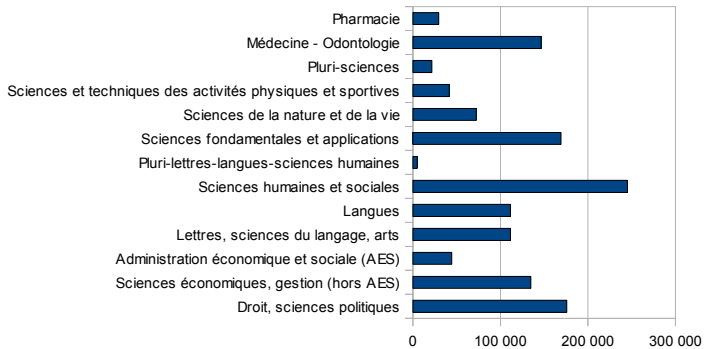
- Dépendance du résumé en le choix de l'échantillon
- Intervalle de confiance pour l'estimation d'une moyenne, d'une fréquence, etc.
- Analyse critique des objectifs 1 et 2 de la statistique descriptive

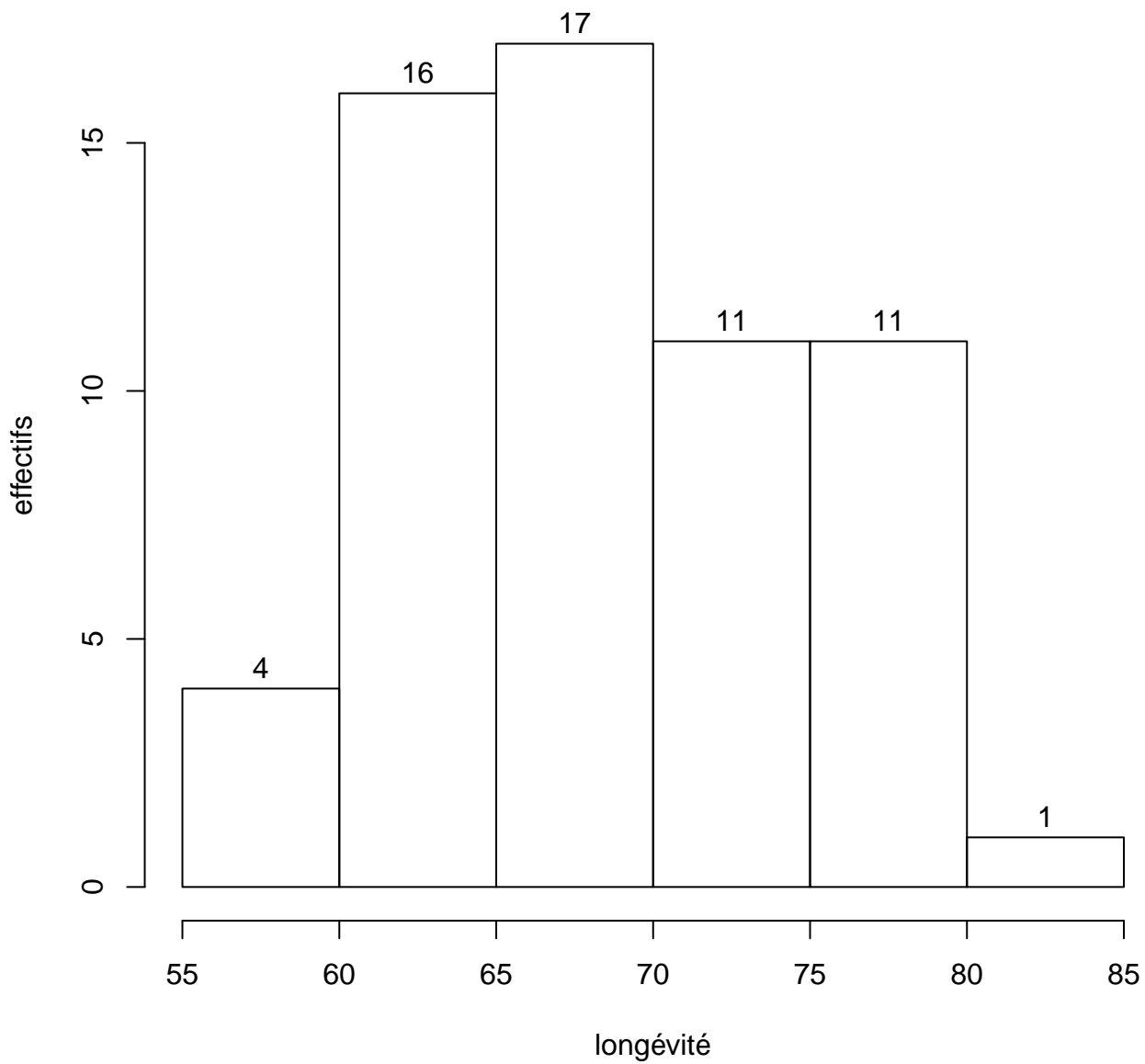
2. Modéliser une population (population étudiée considérée comme un échantillon obéissant au modèle)

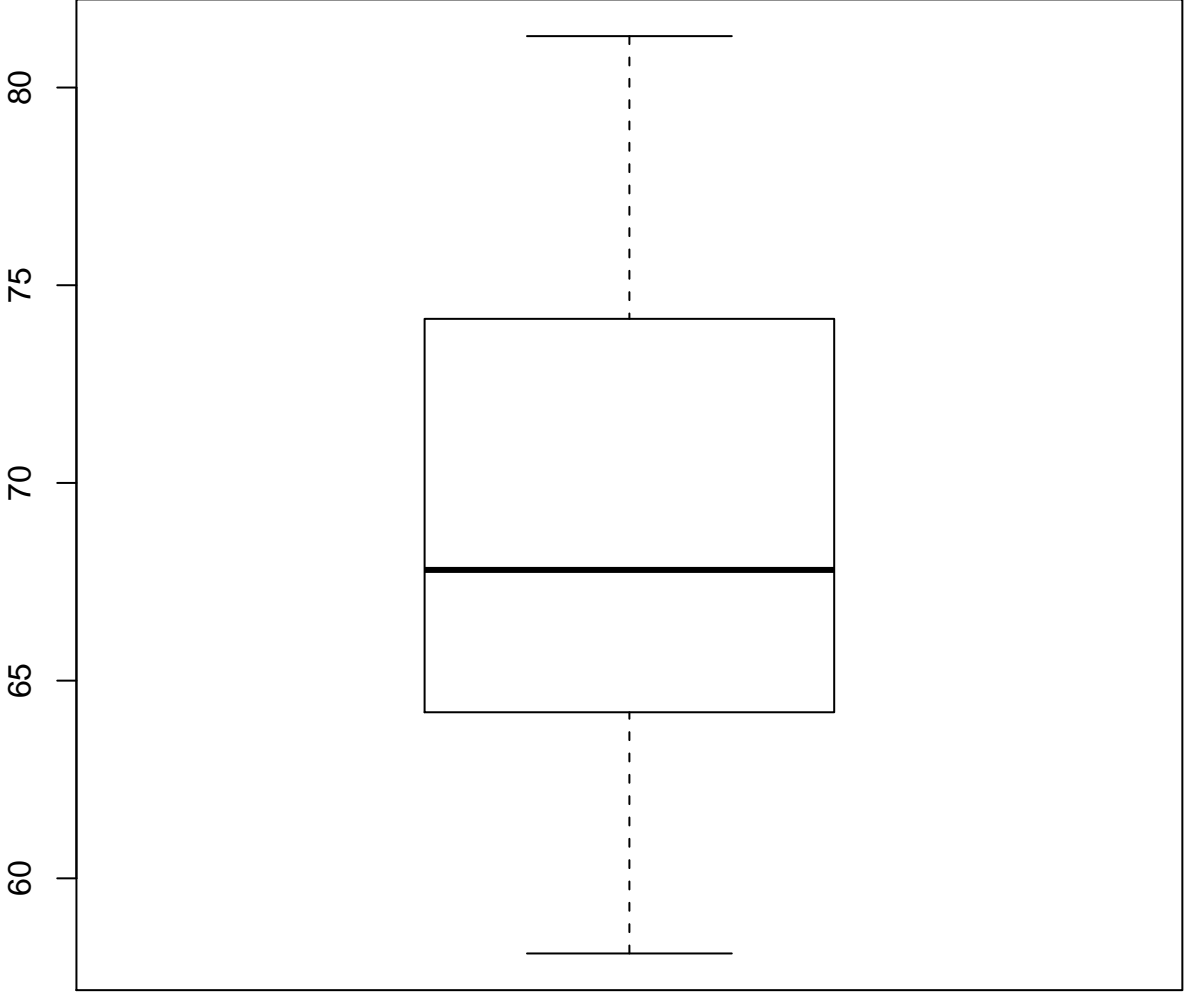
- Mesure de l'adéquation au modèle
- Test d'hypothèses

Rappel du 1er semestre

Étudiants des universités par discipline	
	Année 2005-2006
	Effectif
Droit, sciences politiques	175 853
Sciences économiques, gestion (hors AES)	134 796
Administration économique et sociale (AES)	44 451
Lettres, sciences du langage, arts	111 452
Langues	111 557
Sciences humaines et sociales	245 173
Pluri-lettres-langues-sciences humaines	4 947
Sciences fondamentales et applications	169 158
Sciences de la nature et de la vie	72 389
Sciences et techniques des activités physiques et sportives	41 516
Pluri-sciences	21 617
Médecine - Odontologie	146 589
Pharmacie	29 624
Total hors IUT	1 309 122





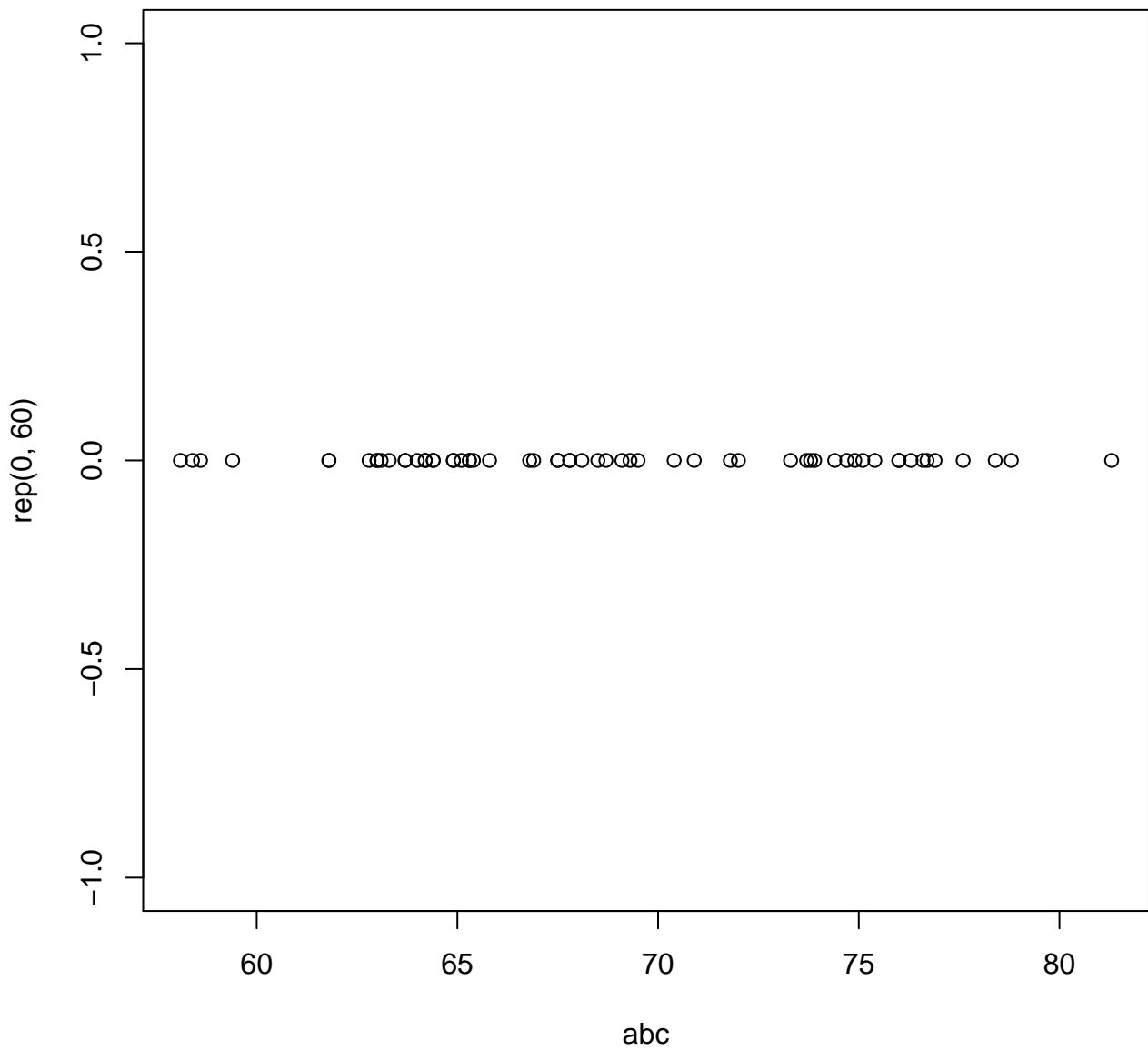


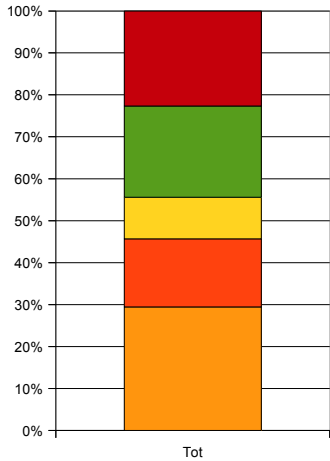
```
> abc
```

```
[1] 65.1 58.4 64.9 76.0 67.8 75.1 76.7 64.2 74.9 77.6 58.1 68.1 73.3 75.4 76.0  
[16] 59.4 65.4 74.7 76.6 81.3 64.4 69.1 66.9 67.5 65.8 70.4 67.8 61.8 68.7 65.3  
[31] 63.7 68.5 72.0 67.5 71.8 64.0 69.5 66.8 64.9 63.0 62.8 58.6 63.3 65.3 78.8  
[46] 63.1 76.3 64.2 61.8 73.9 73.8 76.9 78.4 69.3 63.7 73.7 70.9 63.0 74.4 64.4
```

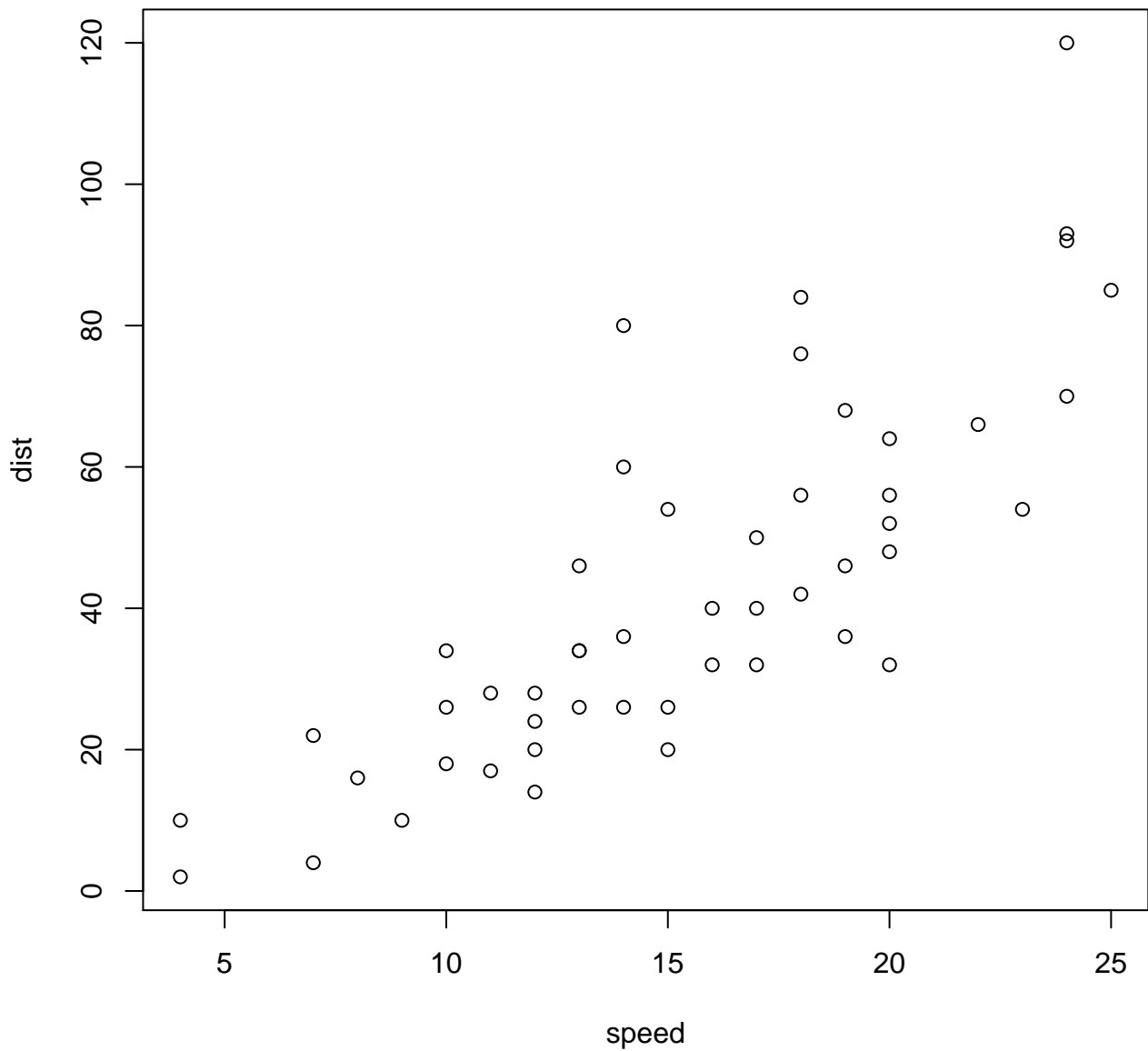
```
> summary(abc)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
58.10	64.20	67.80	68.75	74.03	81.30





- Mathématiques
- Informatique
- Chimie
- Physique
- Sciences de la vie



Cars - data

4 2
4 10
7 4
7 22
8 16
9 10
10 18
10 26
10 34
11 17
11 28
12 14
12 20
12 24
12 28
13 26
13 34
13 34
13 46
14 26
14 36
14 60
14 80
15 20
15 26
15 54
16 32
16 40
17 32
17 40
17 50
18 42
18 56
18 76
18 84
19 36
19 46
19 68
20 32
20 48
20 52
20 56
20 64
22 66
23 54
24 70
24 92
24 93
24 120
25 85

Statistique descriptive en L1AES - semestre 2

TQA, suite du cours du premier semestre

Objectifs du cours :

1. Savoir interpréter les objets mathématiques (tableaux, graphiques ou plus simplement nombres) des rapports publics des observatoires de la vie économique et sociale tels que l'[INSEE](#) (voir des [exemples plus bas](#)).

Pour un tel objet avec sa légende, on répondra aux questions telles que : quelle est la population étudiée, par quels caractères est elle étudiée, quelle est la nature des nombres écrits ou représentés graphiquement. Cf l'exercice 2 de la [feuille de TD2 de 2009-10](#).

2.a. Décrire conjointement plusieurs caractères d'une même population, ou un même caractère dans plusieurs sous-populations. Observer par les fréquences conditionnelles et leurs représentations graphiques l'indépendance ou au contraire la dépendance entre deux caractères ou d'un caractère en la sous-population choisie.

b. Plus abstraitement : notion d'évènements, fréquence d'un évènement, conditionnement, observation de l'indépendance ou d'une corrélation entre évènements.

c. Analyse de la dépendance entre deux caractères quantitatifs par la régression linéaire.

3. Echantillons et modèles statistiques : Echantillons, population entière ou population idéale comme modèle. Représentativité de la fréquence d'un évènement dans un échantillon suivant la taille de l'échantillon, intervalle de confiance, conséquences sur l'observation de l'indépendance entre deux évènements.

Interprétation de la fréquence d'un évènement dans la population entière comme probabilité à l'échelle d'un individu.

Méthodes :

Calcul de fréquences conditionnelles, représentation graphique (barre avec tranches), comparaison des représentations graphiques.

Nuage de points, droite de régression, nuage de points des restes.

Tirage d'échantillons dans une population, histogramme des fréquences observées.

Quelques extraits de rapports publics :

- [Pyramide des ages par pays et par années](#) (histogramme des effectifs par intervalles d'age, comparaison d'histogrammes).
- [Rapport sur les inégalités de niveaux de vie en 2008](#) (quantiles)
- [Etudiants des universités par discipline et par cursus](#) (effectifs conjoints, fréquences conditionnelles)
- [Taux d'activité selon le sexe et la configuration familiale](#) (fréquences conditionnelles)
- [Rapport 2007-2008 de l'Observatoire National de la Pauvreté et de l'Exclusion Sociale, extrait](#) (interprétation probabiliste d'une fréquence p117, relation de dépendance entre deux évènements par comparaison de probabilités conditionnelles ("augmentation du risque") p118, comparaison d'histogrammes p170).
- [Rendement des haricots suivant leur précocité](#) (droite de régression).
- Hérité chez F. Galton : droite de régression (régression vers la moyenne)

Bibliographie pour le cours

A. Hamon & N. Jégou, *Statistique descriptive*, Presse Univ. Rennes 2008.

Lectures :

(livres de médecins, pages web)

[Page du cours en 2009-2010](#)

Données utilisées dans le cours et les TD

1. qq données brutes réelles, ex
 - notes d'examen de TQA
 - tableau de la répartition des étudiants par filière (données déjà agrégées, source INSEE)
 - pyramide des ages (source Insee)

Données utilisées dans le cours et les TD

1. qq données brutes réelles, ex
 - notes d'examen de TQA
 - tableau de la répartition des étudiants par filière (données déjà agrégées, source INSEE)
 - pyramide des ages (source Insee)
2. Données simulées par ordinateur, ex longévité des piles

Données utilisées dans le cours et les TD

1. qq données brutes réelles, ex
 - notes d'examen de TQA
 - tableau de la répartition des étudiants par filière (données déjà agrégées, source INSEE)
 - pyramide des ages (source Insee)

2. Données simulées par ordinateur, ex longévité des piles

Logiciels :

pour l'acquisition de données : un tableur tel LibreOffice,

Données utilisées dans le cours et les TD

1. qq données brutes réelles, ex
 - notes d'examen de TQA
 - tableau de la répartition des étudiants par filière (données déjà agrégées, source INSEE)
 - pyramide des ages (source Insee)

2. Données simulées par ordinateur, ex longévité des piles

Logiciels :

pour l'acquisition de données : un tableur tel LibreOffice,
pour la simulation et le traitement des données : un langage de programmation dédié tel R (r-project.org)

	B	C	D	F	G	H	I	J
1	L1AES TQA Examen session 1 2012							
2				Anal	Stat	Note 1ère session		
20	21001490	BEN HAMOUDA	SAMY	3		3		
21	21108971	BEN M'BAREK	RAHMA	0	1.5	1.5		
22	21107287	BEN-HAMOUDA	RAOUYA	0	1	1		
23	21103315	BEN-SALEM	IMENE	2.5	3.5	6		
24	21103163	BENINATI	WILLIAM	2	4.5	6.5		
25	21003177	BERTHELOT	VALENTIN	0.5	4	4.5		
26	21105739	BICHIOU	SAHBI	6.5	2.5	9		
27	21105440	BORTOLOTTO	AMELIE	0.5	3	3.5		
28	21107928	BOUANGA POATY	JEANNAIS	1	2	3		
29	21005512	BOUCHEIKHI	SHERAZADE	6	3	9		
30	21107787	BOUMEDINE	GHILES	0	1	1		
31	21103434	BOVET	SEBASTIEN	6	7.5	13.5		
32	21005951	CARIDI	DIMITRI	5.5	5	10.5		
33	20903009	CELLARIO	BENJAMIN	4	5.5	9.5		
34	21106098	CHATRY	GWENDOLINE	3.5	4	7.5		
35	21110416	CHENG	XI	5	2	7		
36	21103499	CHOUCHENE	SAYEFF	1.5	4.5	6		
37	21101707	COELHO DIAS	JULIE	9	7.5	16.5		
38	21105671	COLOMES	JEREMIE	3	7.5	10.5		
39	21107404	CUNHA	NADYA	2.5	0	2.5		
40	20903276	DA SILVA PINTO	ODRAN	1	5	6		
41	21102568	DE WAMBERSIE	LISA-MARGOT	2	4.5	6.5		
42	21004750	DELMAS	KEVIN	2.5	4.5	7		
43	21103408	DELOFFRE	FREDERIC	3.5	3.5	7		
44	21103561	DESCAMPS	LAELIA	6	4	10		



Calibri

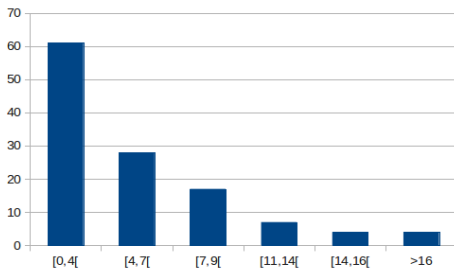
11



B2

 $f(x)$ Σ =

	B	C	D	F	G	H	I	J	
1	L1AES TQA Examen session 1 2012								
2				Anal	Stat	Note 1ère session			
146	<i>Histogramme :</i>								
147	<4	<7	<9	<14	<16				
148	61	89	106	127	131	135			
149	[0,4[[4,7[[7,9[[11,14[[14,16[>16			
150	61	28	17	7	4	4			
151									
152									
153									
154									
155									
156									
157									
158									
159									
160									
161									
162									
163									
164									
165									
166									
167									
168									
169									
170									



File Edit View Search Terminal Help

[18:29 cours1/coursnn2/R] R

R version 2.15.0 (2012-03-30)

Copyright (C) 2012 The R Foundation for Statistical Computing

ISBN 3-900051-07-0

Platform: x86_64-unknown-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

```
> a=read.table(file="A.txt",dec=".",colClasses=c("numeric"))
```

```
>
```

```
> b=read.table(file="B.txt",dec=".",colClasses=c("numeric"))
```

```
>
```

```
> c=read.table(file="C.txt",dec=".",colClasses=c("numeric"))
```

```
>
```

```
> abc=c(t(a),t(b),t(c)) #il faut utiliser la transposee des vecteurs
```

```
>
```

```
> summary(abc)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
58.10	64.20	67.80	68.75	74.03	81.30

```
>
```