MSAG

2ème année - Semestre 3

2021-2022

Bernard.Andruccioli@u-bordeaux.fr

http://aristeri.com

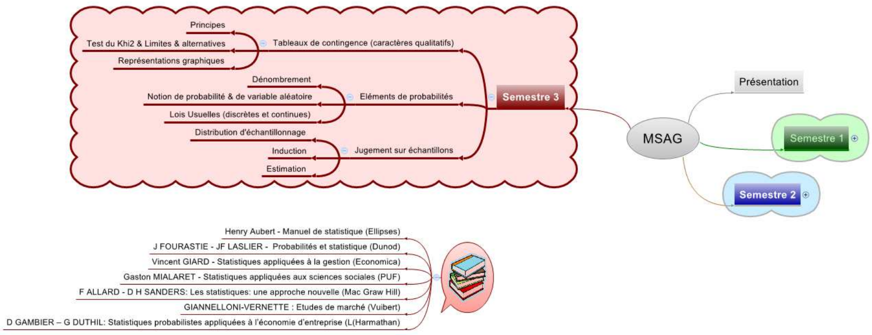
1

✍ Objectif du cours

- Mieux percevoir les difficultés que présente l'analyse des relations entre des caractères.
- Etude des situations liées à l'incertitude.
- Disposer d'outils (complémentaires à ceux vus en marketing et en recherche commerciale) qui permettent d'être plus à l'aise sur des travaux portant sur des échantillons.
- Exploiter les données numériques avec plus de rigueur.
- Liens avec les tableurs, R, et Sphinx.
- MSAG dans une UE parfois délicate pour certains étudiants…mieux vaut éviter les handicaps.

2

Programme & Bibliographie indicatifs



3

✓ Organisation

- Cours en ligne avec vidéos à travailler en amont des séances de TD
- http://aristeri.com
- Login : **etudiants**
- Mot de passe : **secouez_vous**
- **Cas à traiter en TD**
- **Groupes de 3 étudiants**
- **Travail noté**
- **Pas d'utilisation de téléphone ni en amphi, ni en TD**
- **Présence et participation active souhaitée.**

Exclusion automatique du TD sinon

4

Organisation pratique

- Un seul Amphi (si possible)
- Cours en ligne à bosser avant chaque TD,
- http://aristeri.com/MSAG
- Login: **etudiants**
- Mot de passe: **secouez_vous**
- TD: étudiants actifs avec QCM ponctuels

5

Modalités de contrôle

- Participation aux **tests du mois de décembre**: épreuve écrite de **2 heures** (seule la calculatrice du département Tech de Co est autorisée). **Coef: 0,7**.
- QCM + travaux en TD. **Coef: 0,7**
- Les **absences** non ou mal justifiées et les **retards** trop fréquents influenceront de manière **très sensible** la note de contrôle continu => **Malus (Compris entre 0 et 1)**.
- **L'absence de travail**, et/ou de passage au tableau => **Malus (compris entre 0 et 1)**
- **Participation + Travail =>Bonus (Bonus > 1)**
- **Note finale**:
  >Majorité des étudiants:
  Note finale = (Note test x0,7) + (Note QCM x 0,3)
  >Etudiants moins/plus sérieux:
  Note finale = Note test x **Bonus/Malus**

6

Rappel !

Des informations et exercices en ligne sont disponibles sur ma page perso.
(http:/aristeri.com) rubrique Statistiques

7

Partie 1: Compléments sur les séries à deux caractères.

8

Plan Partie 1:

Ordinateurs interdits en amphi!
Merci pour votre compréhension!

Section 1: Rappels tableaux de contingence – caractères quantitatifs

Rappel sur les tableaux de contingence dont les caractères sont quantitatifs:
Exemple: CA (x_i) et Investissement (Y_j): 630 entreprises sont classées selon ces deux critères

$x_i \backslash y_j$	1	3	5	10	ni.
5	78	48	2	0	**128**
10	110	85	3	12	**210**
15	1	3	12	56	**72**
20	2	0	11	125	**138**
50	0	2	1	79	**82**
n.j	**191**	**138**	**29**	**272**	**630**

- ✓ Liens entre le CA et l'investissement ?
- ✓ Impact du CA sur le financement de l'investissement ?
- ✓ Impact de l'investissement sur le CA ?

CA (xi) et Investissement (Yj)

Calculs intermédiaires

$x_i \backslash y_j$	1	3	5	10	ni.	ni.xi	ni.xi²	$\Sigma n_{ij}\, y_j$	$\Sigma n_{ij}\, y_j^2$	$\Sigma n_{ij}\, x_i\, y_j$
5	78	48	2	0	**128**	640	3200	232	560	1160
10	110	85	3	12	**210**	2100	21000	500	2150	5000
15	1	3	12	56	**72**	1080	16200	630	5928	9450
20	2	0	11	125	**138**	2760	55200	1307	12777	26140
50	0	2	1	79	**82**	4100	205000	801	7943	40050
n.j	**191**	**138**	**29**	**272**	**630**	**10680**	**300600**	**3470**	**29358**	**81800**
n.jyj	191	414	145	2720	**3470**					
n.jyj²	191	1242	725	27200	**29358**					
$\Sigma n_{ij}\, x_i$	1545	1235	490	7410	**10680**					
$\Sigma n_{ij}\, x_i^2$	13975	15375	9950	261300	**300600**					
$\Sigma n_{ij} x_i y_j$	1545	3705	2450	74100	**81800**					

11

CA (xi) et Investissement (Yj)

Eléments marginaux

	x	y
Moyennes	16.95	5.51
Variances	189.76	16.26

=3470/630

=29358/630-5,51²

Eléments conditionnels

Caractère x

	x1	x2	x3	x4
Moyennes	8.09	8.95	16.90	27.24
Variances	7.74	31.32	57.61	218.50

=7410/272

=261300/272-27,34²

Caractère y

	y1	y2	y3	y4	y5
Moyennes	1.81	2.38	8.75	9.47	9.77
Variances	1.09	4.57	5.77	2.89	1.45

=801/82

=7943/82-9,77²

12

CA (xi) et Investissement (Yj)

Décomposition des variances

Décomposition variance de x

Var INTER (x)	83.56		R	44.04%
Var INTRA (x)	106.19			

Var totale: 189,76

Décomposition variance de y

Var INTER (y)	13.04		R	80.17%
Var INTRA (y)	3.22			

Var totale: 19,26

CA (xi) et Investissement (Yj)

Corrélation & ajustement

Corrélation

r	0.66	0.66
ρ	0.43	0.43

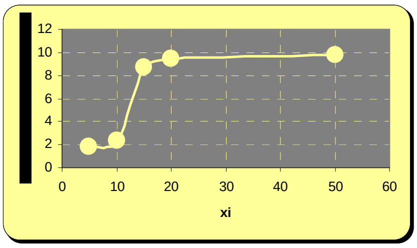
Ajustement

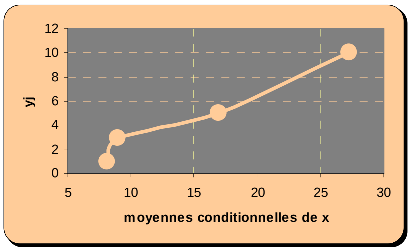
1ère droite	
a	b
0.19	2.25

2ème droite	
a'	b'
2.24	4.60

CA (xi) et Investissement (Yj)

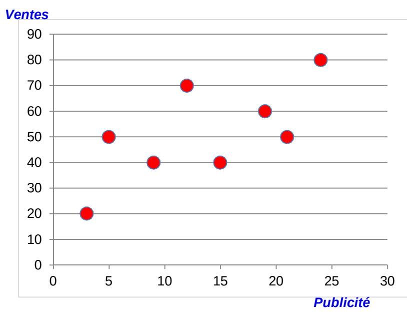
Représentation graphique

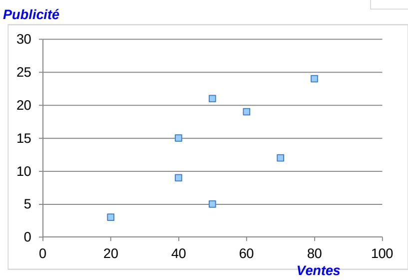




Section 2: Rappels ajustement

Publicité	Ventes
xi	yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8





Section 2: Rappels ajustement – Version 1

Publicité	Ventes
xi	yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8

Version 1				
Xi	Yi	Xi²	Yi²	XiYi
-10,50	-31,25	110,25	976,56	328,13
-8,50	-1,25	72,25	1,56	10,63
-4,50	-11,25	20,25	126,56	50,63
-1,50	18,75	2,25	351,56	-28,13
1,50	-11,25	2,25	126,56	-16,88
5,50	8,75	30,25	76,56	48,13
7,50	-1,25	56,25	1,56	-9,38
10,50	28,75	110,25	826,56	301,88
Sommes				
0	0	404,00	2487,50	685,00

Corrélation

$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \sum Y_i^2}}$$

r	0,683
ρ	0,467

1ère droite d'ajustement

$$y_i = a x_i + b$$

$$a = \frac{\sum X_i Y_i}{\sum X_i^2}$$

a	1,696
b	28,360

$$\overline{y} = a\overline{x} + b$$

$$\Rightarrow b = \overline{y} - a\overline{x}$$

2ème droite d'ajustement

$$x_i = a' y_i + b'$$

$$a' = \frac{\sum X_i Y_i}{\sum Y_i^2}$$

a'	0,275
b'	47,532

$$\overline{x} = a'\overline{y} + b'$$

$$\Rightarrow b' = \overline{x} - a'\overline{y}$$

17

Section 2: Rappels ajustement – Version 2

Publicité	Ventes
xi	yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8

Version 2		
xi²	yi²	xiyi
9	400	60
25	2500	250
81	1600	360
144	4900	840
225	1600	600
361	3600	1140
441	2500	1050
576	6400	1920
Sommes		
1862,00	23500,00	6220,00

Corrélation

$$r = \frac{\frac{\sum x_i y_i}{n} - \overline{x}\,\overline{y}}{\sqrt{\left(\frac{\sum x_i^2}{n} - \overline{x}^2\right) \times \left(\frac{\sum y_i^2}{n} - \overline{y}^2\right)}}$$

r	0,683
ρ	0,467

1ère droite d'ajustement

$$y_i = a x_i + b$$

$$a = \frac{\frac{\sum x_i y_i}{n} - \overline{x}\,\overline{y}}{\frac{\sum x_i^2}{n} - \overline{x}^2}$$

a	1,696
b	28,360

$$\overline{y} = a\overline{x} + b$$

$$\Rightarrow b = \overline{y} - a\overline{x}$$

2ème droite d'ajustement

$$x_i = a' y_i + b'$$

$$a' = \frac{\frac{\sum x_i y_i}{n} - \overline{x}\,\overline{y}}{\frac{\sum y_i^2}{n} - \overline{y}^2}$$

a'	0,275
b'	47,532

$$\overline{x} = a'\overline{y} + b$$

$$\Rightarrow b = \overline{x} - a\overline{y}$$

18

Section 3: Test sur r

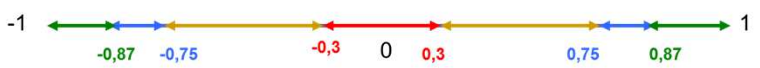
Publicité	Ventes
xi	yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8

Corrélation

r	0,683
ρ	0,467

Problème: Que faire si r est compris entre 0,3 et 0,75 ou entre -0,3 et -0,75?

- **r est petit**, compris entre -0,3 et 0,3. Corrélation insignifiante, indépendance (relative) des caractères
- **r est fort**, >0,87 ou <-0,87=> Corrélation importante; Droites très proches. 75% des variations de y sont expliquées par x. (*Cf B Py*)
- r important, >0,75 ou <-0,75. Liaison relative entre les variables. Présomption de dépendance entre les variables (à tester en fait).
- autres valeurs de r: Impossible de conclure à notre niveau.



19

Section 3: Test sur r

Publicité	Ventes
xi	yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8

Corrélation

r	0,683
ρ	0,467

Méthodologie des tests statistiques

- Fixer et rédiger clairement les hypothèses:
  - H0: hypothèse nulle => r=0 => Aucun lien entre caractères
  - H1: hypothèse alternative =>r ≠0 => Lien entre caractères
- Fixer le risque d'erreur (α)
- Calcul d'un indicateur (dit t calc)
- Détermination d'un indicateur théorique (t theo) (Table du Student)
- Règle de décision permettant le choix entre H0 et H1

20

Section 3: Test sur r

Publicité	Ventes
xi	yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8

Corrélation

r	0,683
ρ	0,467

α = 5%

Indicateur calculé (*t calc*)

$$t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}$$

Donc ici: $$t = \frac{|0.683|\sqrt{8-2}}{\sqrt{1-0.683^2}}$$

Soit *t calc* = 2,2924

Section 3: Test sur r

Publicité	Ventes
xi	yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8

Corrélation

r	0,683
ρ	0,467

α = 5%

Indicateur théorique (*t th*)

=> Table de Student

n-2

ddl \ proba	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693

Soit *t th* = 2,4469 pour un risque d'erreur de 5%

Soit *t th* = 1,9432 pour un risque d'erreur de 10%

Section 3: Test sur r – Règle de décision

Publicité xi	Ventes yi
3	20
5	50
9	40
12	70
15	40
19	60
21	50
24	80
Moyennes	
13,5	51,25
Nombre	
8	8

Corrélation

r	0,683
ρ	0,467

H0 adoptée
si Indicateur théorique (*t th*) > indicateur calculé (*t calc*)

H1 adoptée
si Indicateur théorique (*t th*) < indicateur calculé (*t calc*)

Ici *t calc* = 2,2924

Pour un risque d'erreur de 5%, *t th* = 2,4469
t th > *t calc* => H0 adoptée

Pour un risque d'erreur de 10%, *t th* = 1,9432
t th < *t calc* => H1 adoptée

23

Plan Partie 1:

Chapitre 1- Rappels Tableaux de contingence et ajustement
Chapitre 2- Tableaux de contingence – Caractères qualitatifs
Chapitre 3- Limites et intérêts
Chapitre 4- Test de Kolmogorov
Chapitre 5- Décomposition des variances
Chapitre 6- Rangs
Chapitre 7- Représentations graphiques
Chapitre 8- Liens Excel, OpenOffice, R , et Sphinx.
Chapitre 9– Autre exemple.

24

Section 1 -Tableaux de contingence
Caractères qualitatifs: Présentation

Exemple page :

Résultat d'une élection cantonale

Canton de 3 villes

3 candidats en lice

Caractères qualitatifs (noms des candidats (x_i), noms des villes (y_j)

Effectifs partiels (70 électeurs ont voté pour Ballarac dans la ville de Floiran)

	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	215
Josdur	60	75	55	190
Chipin	50	75	50	175
Total	190	220	170	580

Totaux partiels du caractère xi (nombre de voix par candidat)

Nombre total de votants

Totaux partiels du caractère yj (nombre de votants par ville)

25

Fréquences

Ballarac a obtenu 80 voix dans la ville de Caudérac…

	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	215
Josdur	60	75	55	190
Chipin	50	75	50	175
Total	190	220	170	580

Fréquences absolues

…soit 13,79% des voix

80/580

	Caudérac	Floiran	Pesgnan	Total
Ballarac	13.79%	12.07%	11.21%	37.07%
Josdur	10.34%	12.93%	9.48%	32.76%
Chipin	8.62%	12.93%	8.62%	30.17%
Total	32.76%	37.93%	29.31%	100.00%

Fréquences conditionnelles

Ballarac a obtenu 37,21% du total de ses voix dans la ville de Caudérac

80/215

	Caudérac	Floiran	Pesgnan	Total
Ballarac	37.21%	32.56%	30.23%	100.00%
Josdur	31.58%	39.47%	28.95%	100.00%
Chipin	28.57%	42.86%	28.57%	100.00%
Total	32.76%	37.93%	29.31%	100.00%

En revanche 42,11% des votants de Caudérac ont voté pour lui.

80/190

	Caudérac	Floiran	Pesgnan	Total
Ballarac	42.11%	31.82%	38.24%	37.07%
Josdur	31.58%	34.09%	32.35%	32.76%
Chipin	26.32%	34.09%	29.41%	30.17%
Total	100.00%	100.00%	100.00%	100.00%

Questions: l'environnement des villes influence-t-il le choix des électeurs qui y résident? Existe-il un lien entre ces deux caractères?

Le Test du Khi2 peut donner des éléments de réponse.

26

Section 2: Test du KHI2 (χ^2)

Objectif

- Tester le lien éventuel entre deux caractères =>incontournable dès que l'un des caractères est qualitatif.
- Complément important des tris croisés dans un questionnaire. (Un questionnaire sans tri croisé apporte peu d'informations)
- Permet de conclure sur la base de résultats issus d'une méthodologie pertinente.
- Se méfier des fausses évidences:

	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	**215**
Josdur	60	75	55	**190**
Chipin	50	75	50	**175**
Total	**190**	**220**	**170**	**580**

	Gralence	Lomont	Tagnan	Total
Medelin	80	70	65	**215**
Pinmou	60	**95**	55	**210**
Josrac	50	**65**	**75**	**190**
Total	**190**	**230**	**195**	**615**

Les situations ci-dessus sont-elles différentes? L'influence des villes et de leurs environnements sont-ils plus forts dans un des cantons?

Impossible de répondre sans une méthodologie solide!

- Repose sur la théorie des tests statistiques (non paramétriques)

27

Section 2: Test du KHI2 (χ^2)

Principe général (Page ? poly)

- Fixer et rédiger clairement les hypothèses:
  - H0: hypothèse nulle
  - H1: hypothèse alternative
- Fixer le risque d'erreur (α)
- Calcul d'un indicateur (dit KHI2 calculé: χ^2 calc)
- Détermination d'un indicateur théorique (χ^2 theo) (Table du KHI2)
- Règle de décision permettant le choix entre H0 et H1

28

Section 2: Test du KHI2 (χ^2)

- Fixer et rédiger clairement les hypothèses:

  H0: hypothèse nulle

  => Pas de lien entre les caractères

  Ici: Choix des électeurs pas affecté par le lieu de résidence

  H1: hypothèse alternative

  => Lien entre les caractères

  Ici: le lieu de résidence affecte le choix des électeurs

- Fixer le risque d'erreur (α):

  (α=1%, 2%, 5%; 10%)

Section 2: Test du KHI2 (χ^2)

- Détermination de l'indicateur calculé (χ^2 calc) => 2 étapes:

A -Calcul des effectifs théoriques (H0)

Résultat des élections (Observations)

	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	215
Josdur	60	75	55	190
Chipin	50	75	50	175
Total	190	220	170	580

Ballarac a obtenu 80 voix dans la ville de Caudérac

Sous l'hypothèse H0 il aurait du en obtenir environ 70

70,43=(190 x 215)/580

Effectifs théoriques, donc calculés

	Caudérac	Floiran	Pesgnan	Total
Ballarac	70.43	81.55	63.02	215
Josdur	62.24	72.07	55.69	190
Chipin	57.33	66.38	51.29	175
Total	190	220	170	580

Effectif théorique: ($n_{i.}$ X $n_{.j}$)/n..

Section 2: Test du KHI2 (χ^2)

- Détermination de l'indicateur calculé (χ^2 calc) => 2 étapes:

B - χ2 calculé

Observations (1er tableau) (O)	Calculs (2nd tableau) (C)	O-C	(O - C)²	(O - C)² / C
80	70.43	9.57	91.57	1.300
70	81.55	-11.55	133.44	1.636
65	63.02	1.98	3.93	0.062
60	62.24	-2.24	5.02	0.081
75	72.07	2.93	8.59	0.119
55	55.69	-0.69	0.48	0.009
50	57.33	-7.33	53.69	0.937
75	66.38	8.62	74.32	1.120
50	51.29	-1.29	1.67	0.033
			Somme:	5.296

χ^2 Calculé

$$\chi^2{}_{calc} = \sum \frac{(O-C)^2}{C}$$

31

Section 2: Test du KHI2 (χ^2)

- Détermination de l'indicateur théorique (χ^2 th) => Table du khi2

❖ **Risque d'erreur: α=5%**

❖ **Degré de liberté (ν)**

ν=(nb lignes – 1) x (nb colonnes – 1)

	Caudérac	Floiran	Pesgnan
Ballarac	80	70	65
Josdur	60	75	55
Chipin	50	75	50

ν= (3-1) x (3-1) = 4

Table de la loi du Khi2

ν / α	50%	20%	15%	10%	5%	2%	1%
1	0.455	1.642	2.072	2.706	3.841	5.412	6.635
2	1.386	3.219	3.794	4.605	5.991	7.824	9.210
3	2.366	4.642	5.317	6.251	7.815	9.837	11.345
4	3.357	5.989	6.745	7.779	9.488	11.668	13.277
5	4.351	7.289	8.115	9.236	11.070	13.388	15.086
6	5.348	8.558	9.446	10.645	12.592	15.033	16.812
7	6.346	9.803	10.748	12.017	14.067	16.622	18.475
8	7.344	11.030	12.027	13.362	15.507	18.168	20.090
9	8.343	12.242	13.288	14.684	16.919	19.679	21.666
10	9.342	13.442	14.534	15.987	18.307	21.161	23.209
11	10.341	14.631	15.767	17.275	19.675	22.618	24.725
12	11.340	15.812	16.989	18.549	21.026	24.054	26.217
13	12.340	16.985	18.202	19.812	22.362	25.472	27.688
14	13.339	18.151	19.406	21.064	23.685	26.873	29.141
15	14.339	19.311	20.603	22.307	24.996	28.259	30.578

Ici χ^2 théorique: 9,488 (Table page 15)

32

Section 2: Test du KHI2 (χ^2)

- Règle de décision: confrontation des deux indicateurs.

- χ^2 calculé < χ^2 théorique
  H0 retenu
- χ^2 calculé > χ^2 théorique
  H1 retenu

Ici χ^2calc = 5,296, et χ^2theo = 9,488
Donc nous devons retenir l'hypothèse H0
Pas d'influence de la commune de résidence sur le choix des électeurs (95%)

33

Section 2: Test du KHI2 (χ^2)

Observations	Gralence	Lomont	Tagnan	Total
Medelin	80	70	65	215
Pinmou	60	95	55	210
Josrac	50	65	75	190
Total	190	230	195	615

Effectifs théoriques	Gralence	Lomont	Tagnan	Total
Medelin	66.42	80.41	68.17	215
Pinmou	64.88	78.54	66.59	210
Josrac	58.70	71.06	60.24	190
Total	190	230	195	615

Khi2 calculé	Gralence	Lomont	Tagnan	Total
Medelin	2.78	1.35	0.15	4.270
Pinmou	0.37	3.45	2.02	5.834
Josrac	1.29	0.52	3.61	5.420
Total	4.431	5.314	5.778	15.523

Ici χ^2calc = 15,523, et χ^2theo = 9,488
Donc nous devons retenir l'hypothèse H1
Influence notable de la commune de résidence sur le choix des électeurs (95%)

34

Section 3: χ^2 et mesures d'association

- Tableaux de taille supérieure à 2 x 2.

Le Coefficient de Contingence Cc:

$$Cc = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

1er cas $$Cc = \sqrt{\frac{5{,}296}{5{,}296 + 580}} = 0{,}095$$

2ème cas $$Cc = \sqrt{\frac{15{,}523}{15{,}523 + 615}} = 0{,}156$$

Interprétation délicate: Cc varie de 0 à une valeur qui ne peut jamais atteindre 1. Cc=0: indépendance totale entre les caractères. Plus Cc est fort plus le lien entre les caractères est important. Cc ne peut bien sûr pas être négatif…

- Tableaux de taille 2 x 2.

Le Coefficient d'association ϕ:

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

ϕ varie entre 0 et 1; interprétation plus aisée.

Section 4: Recherche des contributions au χ^2

Khi2 calc

	Caudérac	Floiran	Pesgnan	Total
Ballarac	1.300	1.636	0.062	2.999
Josdur	0.081	0.119	0.009	0.208
Chipin	0.937	1.120	0.033	2.089
Total	2.317	2.875	0.104	5.296

Contribution au Khi2

	Caudérac	Floiran	Pesgnan	Total
Ballarac	24.55%	30.90%	1.18%	56.62%
Josdur	1.52%	2.25%	0.16%	3.94%
Chipin	17.69%	21.14%	0.62%	39.44%
Total	43.76%	54.29%	1.95%	100.00%

Si le test montrait un lien entre ces deux caractères, les cellules en évidence seraient à l'origine de cette relation. C'est le cas dans l'exemple ci-dessous.

Khi2 calculé	Gralence	Lomont	Tagnan	Total
Medelin	2.78	1.35	0.15	4.270
Pinmou	0.37	3.45	2.02	5.834
Josrac	1.29	0.52	3.61	5.420
Total	4.431	5.314	5.778	15.523

Contributions	Gralence	Lomont	Tagnan	Total
Medelin	17.88%	8.68%	0.95%	0.275
Pinmou	2.36%	22.23%	12.99%	0.376
Josrac	8.31%	3.33%	23.28%	0.349
Total	0.285461	0.342348	0.372191	1.000

Plan Partie 1:

37

Section 1: Limites du χ^2

Restrictions:

- Effectif total (n..) >30
- Chaque effectif théorique doit être >5.

Observations

Puissance\Age	<20	>20	Total
<=5	47	5	52
6 et 7	5	5	10
8 et 9	19	15	34
>9	6	10	16
Total	77	35	112

Effectifs calculés

Puissance\Age	<20	>20	Total
<=5	35.75	16.25	52
6 et 7	6.88	3.13	10
8 et 9	23.38	10.63	34
>9	11.00	5.00	16
Total	77	35	112

- Les conditions d'utilisations du test ne sont pas respectées.
- On ne peut donc en tirer aucune conclusion.
- **Intérêt pour justifier la taille d'un échantillon…**

Khi2 calculé

Puissance\Age	<20	>20	Total
<=5	3.540	7.788	11.329
6 et 7	0.511	1.125	1.636
8 et 9	0.819	1.801	2.620
>9	2.273	5.000	7.273
Total	7.143	15.715	22.858

Ici χ^2th=7,815 => H1 ?

38

Section 1: Limites du χ^2

- Que faire si les conditions du test du khi2 ne sont pas remplies?
  - **Tableaux de taille supérieure à 2 x 2.**

➢ Regrouper des modalités quand c'est possible

Observations

Puissance	<20	>20	Total
<=5	47	5	52
6 et 7	5	5	10
8 et 9	19	15	34
>9	6	10	16
Total	77	35	112

Observations

Puissance	<20	>20	Total
<=7	52	10	62
>=8	25	25	50
Total	77	35	112

Effectifs calculés

Puissance	<20	>20	Total
<=7	42.63	19.38	62
>=8	34.38	15.63	50
Total	77	35	112

Khi2 calculé

Puissance	<20	>20	Total
<=7	2.062	4.536	6.598
>=8	2.557	5.625	8.182
Total	4.619	10.161	14.780

χ^2th=3,841… donc…

➢ Utiliser la décomposition des variances ou le test de Kolmogorov

39

Section 2: χ^2 Tableaux 2 x 2

- Que faire si les conditions du test du khi2 ne sont pas remplies?
  - **Tableaux de taille 2 x 2: Correction de YATES**

Observations (O)

	P1	P2	Total
H	6	3	9
F	4	9	13
Total	10	12	22

Effectifs théoriques (C)

	P1	P2	Total
H	4.09	4.91	9
F	5.91	7.09	13
Total	10	12	22

$$\chi^2 = \frac{(|O-C|-1/2)^2}{C}$$

Khi2 Calculé avec Yates

	P1	P2
H	0.49	0.40
F	0.34	0.28

Khi2 calculé	1.5058

40

Section 2: χ^2 Tableaux 2 x 2

☺ Tableaux de taille 2 x 2: Autre procédé de calcul

Calcul précédent:

Observations

	P1	P2	Total
H	6	11	17
F	12	9	21
Total	18	20	38

Effectifs théoriques

	P1	P2	Total
H	8.05	8.95	17
F	9.95	11.05	21
Total	18	20	38

Khi2 Calculé

	P1	P2
H	0.52	0.47
F	0.42	0.38

Khi2 calculé 1.7989

Notations:

	P1	P2	Total
H	a	b	a + b
F	c	d	c + d
Total	a + c	b + d	a+b+c+d=n

Autre formule:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$\chi^2 = \frac{38\times(6\times9-11\times12)^2}{18\times20\times17\times21} = \frac{38\times(54-132)^2}{128520} = 1{,}7989$$

Ici Khi2 Théorique = 3,8414 donc H0 retenue.

41

Section 2: χ^2 Tableaux 2 x 2

Extension: Méthode de Belson (exemple précédent)

Objectif étude de marché: déterminer s'il existe des variables explicatives des comportements.
Test du Khi2 permet d'y arriver =>Technique de segmentation.
Méthode de Belson permet de déterminer si un caractère permet la segmentation.

Principe: Si un caractère n'est pas explicatif on aura une égalité entre les effectifs observés et théoriques. Sinon ce caractère comporte une part d'explication: distance entre effectifs observés et théoriques.

Observations

	P1	P2	Total
H	6	11	17
F	12	9	21
Total	18	20	38

Nb produits P1 achetés par des hommes.

Théoriques

	P1	P2	Total
H	8	9	17
F	10	11	21
Total	18	20	38

(Arrondi à l'entier le plus proche)

Nb produits P1 qu'auraient du acheter les hommes

Distances

	P1	P2	Total
H	-2	2	0
F	2	-2	0
Total	0	0	0

Ici distance (notée B) = 2.

42

Section 2: χ^2 Tableaux 2 x 2
Extension: Méthode de Belson (exemple précédent)

Distances

	P1	P2	Total
H	-2	2	0
F	2	-2	0
Total	0	0	0

Ici distance = 2.

Distance nommée B.

A partir de quelle distance peut-on considérer cette variable comme explicative?

$$B > \frac{49}{\sqrt{n..}}$$

Donc ici:

$$B > \frac{49}{\sqrt{38}}$$

B doit dont être supérieur à 7,95 pour que la variable sexe soit explicative ici et permette de segmenter.

Or B = 2 donc le sexe n'est pas explicatif des comportements ici.

Conclusion qui rejoint celle du test du Khi2 (H0)

Plan Partie 1:

Test de Kolmogorov

Objectif

- Tester le lien éventuel entre deux caractères =>incontournable dès que l'un des caractères est qualitatif.
- Repose sur la théorie des tests statistiques (non paramétriques)
- Permet de conclure dans les cas de petits effectifs

Observations

Puissance	<20	>20	Total
<=5	47	5	52
6 et 7	5	5	10
8 et 9	19	15	34
>9	6	10	16
Total	77	35	112

45

Test de Kolmogorov

Principe général identique au test du Khi2

- Fixer et rédiger clairement les hypothèses:
  - H0: hypothèse nulle
  - H1: hypothèse alternative
- Fixer le risque d'erreur (α)
- Calcul d'un indicateur (dit K calculé: k calc)
- Détermination d'un indicateur théorique (k theo) (Table de Kolmogorov)
- Règle de décision permettant le choix entre H0 et H1

Dans les deux cas (Khi2 et Kolmogorov) on a besoin de déterminer (de manière identique) les effectifs théoriques.

46

Test de Kolmogorov

- Calcul d'un indicateur calculé (dit K calculé: k calc)

Observations	Calculés	Obs C C	Calc C C	|Ecart|
47	35.75	47	35.75	11.25
5	6.88	52	42.625	9.38
19	23.38	71	66	5.00
6	11.00	77	77	0.00
5	16.25	82	93.25	11.25
5	3.13	87	96.375	9.38
15	10.63	102	107	5.00
10	5.00	112	112	0.00
			Max	11.25
			k calc	0.100

K calc=|Ecart| max / n

47

Test de Kolmogorov

- Détermination d'un indicateur théorique (dit K théorique: k theo)

Table de Kolmogorov
(page ?)

$$ktheo = \frac{1,36}{\sqrt{112}} = 0,128$$

Table de Kolmogorov-Smirnov

n	Risque d'erreur 0.2	0.15	0.1	0.05	0.01
1	0.9	0.925	0.95	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.51	0.565	0.669
6	0.41	0.436	0.47	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.36	0.388	0.432	0.514
10	0.322	0.342	0.368	0.41	0.49
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.45
13	0.284	0.302	0.325	0.361	0.433
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.328	0.392
17	0.25	0.266	0.286	0.318	0.381
18	0.244	0.259	0.278	0.309	0.371
19	0.237	0.252	0.272	0.301	0.363
20	0.231	0.246	0.264	0.294	0.356
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.2	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
> 35	$\frac{1,07}{\sqrt{n}}$	$\frac{1,14}{\sqrt{n}}$	$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

48

Test de Kolmogorov

- Règle de décision (Similaire au test du khi2)

k calculé < k théorique
H0 retenu
k calculé > k théorique
H1 retenu

Ici kcalc = 0,1 et ktheo = 0,128
Donc nous devons retenir l'hypothèse H0

49

Plan Partie 1:

50

Décomposition des variances

Même cas (*Effectifs théoriques < 5*)

Observations

Puissance \ Age	<20	>20	Total
<= 5	47	5	52
6 et 7	5	5	10
8 et 9	19	15	34
> 9	6	10	16
Total	77	35	112

	Moyennes	Nombres	Somme carrés	Variances
<= 5	26	2	2234	441
6 et 7	5	2	50	0
8 et 9	17	2	586	4
> 9	8	2	136	4

Puissance

	<20	>20
Nombre	4	4
Moyennes	19,25	8,75
Sommes carrés	2631	375
Variances	287,19	17,19

Age

	Population
Nombre	8
Moyenne	14
Sommes carrés	3006
Variance (Totale)	179,75

En théorie on devrait avoir ici des valeurs proches si la puissances des autos était répartie « équitablement ».

Comptabilisation effectuée en fonction de la puissance des autos utilisées selon l'âge. En théorie on devrait avoir des valeurs très proches si aucun impact… => Si différence sensible entre ces colonnes cela provient de ces puissances …et les parents réfléchissent avant de (vous) confier une voiture puissante.

51

Décomposition des variances

Même cas (*Effectifs théoriques < 5*)

Observations

Puissance \ Age	<20	>20	Total
<= 5	47	5	52
6 et 7	5	5	10
8 et 9	19	15	34
> 9	6	10	16
Total	77	35	112

	<20	>20
Nombre	4	4
Moyennes	19,25	8,75
Sommes carrés	2631	375
Variances	287,19	17,19

Age

	Population
Nombre	8
Moyenne	14
Sommes carrés	3006
Variance (Totale)	179,75

Comptabilisation effectuée en fonction de la puissance des autos utilisées selon l'âge. En théorie on devrait avoir des valeurs très proches si aucun impact… => Si différence sensible entre ces colonnes cela provient de ces puissances …et les parents réfléchissent avant de (vous) confier une voiture puissante.

Var INTER = Variance des Moyennes = (19,25 ² + 8,75 ²) / 2 – 14 ² = 27,56

Var INTRA = Moyenne des Variances = (287,19 + 17,19) / 2 = 152,19

52

Décomposition des variances

Rappel: **Var Totale = Var INTER + Var INTRA**

179,75 = 27,56 + 152,19

R: Rapport de corrélation

R = (Var Inter / Var Totale) x 100

R = (27,65 / 179,75) x 100 = 15,33%

Chiffre faible qui laisse penser que l'âge n'intervient pas directement .

Décomposition des variances

Même cas (*Effectifs théoriques < 5*)

Observations

Puissance \ Age	<20	>20	Total
<= 5	47	5	52
6 et 7	5	5	10
8 et 9	19	15	34
> 9	6	10	16
Total	77	35	112

	Moyennes	Nombres	Somme carrés	Variances
<= 5	26	2	2234	441
6 et 7	5	2	50	0
8 et 9	17	2	586	4
> 9	8	2	136	4

Puissance

	Population
Nombre	8
Moyenne	14
Sommes carrés	3006
Variance (Totale)	179,75

En théorie on devrait avoir ici des valeurs proches si la puissances des autos était répartie « équitablement ». Moins pertinent à priori.

Var INTER = Variance des Moyennes = $(26^2 + 5^2 + 17^2 + 8^2) / 4 - 14^2 = 67,5$

Var INTRA = Moyenne des Variances = (441 + 0 + 4 + 4) / 4 = 112,25

Décomposition des variances

Rappel: Var Totale = Var INTER + Var INTRA

179,75 = 67,5 + 112,25

R: Rapport de corrélation

R = (Var Inter / Var Totale) x 100

R = (67,5 / 179,75) x 100 = 37,55%

Chiffre faible également.

55

Plan Partie 1:

56

Section 2: Rangs et coefficients des rangs de Spearman (Rs ou Rho)

Observations

	Jeunes	Vieux
Puissance \ âge	<20	>20
<=5	47	5
6 et 7	5	5
8 et 9	19	15
>9	6	10

Rangs (préférence)

	Jeunes	Vieux
Puissance \ âge	<20	>20
<=5	1	3
6 et 7	4	3
8 et 9	2	1
>9	3	2

$$Rs = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

Coefficient de corrélation des rangs (Spearman)
=> Interprétation et test comme r,

57

Section 2: Rangs et coefficients des rangs de Spearman

Rangs (préférence)

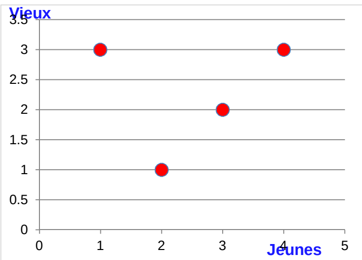
	Jeunes	Vieux		
Puissance \ âge	<20	>20	Di	Di²
<=5	1	3	-2	4
6 et 7	4	3	1	1
8 et 9	2	1	1	1
>9	3	2	1	1
			Total	7

$$Rs = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

Ici:

$$Rs = 1 - \frac{6 \times 7}{4(4^2-1)}$$

Soit Rs = 0,3 => Faiblard.



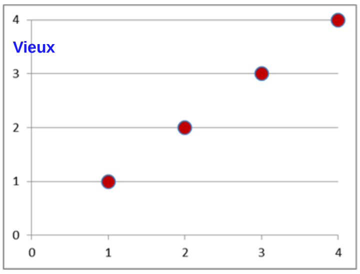
58

Section 2: Rangs et coefficients des rangs de Spearman

Classements identiques

	Jeunes	Vieux		
Puissance \ âge	<20	>20	Di	Di²
<=5	1	1	0	0
6 et 7	2	2	0	0
8 et 9	3	3	0	0
>9	4	4	0	0
			Total	0

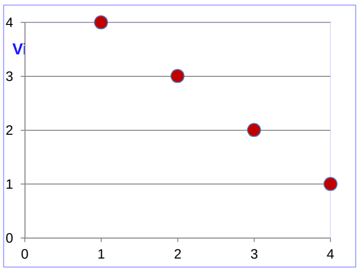
Rs = 1



Classements inversés

	Jeunes	Vieux		
Puissance \ âge	<20	>20	Di	Di²
<=5	1	4	-3	9
6 et 7	2	3	-1	1
8 et 9	3	2	1	1
>9	4	1	3	9
			Total	20

Rs = -1



Section 2: Rangs et coefficients des rangs de Spearman

Marque \ Sexe	F	H	Total
Renault	7	20	27
Peugeot	16	13	29
Citroên	8	10	18
VW	10	9	19
Fiat	19	7	26
Ford	5	2	7
Toyota	6	4	10
Opel	12	8	20
Total	83	73	156

=>

Marque \ Sexe	F	H	Di	Di²
Renault	1	1	0	0
Peugeot	3	2	1	1
Citroên	6	3	3	9
VW	5	4	1	1
Fiat	2	6	-4	16
Ford	8	8	0	0
Toyota	7	7	0	0
Opel	4	5	-1	1
			∑ Di²	28

Ici:

$$Rs = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

$$Rs = 1 - \frac{6 \times 28}{8(8^2 - 1)}$$

Soit Rs = 0,66 => A *tester* donc!

Section 2: Rangs et coefficients des rangs de Spearman

Marque \ Sexe	F	H	Total
Renault	7	20	27
Peugeot	16	13	29
Citroên	8	10	18
VW	10	9	19
Fiat	19	7	26
Ford	5	2	7
Toyota	6	4	10
Opel	12	8	20
Total	83	73	156

=>

Marque \ Sexe	F	H	Di	Di²
Renault	1	1	0	0
Peugeot	3	2	1	1
Citroên	6	3	3	9
VW	5	4	1	1
Fiat	2	6	-4	16
Ford	8	8	0	0
Toyota	7	7	0	0
Opel	4	5	-1	1
			Σ Di²	28

Test Student: risque d'erreur 5% et 10%.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = \frac{|0.666|\sqrt{8-2}}{\sqrt{1-0.666^2}}$$

t calc = 2,191

ddl \ proba	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693

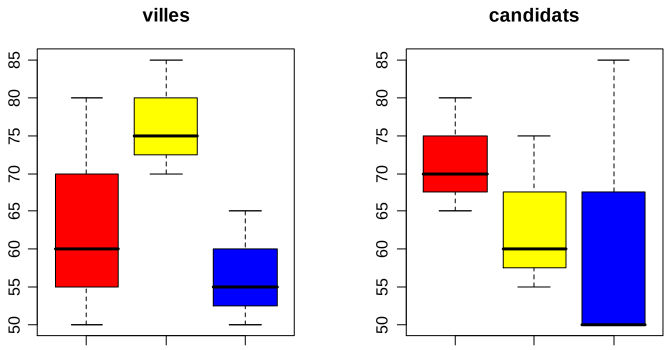
H1 est donc retenue!

Plan Partie 1:

Chapitre 1- Rappels Tableaux de contingence et ajustement
Chapitre 2- Tableaux de contingence – Caractères qualitatifs
Chapitre 3- Limites et intérêts
Chapitre 4- Test de Kolmogorov
Chapitre 5- Décomposition des variances
Chapitre 6- Rangs
Chapitre 7- Représentations graphiques
Chapitre 8- Liens Excel, LibreOffice, R , et Sphinx.
Chapitre 9– Autre exemple.

62

Représentations graphiques: boîte à moustaches de Tukey



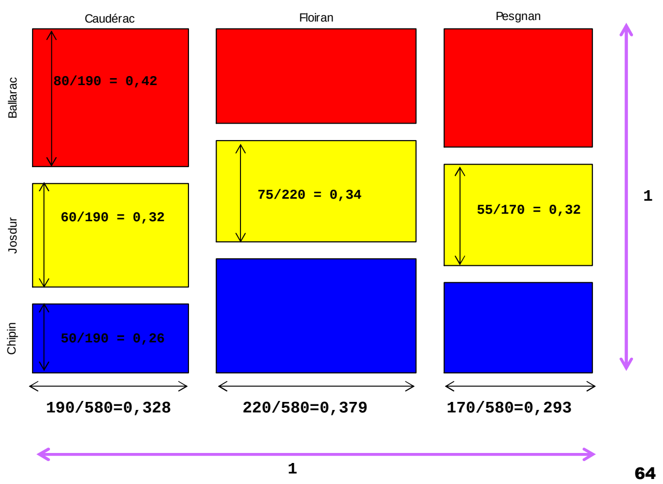
Revoir la méthodologie vue l'an dernier (et poly sur Excel utilisé en TD d'informatique)

Représentations graphiques: Diagramme en mosaïque.

	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	**215**
Josdur	60	75	55	**190**
Chipin	50	75	50	**175**
Total	**190**	**220**	**170**	**580**

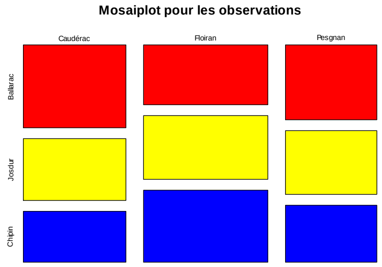
Effectifs observés

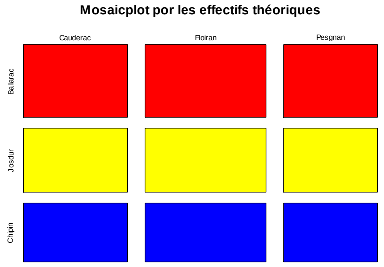
Diagramme en mosaique



Représentations graphiques: Diagramme en mosaïque.

Effectifs observés et théoriques





	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	215
Josdur	60	75	55	190
Chipin	50	75	50	175
Total	190	220	170	580

	Caudérac	Floiran	Pesgnan	Total
Ballarac	70.43	81.55	63.02	215
Josdur	62.24	72.07	55.69	190
Chipin	57.33	66.38	51.29	175
Total	190	220	170	580

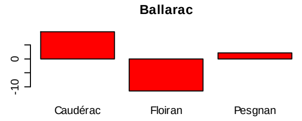
65

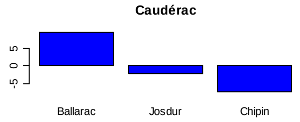
Représentations graphiques: Analyse avec les écarts.

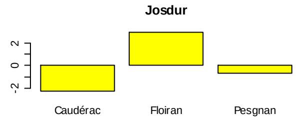
Ecarts=Effectifs observés - Effectifs théoriques

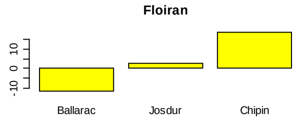
Ecarts

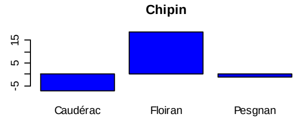
	Caudérac	Floiran	Pesgnan	Total
Ballarac	9.57	-11.55	1.98	0
Josdur	-2.24	2.93	-0.69	0
Chipin	-7.33	8.62	-1.29	0
Total	0	0	0	0

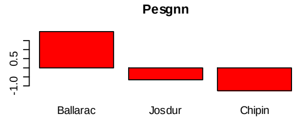












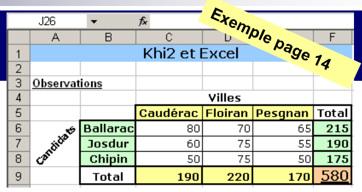
Tous les graphiques ont été créés avec le logiciel R

66

Plan Partie 1:

67

Khi2 et Excel



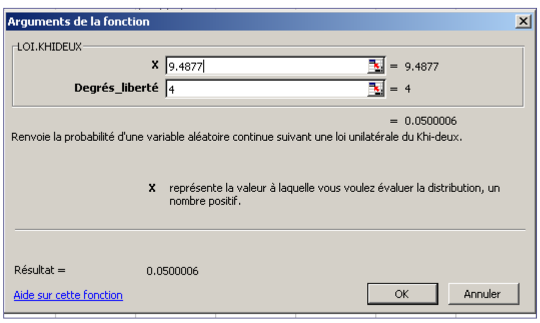
Khi2 et Excel

Observations

		Villes			
		Caudérac	Floiran	Pesgnan	Total
Candidats	Ballarac	80	70	65	215
	Josdur	60	75	55	190
	Chipin	50	75	50	175
	Total	190	220	170	580

3 fonctions relatives au Khi2

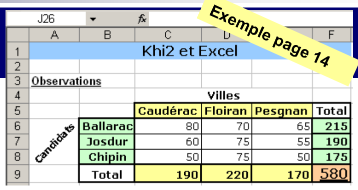
- 1 - LOI.KHIDEUX



Nécessite le ddl et une valeur de x pour renvoyer une probabilité…qui est en fait le risque d'erreur. L'inverse de nos besoins.

68

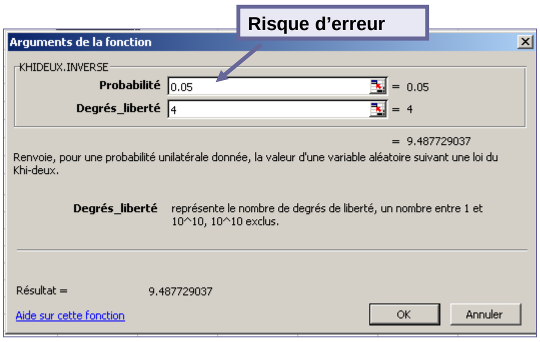
Khi2 et Excel



	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	215
Josdur	60	75	55	190
Chipin	50	75	50	175
Total	190	220	170	580

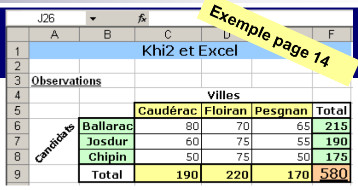
3 fonctions relatives au Khi2

- 2 - KHIDEUX.INVERSE



Nécessite le ddl et une probabilité…qui est en fait le risque d'erreur pour retourner le khi2 théorique. Cela correspond à nos besoins. Nous dispense d'utiliser la table du khi2

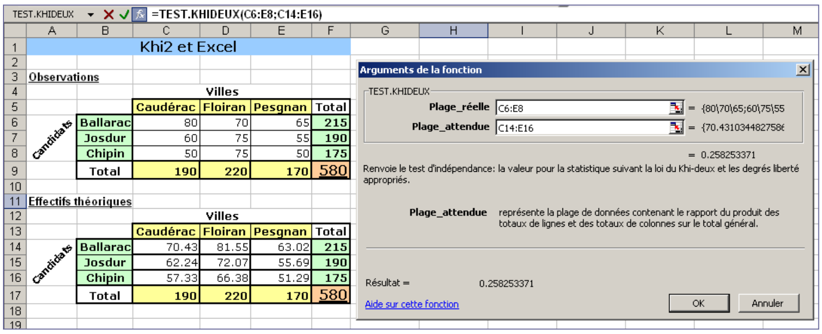
Khi2 et Excel



	Caudérac	Floiran	Pesgnan	Total
Ballarac	80	70	65	215
Josdur	60	75	55	190
Chipin	50	75	50	175
Total	190	220	170	580

3 fonctions relatives au Khi2

- 3 - TEST.KHIDEUX

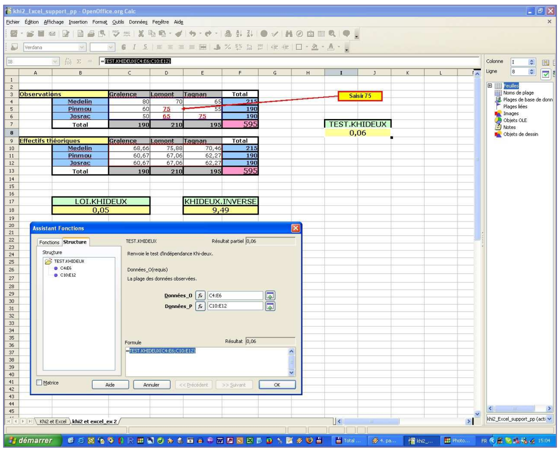


Effectifs théoriques

	Caudérac	Floiran	Pesgnan	Total
Ballarac	70.43	81.55	63.02	215
Josdur	62.24	72.07	55.69	190
Chipin	57.33	66.38	51.29	175
Total	190	220	170	580

Nécessite les effectifs observés (« Plage réelle »), mais aussi (☹) les effectifs théoriques (« Plage attendue ») pour retourner le risque d'erreur qui fait basculer la décision de H1 vers H0.

Khi2 et LibreOffice (Calc)



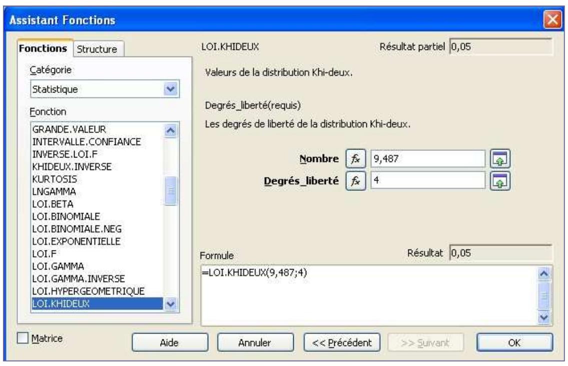
Suite bureautique concurrente de MSOffice. Libre, gratuite, compatible, multiplateforme, fiable, stable, sécurisée, et dotée d'un format de fichier ouvert. Comprend Writer (Traitement de textes), Calc (Tableur), Impress (PréAO), Base (Base de données) et Draw (Dessin vectoriel).

71

Khi2 et LibreOffice.org

3 fonctions relatives au Khi2

- 1 - LOI.KHIDEUX



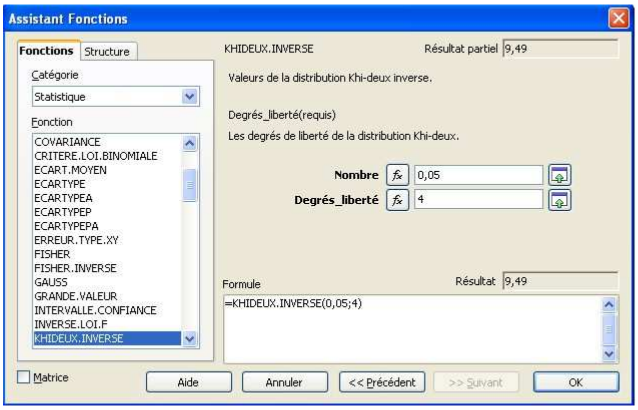
Nécessite le ddl et une valeur de x pour renvoyer une probabilité…qui est en fait le risque d'erreur. L'inverse de nos besoins. Même principe d'utilisation qu'Excel.

72

Khi2 et LibreOffice

3 fonctions relatives au Khi2

- 2 - KHIDEUX.INVERSE



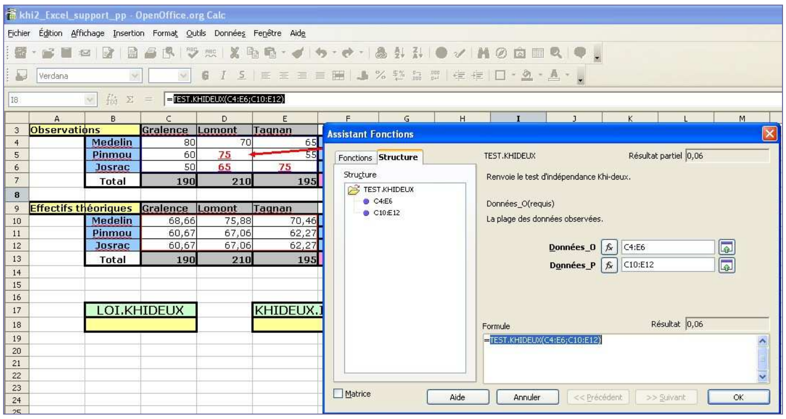
Nécessite le ddl et une probabilité…qui est en fait le risque d'erreur pour retourner le khi2 théorique. Cela correspond à nos besoins. Nous dispense d'utiliser la table du khi2. Idem qu'Excel.

73

Khi2 et LibreOffice

3 fonctions relatives au Khi2

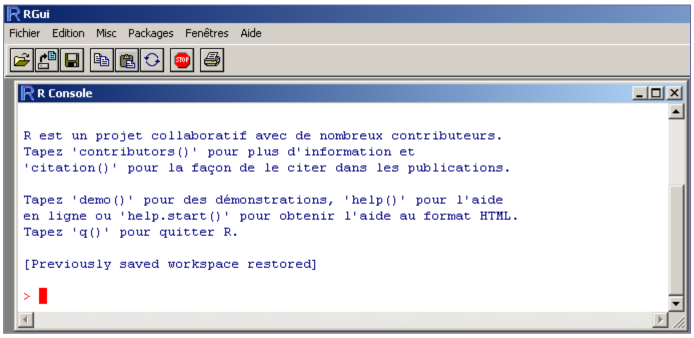
- 3 - TEST.KHIDEUX



Nécessite les effectifs observés (« Plage réelle »), mais aussi (☹) les effectifs théoriques (« Plage attendue ») pour retourner le risque d'erreur qui fait basculer la décision de H1 vers H0. Comme sur Excel…

74

Khi2 et R

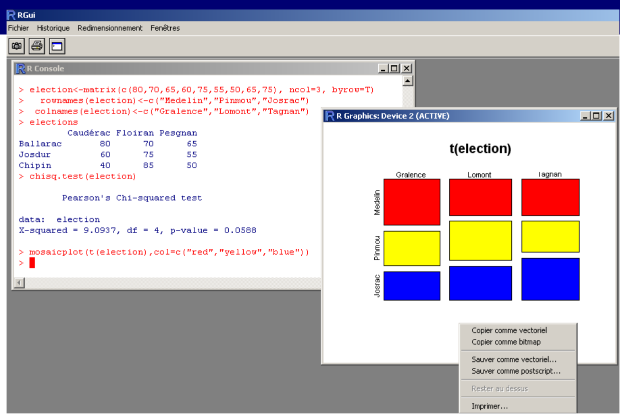


R: logiciel de statistiques libre gratuit, performant, et compatible. Fonctionne en ligne de commandes, et permet la réalisation de graphiques complexes.

75

Khi2 et R

Syntaxe: 1 ligne pour saisir les valeurs, 2 pour nommer les lignes et colonnes (facultatif), 1 pour faire afficher le tableau, 1 ligne pour le test, et 1 dernière pour le diagramme.. Retour du test: seuil de basculement (comme pour la fonction TEST.KHI2 des tableurs).




```
> election<-matrix(c(80,70,65,60,75,55,50,65,75), ncol=3, byrow=T)
> rownames(election)<-c("Medelin","Pinmou","Josrac")
> colnames(election)<-c("Gralence","Lomont","Tagnan")
> chisq.test(election)
 Pearson's Chi-squared test
data:  election
X-squared = 9.0937, df = 4, p-value = 0.0588
> mosaicplot(t(election),col=c("red","yellow","blue"))
```

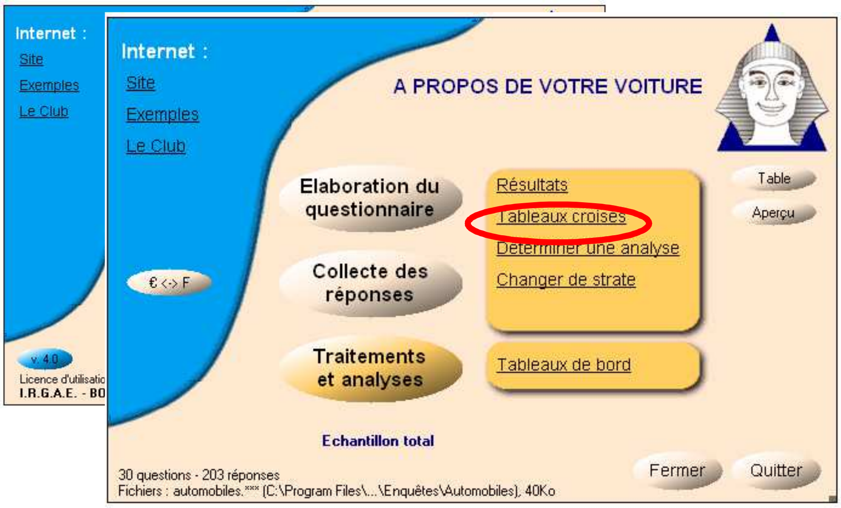
76

Khi2 et Sphinx

- Le Sphinx: logiciel de traitement et d'analyses de questionnaires => Enquêtes **quantitatives**.
- Présent en salle TC300, et dans la salle M108 du pôle.
- Gère toutes les étapes de ces enquêtes: saisie et rédaction du questionnaire, saisie des réponses, traitements et analyses.
- Parmi ces traitements, il est capable de construire des tableaux de contingence (« tris croisés », ou « tableaux croisés »), et de proposer des de pratiquer le test du Khi2 (entre autre…)
- Enquête Automobile (livrée avec le Sphinx) permet de le manipuler.
- Tris croisé entre les questions 3 (Marque) et 26 (Sexe). **Idée**: le choix de la marque d'une automobile est-il lié au sexe?

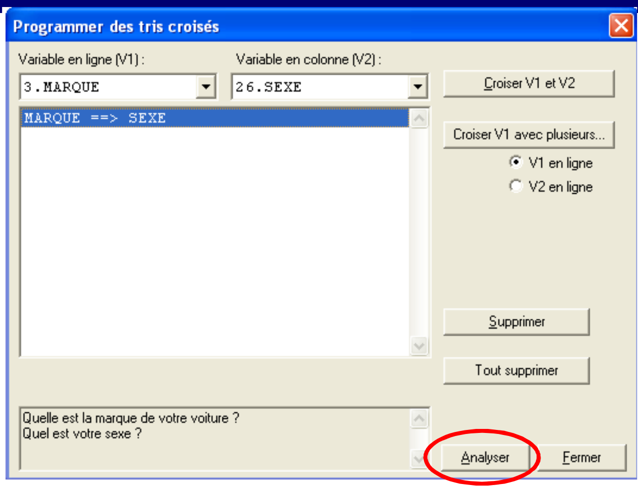
77

Khi2 et Sphinx



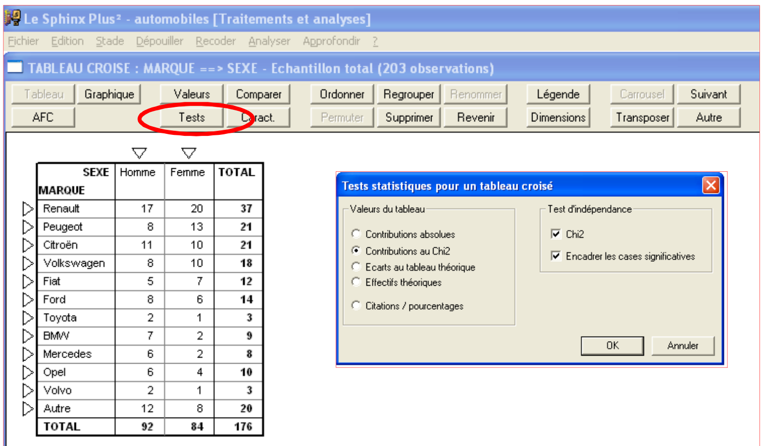
78

Khi2 et Sphinx



79

Khi2 et Sphinx



MARQUE \ SEXE	Homme	Femme	TOTAL
Renault	17	20	37
Peugeot	8	13	21
Citroën	11	10	21
Volkswagen	8	10	18
Fiat	5	7	12
Ford	8	6	14
Toyota	2	1	3
BMW	7	2	9
Mercedes	6	2	8
Opel	6	4	10
Volvo	2	1	3
Autre	12	8	20
TOTAL	92	84	176

La dépendance n'est pas significative. chi2 = 8,62, ddl = 11, 1-p = 34,32%.
Les cases encadrées en bleu (rose) sont celles pour lesquelles l'effectif réel est nettement supérieur (inférieur) à l'effectif théorique.
Attention, 9 (37.5%) cases ont un effectif théorique inférieur à 5, les règles du chi2 ne sont pas réellement applicables.

Les valeurs du tableau sont les nombres de citations de chaque couple de modalités.

80

Khi2 et Sphinx

Fichier Edition Stade Dépouiller Recoder Analyser Approfondir ?

TABLEAU CROISE : MARQUE ==> SEXE - Echantillon total (203 observations)

Tableau Graphique Valeurs Comparer Ordonner Regrouper Renommer Légende Carrousel Suivant
AFC Tests Caract. Permuter Supprimer Revenir Dimensions Transposer Autre

MARQUE \ SEXE	Homme	Femme	TOTAL
Renault	-3	+3	**37**
Peugeot	-9	+10	**21**
Citroën	+0	+0	**21**
Volkswagen	-2	+2	**18**
Fiat	-2	+3	**12**
Ford	+0	+0	**14**
Toyota	+1	-1	**3**
BMW	+12	-14	**9**
Mercedes	+9	-10	**8**
Opel	+1	-1	**10**
Volvo	+1	-1	**3**
Autre	+2	-2	**20**
TOTAL	**92**	**84**	**176**

Attention, 9 (37.5%) cases ont un effectif théorique inférieur à 5, les règles du chi2 ne sont pas réellement applicables.

La dépendance n'est pas significative. chi2 = 8,62, ddl = 11, 1-p = 34,32%.
Les cases encadrées en bleu (rose) sont celles pour lesquelles l'effectif réel est nettement supérieur (inférieur) à l'effectif théorique.
Attention, 9 (37.5%) cases ont un effectif théorique inférieur à 5, les règles du chi2 ne sont pas réellement applicables.

Les valeurs du tableau sont les pourcentages chi2 partiel / chi2 total. Le signe représente l'écart à l'indépendance.

Khi2 et Sphinx

Le Sphinx Plus² - automobiles [Traitements et analyses]

Fichier Edition Stade Dépouiller Recoder Analyser Approfondir ?

TABLEAU CROISE : MARQUE ==> SEXE - Echantillon total (203 observations)

Tableau Graphique Valeurs Comparer Ordonner Regrouper Renommer Légende Carrousel Suivant
AFC Tests Caract. Permuter Supprimer Revenir Dimensions Transposer Autre

MARQUE \ SEXE	Homme	Femme	TOTAL
France	-26	+28	**79**
Reste du monde	+21	-23	**97**
TOTAL	**92**	**84**	**176**

La dépendance est peu significative. chi2 = 2,58, ddl = 1, 1-p = 89,19%.
Les cases encadrées en bleu (rose) sont celles pour lesquelles l'effectif réel est nettement supérieur (inférieur) à l'effectif théorique.

Les valeurs du tableau sont les pourcentages chi2 partiel / chi2 total. Le signe représente l'écart à l'indépendance.

Plan Partie 1:

Autre exemple.

Questionnaire/Etude auprès de 199 clients.

Points abordés: CA, Nb Visites, Nb achats, CSP, Département, Gamme achetée, Age.

Eléments qualitatifs: CSP, Département, Gamme.

Lien CSP > CA et CSP > Gamme?

Fichier disponible sur ma page perso
(*Supports > Excel: tris_croises_vierge_200_reponses.xls*)

n° répondant	nb_visites	CA	Age	nb_achats	CSP	Département	Gamme achetée	Recodage nb visites	Recodage CA	Recodage Age	Recodage Nb achats
1	6	145	38	9	Employé	24	B	7	100	40	8
2	0	135	21	8	Cadre	24	D	1	100	20	8
3	10	32	19	10	Cadre	24	E	9	50	20	11
4	6	120	38	6	Employé	40	C	7	100	40	8
5	5	75	48	10	Inactif	33	B	4	100	50	11
6	3	175	45	12	Chef d'entreprise	33	E	4	200	50	11
7	3	275	40	0	Cadre	33	D	4	200	40	2
8	7	65	42	10	Inactif	33	B	7	50	40	11
9	4	20	18	14	Inactif	40	C	4	50	20	14
10	10	190	41	10	Employé	40	B	9	200	40	11

Fin de la partie 1…déjà!

85

Partie 2 Eléments de probabilités

Dénombrement, Notion de Variable Aléatoire, Lois Usuelles et caractéristiques

86

Statistiques Probabilités, Hasard

« *S'il est vrai que la statistique est la façon moderne de mentir, il est donc nécessaire de savoir l'utiliser pour ne pas être dupe et savoir rétablir la vérité.* » Gaston Mialaret

« *Le hasard gouverne un peu plus de la moitié de nos actions, et nous dirigeons le reste.* » Nicolas Machiavel

« *Ne rien livrer au hasard, c'est économiser du travail.* » Antoine Albalat

« *Ce que nous appelons le hasard n'est et ne peut être que la cause ignorée d'un effet connu.* »

« *Il n'y a point de hasard.* »

Voltaire

« *Le hasard dans certains cas, c'est la volonté des autres.* » .Alfred Capus

« *Toute connaissance dégénère en probabilité.* » David Hume

« ... *la valeur d'un hasard est égale à son degré d'improbabilité.* » Milan Kundera

« *Calculer la probabilité d'un événement n'a aucun sens une fois que l'on sait qu'il s'est produit* » Albert Jacquard

« *Le problème, c'est que si l'on ne prend pas de risque, on risque encore davantage.* » Erika Jong

Plan

Partie 2 – Probabilités

Probabilités?

- ✓ Probabilités: L'ensemble des règles d'après lesquelles on peut calculer les chances relatives des événements futurs.
- ✓ A l'opposé d'une approche passive des évènements
- ✓ A l'opposé d'une approche théocratique (Dieu)
- ✓ Pas du tout maîtrisées par les professionnels de la voyance...
- ✓ Etroite relation avec les statistiques. Poursuite logique.
- ✓ Etroite relation avec la théorie des jeux
- ✓ Etroite relation avec les mathématiques, dont certains résultats sont utilisés.
- ✓ Etroite relation avec les statistiques
- ✓ Objectif: limiter le hasard, donc maîtriser, connaître et limiter le risque.

89

Probabilités?

- On s'intéresse aux phénomènes et aux situations dans lesquels intervient l'incertitude.
- Si le hasard (aléa) est bien présent cela ne signifie pas que tout ce qui est entrepris est hasardeux...
- On parle d'expérience aléatoire. Si le résultat de cette expérience n'est pas connu, les conditions de son déroulement le sont.

Epreuve: action dont le résultat est aléatoire
=> Participer à une jeu chez un commerçant.

Evènement: Résultat obtenu après l'épreuve
=> Gagner le gros lot.

Probabilité d'un évènement: nb cas favorables / nb cas possibles
=> Proba gain: 1 / Nb billets

La probabilité d'un évènement X mesure donc ses chances de réalisation (P(X)

90

Probabilités?

Probabilité: nombre compris entre 0 et 1.Fréquence probable d'apparition d'un évènement.

$$Probabilité = \frac{Nombre\, de\, cas\, Favorables}{Nombre\, de\, cas\, Possibles}$$

Probabilité d'un évènement certain = 1

Probabilité d'un évènement impossible = 0

Probabilité εστ χομπρισε δανσ λ ιντερϖαλλε **[0 , 1]**

=> **Nécessite de savoir compter: Dénombrement.**

91

Probabilités?

<u>Notations:</u>

- **Ensemble des résultats possibles:** Ω
- **Evènement: Sous-ensemble de** Ω
- **Evènement noté: A, B, …C**
- **Nombre de cas possibles: Cardinal de** Ω **(card** Ω**)**
- **Nombre de cas favorables: Cardinal de A. (card A)**
- **Probabilité de réalisation d'un évènement X: P(X)**

$$Probabilité = \frac{Nombre\, de\, cas\, Favorables}{Nombre\, de\, cas\, Possibles}$$

$$Probabilité = \frac{card\, A}{card\, \Omega}$$

92

Plan

Partie 2 – Probabilités

93

Dénombrement

Puissance

$$a^n = \overbrace{a \times a \times a \times \ldots \times a}^{n \text{ fois}}$$

Exemple: **Combien de mots de 3 lettres peut-on écrire avec l'alphabet?**

26^3 = 26 x 26 x 26 = 17 576 mots différents dont la plupart n'ont aucun sens. *(aaa* par exemple)

Factorielle

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$$

Exemple: **Combien de mots de 5 lettres *différentes* peut-on écrire avec les lettres A, B, C, D, E? (le mot *abcde* est donc différent du mot *bacde*)**

5!= 5 x 4 x 3 x 2 x 1 = 120 mots dont la majorité n'a aucun sens (*abcde*)

On admet que: 0!=1

94

Dénombrement

Arrangements

***Exemple:* Combien de mots de 3 lettres différentes peut-on écrire ?**

=> 15600 mots différents dont la plupart n'ont aucun sens. *(abc)*

L'ordre des lettres intervient ici. *abc* est un mot différent de *bac*

$$A_{26}^{3} = \frac{26!}{(26-3)!}$$

$$A_{26}^{3} = 26 \times 25 \times 24$$

$$A_{n}^{p} = \frac{n!}{(n-p)!}$$

$$A_{n}^{p} = n \times (n-1) \times (n-2) \times ... \times (n-p+1)$$

Dénombrement

Combinaisons

***Exemple:* Combien de groupes de 3 lettres différentes *et non ordonnées* peut-on écrire?**

=> 2600 groupes. L'ordre n'intervient pas. Le groupe *abc* est identique au groupe *bac*, *acb*,…

$$C_{26}^{3} = \frac{A_{26}^{3}}{3!}$$

$$C_{26}^{3} = \frac{26 \times 25 \times 24}{3 \times 2 \times 1}$$

$$C_{26}^{3} = \frac{26!}{(26-3)!3!}$$

$$C_{n}^{p} = \frac{A_{n}^{p}}{p!}$$

$$C_{n}^{p} = \frac{n!}{(n-p)!p!}$$

Plan

Notion de probabilité

$$Probabilit\acute{e} = \frac{Nombre\,de\,cas\,Favorables}{Nombre\,de\,cas\,Possibles}$$

$$Probabilit\acute{e} = \frac{card\,A}{card\,\Omega}$$

$$=> Pr(\Omega) = 1 \text{ et } Pr(\emptyset) = 0$$

Probabilités totales

Cas des évènements incompatibles.

A et B sont des évènements incompatibles si A se produit B ne se produit pas. On a alors:

$$P(AetB) = P(A \cap B) = 0$$

$$P(AouB) = P(A \cup B) = P(A) + P(B) puisque P(A \cap B) = 0$$

***Exemple*: Probabilité de tirer 1 et 6 sur un seul jet de dé.**
P(1 et 6) = 0

***Exemple:* Probabilité de tirer 1 ou 6 sur un seul jet de dé.**
P(1 ou 6) = P(1) + P(6) = 1/6 + 1/6 = 1/3

Notion de probabilité

Probabilités totales

Cas des évènements incompatibles.

A et B sont des évènements incompatibles si A se produit B ne se produit pas. On a alors:

$$P(AetB) = P(A \cap B) = 0$$

$$P(AouB) = P(A \cup B) = P(A) + P(B) puisque P(A \cap B) = 0$$

***Exemple*: Probabilité de tirer 1 et 6 sur un seul jet de dé.**
P(1 et 6) = 0

***Exemple*: Probabilité de tirer 1 ou 6 sur un seul jet de dé.**
P(1 ou 6) = P(1) + P(6) = 1/6 + 1/6 = 1/3

99

Notion de probabilité

Probabilités totales

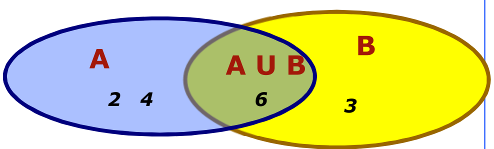
Cas des évènements compatibles.

Tenir compte de leur intersection puisque P(A et B) n'est alors pas nul.

$$P(AouB) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(AetB) = P(A \cap B)$$

***Exemple* avec le lancer de dé.**
A: Résultat est pair.
B: Résultat est un multiple de 3.



Donc A={2,4,6} et B={3,6}. On a:

(A et B)={6} et (A ou B) = {2,3,4,6}

P(A ou B) = P (2,3,4,6) = P(A) + P(B) –P(A ou B)

= P(2,4,6) +P(3,6) –P(6)

= 3/6 + 2/6 – 1/6 = 4/6

= P(2,3,4,6)

100

Notion de probabilité

$$Pr\{A|E\} = \frac{Pr\{AetE\}}{Pr\{E\}}$$

Probabilités Composées - Exemple

Restriction sur l'ensemble des résultats possibles => Constructions possibles de nouvelles probabilités. Exemple lancer de dé:Ω={1,2,3,4,5,6}

On lance le dé en sachant que le résultat est pair. L'ensemble des résultats possible devient alors: E={2,4,6} qui est un évènement de Ω.

On s'intéresse à l'évènement A={6}.

Probabilité que A se produise, sachant que le résultat du dé est pair? P(A) sachant que le résultat est pair.

Intuitivement (Cas très simple)

Parmi les possibilités de E, on retient celle(s) qui corresponde(nt) à A: le 6

Donc P(A)=1/3.

Notion de probabilité

Probabilités Composées - Indépendance

Soit un espace probabilisé. Soit A, et E deux évènements , et P(E) ≠0.

On a: $P(A/E) = \frac{P(A \cap E)}{P(E)}$

Donc : $P(A \cap E) = P(A/E)xP(E)$

Donc ici:

P(E)= P(obtenir un résultat pair) = 3/6 => 1/2

P(A)=1/6

P(A et E)= $P(A \cap E) = \frac{1}{6}$ ***(Obtenir un 6 et un nombre pair). (=P(A))***

$P(A/E) = \frac{P(A \cap E)}{P(E)} = \frac{\frac{1}{6}}{\frac{1}{2}}$

1

Notion de probabilité

Probabilités Composées - Indépendance

Deux évènements A et B sont indépendants si et seulement si:

P(A/B)=P(A) ou P(B/A)=P(B).

On a alors: $P(A \cap B) = P(A)xP(B)$

Dans notre cas A (Obtenir le 6) et E (résultat pair) sont-ils indépendants?

P(E)=1/2 et P(A)= 1/6

On a vu que $P(A \cap E) = \frac{1}{6}$

Ici

- P(A/E)=1/3 => 1/6 (=P(A))
- P(A) x P(E)=1/6 x 1/2 => 1/3

A et E sont donc dépendants.

Notion de probabilité

Probabilités Composées – Autre cas

❖ Dans une famille de 2 enfants quelle est la probabilité d'avoir 2 filles (Evènement A)?

Ω ={(G,G), (G,F), (F,G), (F,F)} *avec F: Fille et G: garçon*. 4 cas possibles.

A={(F,F)} =>1 seul cas favorable. Donc P(A)=1/4.

❖ Probabilité composée ou conditionnelle: renseignement complémentaire.

Ici: dans la famille il y a au moins une fille (B).

B= {(G,F), (F,G), (F,F)} . 3 cas possibles seulement.

1 seul cas favorable. Donc P(A/B)=1/3

Avec la formule:

On a P(B)=3/4. $P(A \cap B) = \frac{1}{4}$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Probabilités – Fréquences: différence de nature.

On a lancé 30 fois un dé. Les résultats obtenus sont les suivants:

x_i	1	2	3	4	5	6	Total
n_i	5	4	5	7	6	3	30
f_i	0,17	0,13	0,17	0,23	0,20	0,10	1

Série statistique à un caractère (x_i qui est discret). Les effectifs mesurent les résultats qui ont été obtenus pour chaque valeur de x_i. Les fréquences f_i expriment ceux-ci de manière relative (Le 6 est donc sorti dans 10% des cas.)
On peut calculer la moyenne (3,47), la variance (2,52) et l'écart type (1,59).

=> Eléments qui reflètent une expérience passée dont les résultats sont certains et connus à présent.

On va lancer une nouvelle fois ce dé. Quel résultat pouvons-vous obtenir?

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Probabilité. On ne connaît pas à l'avance le résultat. On ne peut qu'évaluer les différents résultats possible et les probabilités associées.

=> Expérience à venir dont les résultats sont incertains et inconnus à présent.

Plan

Partie 2 – Probabilités

Notion de Variable aléatoire

- **Variable dont les valeurs possibles sont connues et auxquelles on peut rattacher une probabilité.**
- **Si le nombre de valeurs possibles est fini (donc limité) on parlera de variable aléatoire discrète. Dans le cas contraire (Intervalle) on parlera de variable aléatoire continue.**
- **Les variables aléatoires sont notées avec des majuscules (X). Ne pas confondre avec leur(s) réalisation(s) notées avec des minuscules (x)**
- **Des valeurs typiques (espérance mathématique, variance) peuvent être calculées pour ces variables aléatoires.**
- **Des représentations graphiques peuvent être associées à celles-ci: fonctions de distribution (ou densité de probabilité) et fonctions de répartition**

107

Variables aléatoires discrètes

On lance un dé. On s'intéresse au numéro qui apparaît.

X, variable aléatoire, représente donc le numéro obtenu à ce lancer et peut prendre les valeurs 1, 2, 3, 4, 5, et 6. Le nombre de valeurs possibles est donc fini. Cette variable aléatoire est discrète.

Loi de probabilité de X:
(x: réalisation possible de X, p: probabilité associée)

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

(Equiprobabilité)

108

Variables aléatoires discrètes: Valeurs typiques

- **Espérance mathématique**

Loi de X

X: lancer d'un dé

x	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Espérance Mathématique E(X) => Caractéristique de tendance centrale

$$E(X)=\sum_{i=1}^{n} x_i\, p_i$$

Soit ici $$E(X)=\frac{1}{6}(1+2+3+4+5+6)=\frac{21}{6}=\frac{7}{2}$$

$\boxed{E(X) = 3,5}$

- **Propriétés**

E(X + a) = E(X) + a
E(aX) = a E(X)
E(X-E(X))=0

109

Variables aléatoires discrètes: Valeurs typiques

- **Variance – Ecart type**

Loi de X

X: lancer d'un dé

x	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Variance var(X), Ecart type σ(X) => Caractéristiques de dispersion

$$Var(X)=E(X-E(X))^2 \qquad Var(X)=E(X^2)-E(X)^2 \qquad Var(X)=\sum_{i=1}^{n} x^2_i\, p_i - E(X)^2$$

$$\sigma(X)=\sqrt{Var(X)}$$

Soit ici

$$Var(X)=\frac{1}{6}[(1-3,5)^2+(2-3,5)^2+\ldots+(6-3,5)^2]=2,92$$

$$Var(X)=\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)-3,5^2=2,92$$

$$\sigma(X)=\sqrt{2,92)}=1,71$$

- **Propriétés**

Var(X + a) = Var(X)
Var(aX) = a² Var(X)

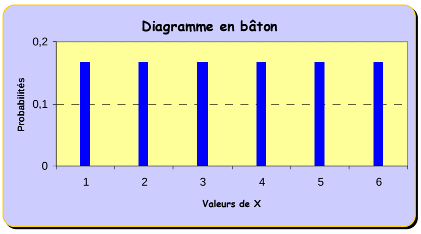
σ(X+a) = σ(X)
σ(aX) =|a| σ(X)

110

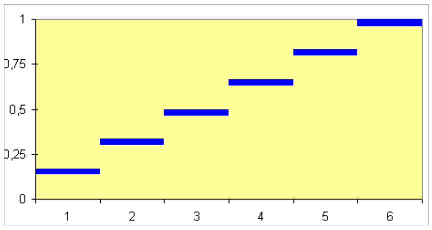
Représentations graphiques

x	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Fonction de distribution
(Diagramme à bâtons)



Fonction de répartition
(Cumuls)



Variables aléatoires: autre exemple

Anatole et Bérangère jouent au <u>jeu suivant</u> avec un dé:

Si le résultat est pair Anatole donne 1€ à Bérangère
Si le 1 ou le 5 sort Bérangère donne 3€ à Anatole.
Si le 3 sort la partie est nulle.

Le jeu est-il équitable?

Jeu équitable => Espérance de gain égale pour les deux personnes

$\Rightarrow$ Espérance mathématique nulle.

2

Loi de probabilité (Bérangère)

x	-3	0	1
p	2/6	1/6	3/6

2

$$E(X) = (-3)\times\frac{2}{6} + 0\times\frac{1}{6} + 1\times\frac{3}{6} = -0{,}5$$

0

$$Var(X) = [(-3)^2\times\frac{2}{6} + 0 + 1^2\times\frac{3}{6}] - (-0{,}5)^2$$

$$Var(X) = (\frac{18}{6} + \frac{3}{6}) - 0{,}5^2 = 3{,}25$$

$$\sigma(X) = \sqrt{3{,}25} = 1{,}80$$

Bérangère a une espérance de gain négative => elle est donc désavantagée par ce jeu non équitable.

Plan

Partie 2 – Probabilités

113

Lois de probabilité

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Recensement de tous les résultats possibles…
=> Pas toujours faisable
Calcul des probabilités associées
=> Pas toujours utile
Variables Continues
=> Impossible à réaliser

Types de situations connues qui ont permis de dégager et d'étudier des lois de probabilité qui apportent des réponses pertinentes tout en évitant de construire un tel tableau:
- ✓ Loi Binomiale, Loi de Poisson (Discrètes)
- ✓ Loi Normale (Continue)

114

Loi de Bernoulli

On jette un dé une seule fois et on souhaite obtenir le 6. On s'intéresse à la variable aléatoire suivante:

X=1 => Succès: le 6 est sorti.
X=0 => Echec: un autre numéro que le 6 apparaît.

Loi de probabilité de X

x_i	0	1
p_i	5/6	1/6

- **E(X)= (0 x 5/6) + (1 x 1/6) = 1/6**
  => Probabilité du succès
- **Var(X)= (0² x 5/6) + (1² x 1/6) – (1/6)² =1/6 – (1/6)² = 5/36**
  => Probabilité du succès x Probabilité de l'échec

$$\sigma(X) = \sqrt{5/36}$$

- ✓ **Expérience avec un seul essai**
- ✓ **Deux résultats possibles: succès ou échec**
- ✓ **Probabilité de succès et d'échec connues.**

Loi Binomiale

- **On jette un dé plusieurs fois (*n* fois). Les résultats des jets sont indépendants.**
- **Pour chaque jet on a deux possibilités:**
  Succès: le 6 est sorti.
  Echec: un autre numéro que le 6 apparaît.
- **Pour chaque jet on connaît:**
  La probabilité de succès: *p*=1/6
  La probabilité d'échec: q=5/6 *(= 1 – p)*
- **On s'intéresse au nombre (k) de succès. X représente cette variable aléatoire.**

X suit une loi Binomiale B(n,p). On démontre que:

$$P(X=k) = C_n^k p^k q^{n-k}$$

Valeurs typiques:

$$E(X) = n \times p$$

$$Var(x) = n \times p \times q$$

$$\sigma(X) = \sqrt{n \times p \times q}$$

Loi Binomiale

On jette un dé plusieurs fois (*4* fois). Les résultats des jets sont indépendants.

Pour chaque jet on a deux possibilités:

Succès: le 6 est sorti.

Echec: un autre numéro que le 6 apparaît.

Pour chaque jet on connaît:

La probabilité de succès: *p*=1/6

La probabilité d'échec: q=5/6 *(= 1 – p*)

On s'intéresse au nombre (k) de succès. X représente cette variable aléatoire. Bien entendu k ≤ n.

Valeurs typiques:

$$E(X) = 4 \times \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

$$Var(x) = 4 \times \frac{1}{6} \times \frac{5}{6} = \frac{20}{36} = \frac{5}{9}$$

$$\sigma(X) = \sqrt{\frac{20}{36}}$$

$$P(X=0) = C_4^0 \frac{1}{6}^0 \frac{5}{6}^{4-0} = 0{,}482$$

$$P(X=3) = C_4^3 \frac{1}{6}^3 \frac{5}{6}^{4-3} = 0{,}015$$

117

Loi Binomiale

Tableau de cette expérience

Nb Essais	**n=**	**4**
Proba succès	**p=**	**1/6**
Proba Echec	**q=**	**5/6**

xi	0	1	2	3	4	Somme probas
pi	0,482	0,386	0,116	0,015	0,001	1

E(X)	= 0 x 0,482 + … + 4 x 0,001 =	2/3	= 4 x (1/6)	= n x p
Var(X)	= (0² x 0,482 + … + 4² x 0,001) - (2/3)² =	5/9	= 4 x (1/6) x (5/6)	= n x p x q

Résultats appréciables quand le nombre d'essais est importants

Il est fréquent, et souvent aisé, de se ramener à des situations de type binomial pour des caractères quantitatifs et mais aussi qualitatifs (*Le client achète ou n'achète pas*)

Loi qui peut être supplée par une loi discrète (Loi de Poisson) continue (Loi Normale) quand les situations le permettent.

=> Important

118

Loi Binomiale

Loi qui est tabulée…Moins intéressant depuis l'apparition des ordinateurs personnels…sauf pour les intervalles P(k1<X<k2)

$P(X = k)$

Table Loi Binomiale (Extrait)

n= 10

k \p	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
0	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
1	0,3151	0,3874	0,3474	0,2684	0,1877	0,1211	0,0725	0,0403	0,0207	0,0098
2	0,0746	0,1937	0,2759	0,3020	0,2816	0,2335	0,1757	0,1209	0,0763	0,0439
3	0,0105	0,0574	0,1298	0,2013	0,2503	0,2668	0,2522	0,2150	0,1665	0,1172
4	0,0010	0,0112	0,0401	0,0881	0,1460	0,2001	0,2377	0,2508	0,2384	0,2051
5	0,0001	0,0015	0,0085	0,0264	0,0584	0,1029	0,1536	0,2007	0,2340	0,2461
6	0,0000	0,0001	0,0012	0,0055	0,0162	0,0368	0,0689	0,1115	0,1596	0,2051
7	0,0000	0,0000	0,0001	0,0008	0,0031	0,0090	0,0212	0,0425	0,0746	0,1172
8	0,0000	0,0000	0,0000	0,0001	0,0004	0,0014	0,0043	0,0106	0,0229	0,0439
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0016	0,0042	0,0098
10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010

$P(X \leq k)$

Table Loi Binomiale Cumulative (Extrait)

n= 10

k \p	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
0	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
1	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,0860	0,0464	0,0233	0,0107
2	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2616	0,1673	0,0996	0,0547
3	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5138	0,3823	0,2660	0,1719
4	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7515	0,6331	0,5044	0,3770
5	1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9051	0,8338	0,7384	0,6230
6	1,0000	1,0000	0,9999	0,9991	0,9965	0,9894	0,9740	0,9452	0,8980	0,8281
7	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9952	0,9877	0,9726	0,9453
8	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9983	0,9955	0,9893
9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

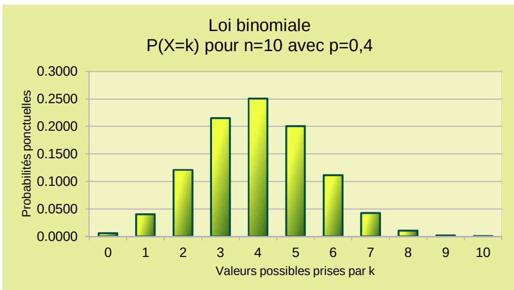
Tables en fin de document

Loi Binomiale

$$P(X = k) = C_n^k p^k q^{n-k}$$

n= 10

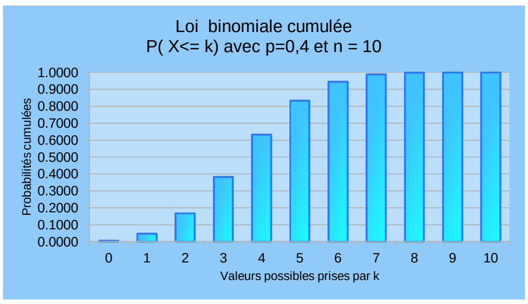
k \ p	0,4
0	0,0060
1	0,0403
2	0,1209
3	0,2150
4	0,2508
5	0,2007
6	0,1115
7	0,0425
8	0,0106
9	0,0016
10	0,0001
Total	1



$P(X \leq k)$

n= 10

k \ p	0,4
0	0,0060
1	0,0464
2	0,1673
3	0,3823
4	0,6331
5	0,8338
6	0,9452
7	0,9877
8	0,9983
9	0,9999
10	1,0000



Plan

Partie 2 – Probabilités

121

Loi de Poisson

Loi des phénomènes rares.

Exemple: ***nombre de personnes faisant la queue à la caisse d'un supermarché pendant quelques minutes.***

Loi dont les résultats dépendent d'un paramètres Λαμβδα qui est un nombre positif.

Si X est une variable aléatoire qui obéit à une loi de Poisson, X prend donc les valeurs 0, 1, 2, …(*k* de manière générale)

On démontre que:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Valeurs typiques: $E(X) = \lambda$ $Var(X) = \lambda$

Les phénomènes étudiés étant peu nombreux, leur dispersion est donc de fait importante.

122

Loi de Poisson

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$$

Loi des phénomènes rares.

<u>Exemple:</u> On admet que le nombre de clients qui attendent à une caisse d'un supermarché pendant 5 minutes obéit à une loi de Poisson de paramètre Λαμβδα=4.

On se propose d'affecter du personnel si nécessaire, en fonction du nombre effectif de personnes qui attendent.

Probabilité qu'il n'y ait aucun client: $P(X=0)=e^{-4}\frac{4^0}{0!}=e^{-4}=0{,}0183$

Probabilité qu'il y ait 5 clients: $P(X=5)=e^{-4}\frac{4^5}{5!}=0{,}1563$

Cette loi est tabulée (diapo suivante) ce qui permet de retrouver ces résultats très vite (1 tableau)

Probabilité qu'il y ait moins de 4 (<4) clients :
P(X<4)= P(X ≤ 3) = P(X=0) + p(X=1) + P(X=2) + P(X=3).

La table cumulée (2ème tableau diapo suivante) donne le résultat : P(X ≤ 3)= 0,4335

Loi de Poisson

Loi qui est tabulée…Moins intéressant depuis l'apparition des ordinateurs personnels… sauf pour les intervalles: P(k1<X<k2)

Table de la loi de Poisson (Extraits)

$P(X=k)$

k \ λ	1	2	3	4	5	6	7	8	9	10
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0573	0,0337	0,0189
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
7	0,0001	0,0034	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171	0,0901

Table de la loi de Poisson Cumulative (Extraits)

$P(X\leq k)$

k \ λ	1	2	3	4	5	6	7	8	9	10
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000
1	0,7358	0,4060	0,1991	0,0916	0,0404	0,0174	0,0073	0,0030	0,0012	0,0005
2	0,9197	0,6767	0,4232	0,2381	0,1247	0,0620	0,0296	0,0138	0,0062	0,0028
3	0,9810	0,8571	0,6472	0,4335	0,2650	0,1512	0,0818	0,0424	0,0212	0,0103
4	0,9963	0,9473	0,8153	0,6288	0,4405	0,2851	0,1730	0,0996	0,0550	0,0293
5	0,9994	0,9834	0,9161	0,7851	0,6160	0,4457	0,3007	0,1912	0,1157	0,0671
6	0,9999	0,9955	0,9665	0,8893	0,7622	0,6063	0,4497	0,3134	0,2068	0,1301
7	1,0000	0,9989	0,9881	0,9489	0,8666	0,7440	0,5987	0,4530	0,3239	0,2202

Tables en fin de document

<u>Avec les données précédentes</u>: Probabilité d'avoir entre 2 et 5 clients (compris).
P(X compris entre 2 et 5)=P(X=2) + … + P(X=5)
P(X<=5)-P(X<=1) = 0,7851 – 0,0916 = 0,6936

Convergence de la loi Binomiale vers la loi de Poisson

La loi Binomiale B(n,p) converge vers la loi de Poisson avec lαμβδα=np quand n devient important.

Les condition d'approximation sont les suivantes:

- ✓ **n>30 (n *grand*)**
- ✓ **p<=0,1 ou p ≥ 0,9 (p *très petit*)**
- ✓ **np<5**

Exemple: X suit une loi binomiale <u>B(40; 0,05)</u>

On a donc n=40 donc suffisamment grand; p <0,1, et np=40 x 0,05 = 2.

Donc X peut donc être approximée par une variable aléatoire Y qui obéit à la loi de Poisson suivante: P(2).

On a alors:

$$P(X=3) = C_{40}^{3} 0{,}05^{3} \times 0{,}95^{40-3} = 0{,}185 \qquad P(Y=3) = e^{-2}\frac{2^{3}}{3!} = 0{,}181$$

125

Convergence de la loi Binomiale vers la loi de Poisson

La société DOVITEL constate que 0,75 % des milliers de contrats d'abonnements passés avec ses clients finissent en contentieux 6 mois après leur signature.
Quelle est la probabilité pour qu'aucun des 400 derniers ne soit concerné ? Qu'il y en ait plus d'un ? 5 ? moins de 3 ? Entre 4 et 7 ?

Loi exacte : B(400 ; 0,75%), avec E(X) = 400 x 0,75 % = 3 et Var(X) = 800 x 0,75 % x 99,25 % = 2,977, soit presque 3.

On aurait donc : $P(X=0) = C_{400}^{0} 0{,}75\%^{0} 99{,}25\%^{400-0}$

<u>On a ici:</u> np < 5, n > 30, p < 0,1.

Cette loi exacte peut donc **approximer cette loi exacte par une loi de Poisson de paramètre Lambda = 3**

$$P(X=0) = e^{-3}\frac{3^{0}}{0!} = e^{-3} = 0{,}0498$$

P(X ≥ 1) = 1 – P(X=0) = 1 – 0,0498 = 0,9502

$$P(X=5) = e^{-3}\frac{3^{5}}{5!} = 0{,}1008$$

P(X<3) = P(X ≤ 2) = P(X=0) + P(X=1) + P(X=2) = 0,4232

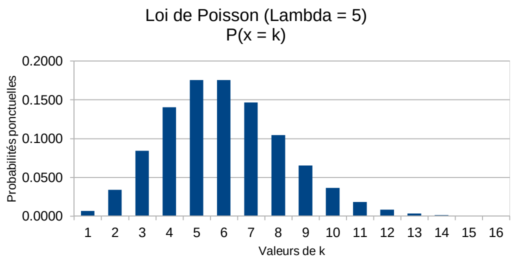
P(4 ≤ X ≤ 7) = P(X ≤ 7) – P(X < 4) = P (X ≤ 7) – P (X ≤ 3) = 0,9881 – 0,6472 = 0,3409

126

Loi de Poisson

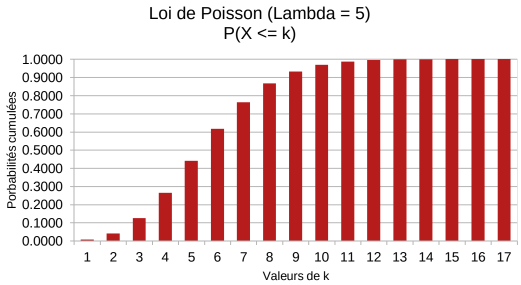
$P(X = k)$
5^k

K \ lambda	5
0	0,0067
1	0,0337
2	0,0842
3	0,1404
4	0,1755
5	0,1755
6	0,1462
7	0,1044
8	0,0653
9	0,0363
10	0,0181
11	0,0082
12	0,0034
13	0,0013
14	0,0005
15	0,0002
16	0,0000

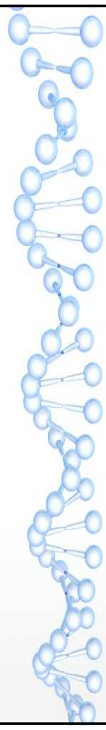


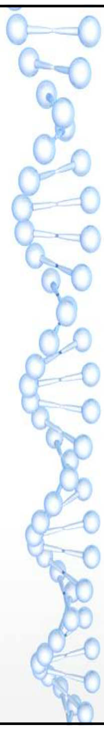
$P(X \leq k)$

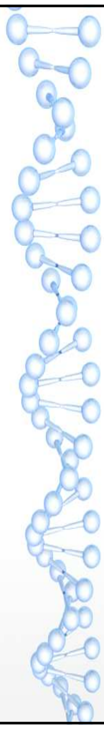
K \ lambda	5
0	0,0067
1	0,0404
2	0,1247
3	0,2650
4	0,4405
5	0,6160
6	0,7622
7	0,8666
8	0,9319
9	0,9682
10	0,9863
11	0,9945
12	0,9980
13	0,9993
14	0,9998
15	0,9999
16	1,0000

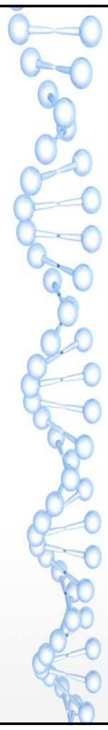


127









Plan

Plan Loi Normale

Généralités et caractéristiques

Utilisation de la table avec N(0,1)

Utilisation de la table Cas Général

Autre utilisation de la table

Approximation des lois discrètes par la loi Normale

Adéquation d'une série statistique à une loi Normale

135

Loi Normale ou Laplace Gauss

- Loi de probabilité continue et tabulée.
- Une telle loi est caractérisée par sa moyenne et son écart-type.
- L'évènement ponctuel est sans intérêt (donc considéré comme nul).
- La table concerne un cas particulier: moyenne=0, et écart-type=1. On parle de loi Normale Centré Réduite.
- On peut donc toujours se ramener à ce cas particulier par un simple changement de variable.
- La table renvoie des probabilités cumulées en fonction d'un paramètre.
- Le modèle mathématique ne nous est pas utile. Il nous suffit de connaître les propriétés de base et la manipulation de la table.
- On désigne par X, Y, Z ou W… un phénomène obéissant à une loi normale quelconque.
- Et par T un phénomène obéissant à la loi normale tabulée (centrée réduite)
- Une probabilité est comprise entre 0 (évènement impossible) et 1 (évènement certain)
- Apprendre à utiliser la table est donc important pour la suite.

136

Loi Normale

- **Loi d'une variable aléatoire X dont la fonction de densité est la suivante dans le cas général, en fonction de deux paramètres: m et σ**

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

- **Cas particulier de la loi normale centrée réduite: m=0 et σ=1**

$$f(x) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{x^2}{2}}$$

137

Loi Normale

- **Calcul de probabilité dans le cas général,** en fonction de deux paramètres: m et σ

$$P(u < X < v) = \int_u^v \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

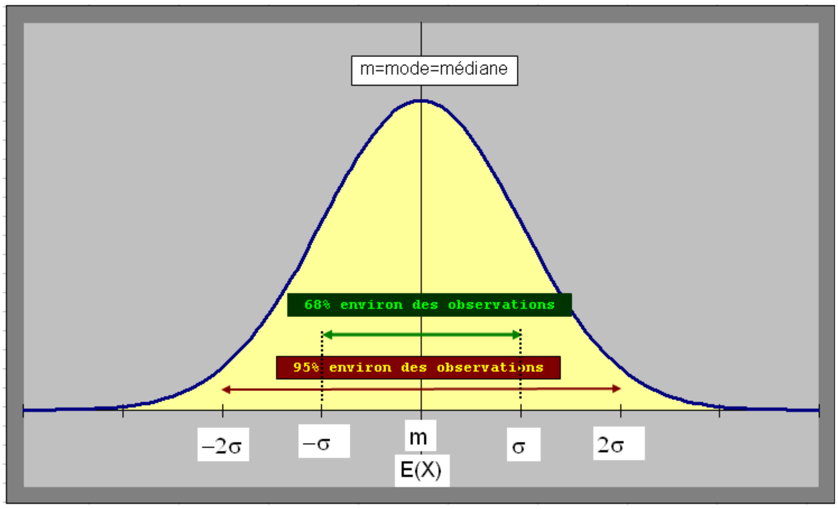
- **Cas particulier de la loi normale centrée réduite: m=0 et σ=1**

$$P(u < X < v) = \int_u^v \frac{1}{\sqrt{2\Pi}} e^{-\frac{x^2}{2}} dx$$

…Ou alors utilisation de la table…

138

Loi Normale - Caractéristiques



139

Plan Loi Normale

Généralités et caractéristiques

Utilisation de la table avec N(0,1)

Utilisation de la table Cas Général

Autre utilisation de la table

Approximation des lois discrètes par la loi Normale

Adéquation d'une série statistique à une loi Normale

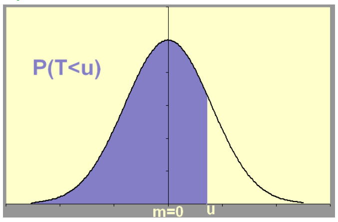
140

Loi Normale: utilisation de la table (Page).

Table de la loi normale centrée réduite

υ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

- La table renvoie p(T<u) avec u positif ou nul.
- u se lit en colonne et en ligne
- Ainsi P(T<1.35)=0.91149
- Et P(T<0.52)=0.69847
- P(T<3.15)=1
- P(T<0)=0.5=P(T>0)



La table découle du modèle mathématique ainsi qu'une courbe. Une probabilité recherchée (P(T<u) est égale à la surface sous la courbe. La surface totale sous la courbe est donc égale à 1. La fonction associée est paire, la courbe est donc symétrique autour de la moyenne (0 pour la table).

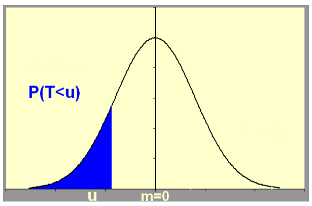
141

Loi Normale: utilisation de la table.

Table de la loi normale centrée réduite

υ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

- La table renvoie p(T<u) avec u positif ou nul. Autres cas?
- u>0, mais P(T>u)?
- P(T>1,25)=
- 1-P(T<1.25)=1-0.89435=0.10565
- u<0, mais P(T<u)?
- P(T<-0.24)=
- 1-P(T<0.24)=1-0.59483=0.40517



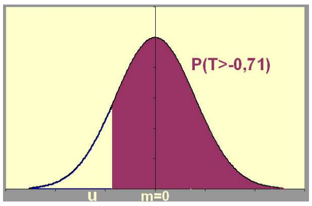
142

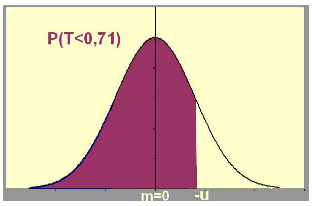
Loi Normale: utilisation de la table.

- La table renvoie p(T<u) avec u positif ou nul. Autres cas?
- u<0, mais P(T>u)?
- P(T>-0.71)=P(T<0,71)=0,76115

Table de la loi normale centrée réduite

υ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861





143

Loi Normale: utilisation de la table.

- La table renvoie p(T<u) avec u positif ou nul. Autres cas?
- Intervalle borné?

P(-0.43<T<0,43)

=P(T<0,43)-P(T<-0,43)

- =P(T<0,43)-[1-P(T<0,43)]

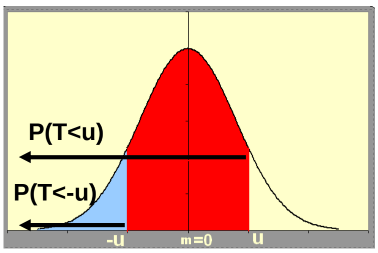
=2P(T<0,43)-1

=2 x 0,66640-1=0,3328

Table de la loi normale centrée réduite

υ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

Principe identique si l'intervalle n'est pas symétrique, mais on lit 2 valeurs dans la table.



144

Plan Loi Normale

Généralités et caractéristiques
Utilisation de la table avec N(0,1)
Utilisation de la table Cas Général
Autre utilisation de la table
Approximation des lois discrètes par la loi Normale
Adéquation d'une série statistique à une loi Normale

145

Loi Normale: Cas général

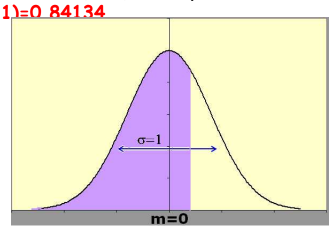
u	0	0.01	0.02
0	0.50000	0.50399	0.50798
0.1	0.53983	0.54380	0.54776
0.2	0.57926	0.58317	0.58706
0.3	0.61791	0.62172	0.62552
0.4	0.65542	0.65910	0.66276
0.5	0.69146	0.69497	0.69847
0.6	0.72575	0.72907	0.73237
0.7	0.75804	0.76115	0.76424
0.8	0.78814	0.79103	0.79389
0.9	[illegible]	0.81859	0.82121
1	0.84134	0.[illegible]4375	0.84614

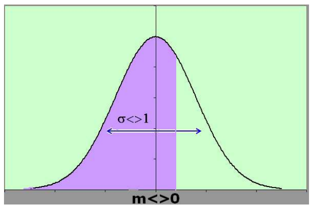
- La table concerne un cas bien particulier. T suit une loi normale de moyenne=0, et d'écart type=1. Or m et σ peuvent prendre une infinité de valeurs.
- Bien souvent le phénomène étudié (X) suit une loi normale: N(m, σ), avec m non nul et σ différent de 1. Exemple m=5 et σ=3.
- P(X<8)=?
- Appliquer le changement de variable suivant pour pouvoir utiliser la table (T loi normale centrée réduite):

$$T = \frac{X-m}{\sigma}$$

$$P(X < 8) = P(\frac{X-5}{3} < \frac{8-5}{3}) = P(T < 1)$$

- Rechercher P(X<8) est équivalent en terme de résultat à rechercher P(T<1). Et P(T<1)=0.84134





146

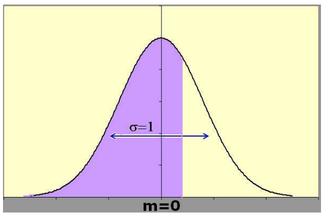
Loi Normale: Cas général

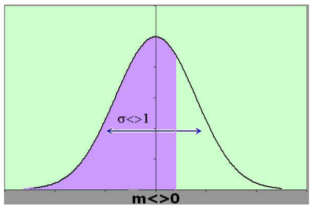
υ	0	0.01	0.02
0	0.50000	0.50399	0.50798
0.1	0.53983	0.54380	0.54776
0.2	0.57926	0.58317	0.58706
0.3	0.61791	0.62172	0.62552
0.4	0.65542	0.65910	0.66276
0.5	0.69146	0.69497	0.69847
0.6	0.72575	0.72907	0.73237
0.7	0.75804	0.76115	0.76424
0.8	0.78814	0.79103	0.79389
0.9	[illegible]	0.81859	0.82121
1	0.84134	[illegible]	0.84614

- **Exemple m=5 et σ=3.** P(X>2)?
- **Appliquer le changement de variable suivant pour pouvoir utiliser la table (T loi normale centrée réduite):** $T = \frac{X - m}{\sigma}$

$$P(T > 1) = 1 - P(T < 1) = 1 - 0{,}84134 = 0{,}15866$$

- **Rechercher P(X>2) est équivalent en terme de résultat à rechercher P(T>1). Et P(T>1)=0,15866**





147

Plan Loi Normale

Généralités et caractéristiques
Utilisation de la table avec N(0,1)
Utilisation de la table Cas Général
Autre utilisation de la table
Approximation des lois discrètes par la loi Normale
Adéquation d'une série statistique à une loi Normale

148

Loi Normale: Autre Utilisation de la table:

- **T loi Normale centrée réduite N(0;1)**

Table de la loi normale centrée réduite

υ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	[illegible]	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	[illegible]	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	[illegible]	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	[illegible]	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	[illegible]	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	[illegible]	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	[illegible]	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	[illegible]	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	[illegible]	0.81327
0.9	0.81594	0.81859	0.82121	[illegible]	0.82639	[illegible]	0.83147	0.83398	[illegible]	0.83891
1	[illegible]	[illegible]	[illegible]	[illegible]	0.85083	[illegible]	0.85543	0.85769	[illegible]	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	[illegible]	[illegible]	[illegible]
1.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	[illegible]	[illegible]	[illegible]
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

- Valeur de u sachant que P(T<u)=0.90 ?
- La table va donc s'utiliser de manière différente:
- 1°) Rechercher la probabilité (la plus proche) à l'intérieur de la table.
- 2°) En déduire la valeur de u.
- Donc u=1,28
- Autre exemple: Valeur de u sachant que P(T<u)=0.85 ?
- Donc u=1,04

149

Loi Normale: Utilisation de la table:

- **T loi Normale générale N(m;σ)**

Table de la loi normale centrée réduite

υ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	[illegible]	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	[illegible]	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	[illegible]	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	[illegible]	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	[illegible]	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	[illegible]	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	[illegible]	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	[illegible]	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	[illegible]	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	[illegible]	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	[illegible]	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	[illegible]	[illegible]	0.88298
1.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.89973	[illegible]
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	[illegible]	[illegible]	[illegible]
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

- X suit une loi normale N(100;20)
- Valeur de b sachant que P(X<b)=0,90 ?
- Changement de variable pour se ramener à T:

$$T = \frac{X - m}{\sigma} = \frac{X - 100}{20}$$

- P(X<b)=0,90 => P(T< (b-100)/20)=0,9
- Après lecture dans la table:
- (b-100)/20= 1,28
- b-100 = 20 x 1,28
- b= 100 + 25,6 = 125,6

150

Plan Loi Normale

Généralités et caractéristiques
Utilisation de la table avec N(0,1)
Utilisation de la table Cas Général
Autre utilisation de la table
Approximation des lois discrètes par la loi Normale
Adéquation d'une série statistique à une loi Normale

Approximation d'une loi binomiale par une loi normale

Soit une variable aléatoire X qui suit une loi binomiale B(n , p)
X peut être approximé par Y qui suit une loi normale N(np, $npq^{0.5)}$
Si $n \geq 15$ et p et q non voisins de 0.
Dans les faits dès que np>15 et nq>15 cela suffit.

Exemple: X suit B(50; 0,4).
On a n=50, p=0,4 et q=0,6.
np= 50 x 0.4 =20, et nq= 50 x 0.6 = 30.
npq= 50 x 0.4 x 0.6 =12 pour la variance, ce qui donne 3,46.
Donc X peut être approximée par Y qui suit N(20; 3,46)

Approximation d'une loi Poisson par une loi Normale

Soit une variable aléatoire X qui suit une loi de Poisson P(λ)
X peut être approximé par Y qui suit une loi normale N(λ, $\lambda^{0.5)}$
Si λ est grand
En pratique si $\lambda > 25$

Exemple: **X suit P(36).**
On a $\lambda > 25$..
Donc X peut être approximée par Y qui suit N(36; 6)

153

Plan Loi Normale

Généralités et caractéristiques
Utilisation de la table avec N(0,1)
Utilisation de la table Cas Général
Autre utilisation de la table
Approximation des lois discrètes par la loi Normale
Adéquation d'une série statistique à une loi Normale

154

Adéquation série statistique à une loi normale théorique

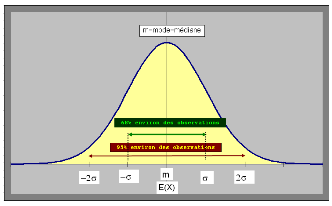
Classes					ni
[	25	-	26	[	3
[	26	-	27	[	8
[	27	-	28	[	18
[	28	-	29	[	35
[	29	-	30	[	85
[	30	-	31	[	100
[	31	-	32	[	79
[	32	-	33	[	41
[	33	-	34	[	23
[	34	-	35	[	6
[	35	-	36	[	2
					400

Problème simple en apparence: on dispose d'une série statistique. Obéit-elle à une loi de probabilité connue? Par exemple la série ci-dessous obéit-elle à une loi Normale?

Plusieurs méthodes pour le montrer:

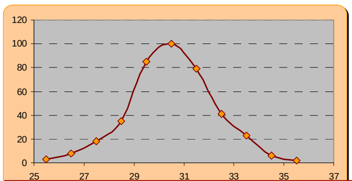
Empirique: Vérifier que cette série respecte les caractéristiques d'une loi normale.
(*Cf diapo 44*).

- **Courbe en cloche**
- **Moyenne, mode et médiane très voisins**
- **68% des observations contenues dans l'intervalle d'un σ autour de la moyenne.**
- **95% des observations contenues dans l'intervalle de deux σ autour de la moyenne.**



Adéquation série statistique à une loi normale théorique: (*Vérification des propriétés: méthode empirique*)

Classes					xi	ni	ECC	nixi	nixi²	di
[	25	-	26	[	25,5	3	3	76,5	1950,75	3
[	26	-	27	[	26,5	8	11	212	5618	8
[	27	-	28	[	27,5	18	29	495	13612,5	18
[	28	-	29	[	28,5	35	64	997,5	28428,75	35
[	29	-	30	[	29,5	85	149	2507,5	73971,25	85
[	30	-	31	[	30,5	100	249	3050	93025	100
[	31	-	32	[	31,5	79	328	2488,5	78387,75	79
[	32	-	33	[	32,5	41	369	1332,5	43306,25	41
[	33	-	34	[	33,5	23	392	770,5	25811,75	23
[	34	-	35	[	34,5	6	398	207	7141,5	6
[	35	-	36	[	35,5	2	400	71	2520,5	2
						400		**12208**	**373774**	



La courbe en cloche évoque bien la courbe de Gauss.

Moyenne	**30,52**
Mode	**30,50**
Médiane	**30,5**
Var	2,96
σ	**1,72**

La moyenne, la médiane et le mode sont très proches.

On peut donc admettre que cette série obéit à une loi normale N(30,5 ; 2,72)

			Effectif	%
$\left[\bar{x}-\sigma, \bar{x}+\sigma\right]$	**Intervalle 1**	28,80 32,24	**300,2**	**75,05%**
$\left[\bar{x}-2\sigma, \bar{x}+2\sigma\right]$	**Intervalle 2**	27,08 33,96	**378,82**	**94,71%**

Les deux intervalles contiennent 75% et près de 95% des observations.

Adéquation série statistique à une loi normale théorique

Classes	ni
[25 - 26 [	3
[26 - 27 [	7
[27 - 28 [	18
[28 - 29 [	40
[29 - 30 [	83
[30 - 31 [	98
[31 - 32 [	79
[32 - 33 [	43
[33 - 34 [	21
[34 - 35 [	6
[35 - 36 [	2
	400

Problème simple en apparence: on dispose d'une série statistique. Obéit-elle à une loi de probabilité connue? Par exemple la série ci-dessous obéit-elle à une loi Normale?

Plusieurs méthodes pour le montrer:

Visuelle:

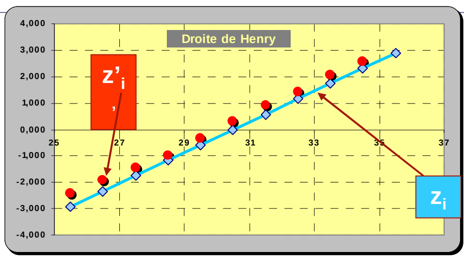
- **Droite de HENRY**

Adéquation série statistique à une loi normale théorique
Droite de Henry

Solution graphique basée sur les éléments suivants:

- **La représentation graphique de la série $z_i = (x_i - m)/\sigma$ est une droite.**
  => Calcul des valeurs réelles de z_i
- **Calcul des FCC à partir des effectifs observés.**
  Fréquences Cumulées Croissantes
- **Calcul des valeurs observées de z_i (z_i') à l'aide des FCC et de la table N(0,1)**
  => P(T< chaque FCC)=z_i'

$z_i=(x_i-m)/\sigma$	fcc	z_i'
-2,916	0,0075	-2,432
-2,335	0,0275	-1,919
-1,754	0,0725	-1,457
-1,173	0,16	-0,994
-0,592	0,3725	-0,325
-0,012	0,6225	0,312
0,569	0,82	0,915
1,150	0,9225	1,422
1,731	0,98	2,054
2,312	0,995	2,576
2,892	1	



P(T<t)=0,82 =>t=0,915 avec T suit N(0,1)

Le nuage rouge suit une droite, près de la droite de Henry. => Normalité

Adéquation série statistique à une loi normale théorique

Classes	ni
[25 - 26 [	3
[26 - 27 [	7
[27 - 28 [	18
[28 - 29 [	40
[29 - 30 [	83
[30 - 31 [	98
[31 - 32 [	79
[32 - 33 [	43
[33 - 34 [	21
[34 - 35 [	6
[35 - 36 [	2
	400

Problème simple en apparence: on dispose d'une série statistique. Obéit-elle à une loi de probabilité connue? Par exemple la série ci-dessous obéit-elle à une loi Normale?

Plusieurs méthodes pour le montrer:

Scientifiques:
- **Test du Khi2**
- **Test de Kolmogorov**

159

Adéquation série statistique à une loi normale théorique: Tests du Khi2 et de kolmogorov.

Classes	ni
[25 - 26 [	3
[26 - 27 [	8
[27 - 28 [	18
[28 - 29 [	35
[29 - 30 [	85
[30 - 31 [	100
[31 - 32 [	79
[32 - 33 [	41
[33 - 34 [	23
[34 - 35 [	6
[35 - 36 [	2
	400

Hypothèses communes aux deux tests:

H0: la série obéit à une loi normale N(30,5 ; 1,72)

H1: la série n'obéit pas à une telle loi.

Risque d'erreur:

α **= 5% et** α **= 1%**

160

Adéquation série statistique à une loi normale théorique: Tests du Khi2 et de kolmogorov.

Indicateurs calculés
Effectifs théoriques (Commun aux deux tests)

On suppose que la série de 400 observations obéit à N(30,5 ; 1,72). Ces éléments sont donc utilisés ci-dessous:

Classes	P(X<borne sup)	Proba(classe)	Eff Th
[25 - 26 [	0,004	0,004	1,46
[26 - 27 [	0,020	0,016	6,45
[27 - 28 [	0,072	0,051	20,48
[28 - 29 [	0,189	0,117	46,81
[29 - 30 [	0,381	0,193	77,06
[30 - 31 [	0,610	0,228	91,39
[31 - 32 [	0,805	0,195	78,08
[32 - 33 [	0,925	0,120	48,05
[33 - 34 [	0,978	0,053	21,30
[34 - 35 [	0,995	0,017	6,80
[35 - 36 [	0,999	0,004	1,56

1,56 = 400 x 0,004 (aux arrondis près)

$P(X < 36) = 0{,}999$

$P(34 < X < 36) = 0{,}004$

161

Adéquation série statistique à une loi normale théorique: Tests du Khi2 et de kolmogorov.

Eff Th (C)	ni (O)
1,46	3
6,45	8
20,48	18
46,81	35
77,06	85
91,39	100
78,08	79
48,05	41
21,30	23
6,80	6
1,56	2

Indicateurs calculés

Khi2

$$\chi^2_{calc} = \sum \frac{(O-C)^2}{C}$$

(O-C)²/C
1,615
0,372
0,300
2,979
0,818
0,812
0,011
1,035
0,136
0,094
0,123
Khi2 Calc 8,293

Kolmogorov

OCC	CCC	|Ecarts|
3	1,46	1,537
11	7,91	3,086
29	28,39	0,608
64	75,20	11,200
149	152,26	3,260
249	243,65	5,352
328	321,72	6,275
369	369,78	0,777
392	391,08	0,922
398	397,88	0,123
400	399,44	0,561

Max	11,200
Kcalc	0,028

Indicateurs théoriques

ν (Nb classes - 1):	**10**
α:	**5%**
Khi2 Th	**18,307**

=> H0 Retenu puisque **khi²** calc < khi² Th

K Th	0,068

$$\frac{1{,}36}{\sqrt{400-1}}$$

=> H0 Retenu puisque K calc < K Th

Les deux tests conduisent à accepter la normalité de cette série.

162

Plan

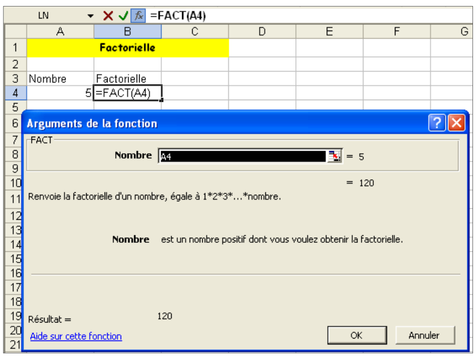
Partie 2 – Probabilités

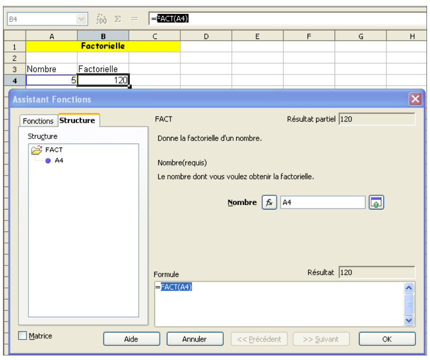
163

Dénombrement - Factorielle

n! = n x (n-1) x (n-2) x …x 3 x 2 x 1

5! = 5 x 4 x 3 x 2 x 1 = 120



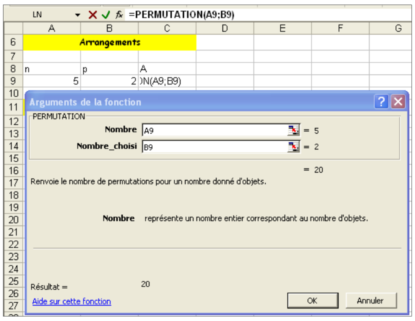


Excel **LibreOffice.org**

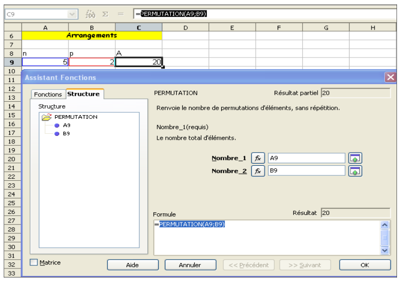
164

Dénombrement - Arrangements

$$A_n^p = \frac{n!}{(n-p)!} \qquad A_5^2 = \frac{5!}{(5-2)!} = 5 \times 4 = 20$$



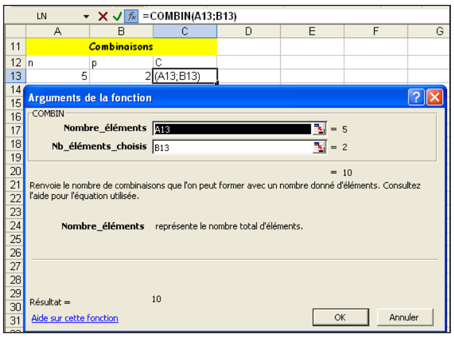
Excel



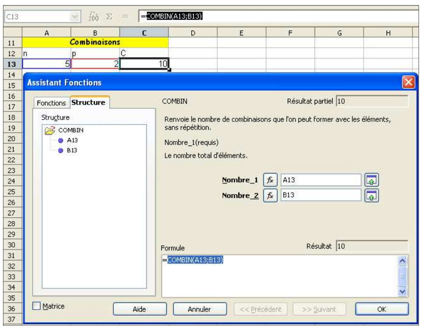
LibreOffice.org

Dénombrement - Combinaisons

$$C_n^p = \frac{A_n^p}{p!} \qquad C_n^p = \frac{n!}{(n-p)!p!} \qquad C_5^2 = \frac{5!}{(5-2)!2!} = 10$$



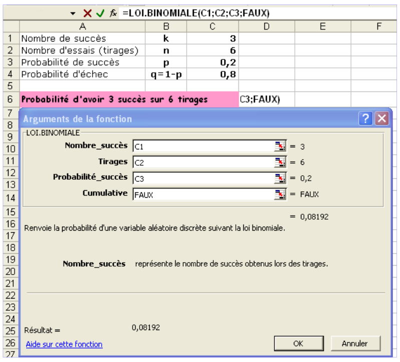
Excel



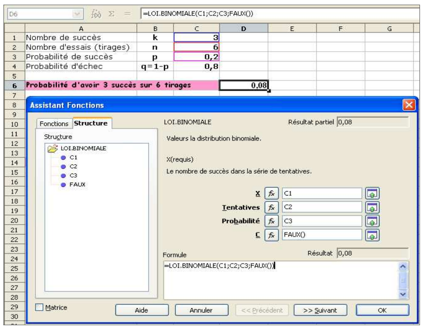
LibreOffice.org

Loi Binomiale

$$P(X = k) = C_n^k \, p^k \, q^{n-k}$$



Excel



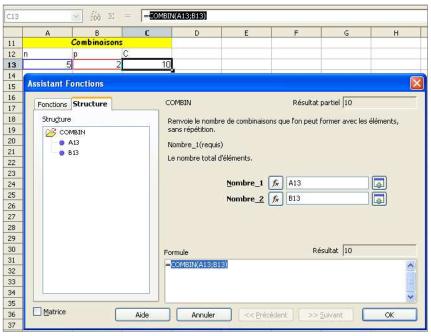
LibreOffice.org

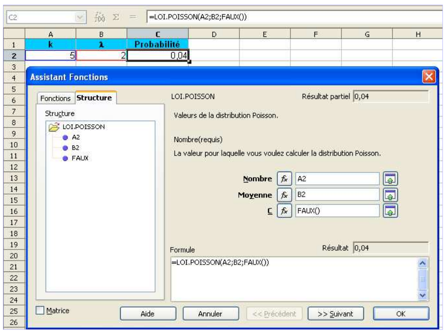
LOI.BINOMIALE()

167

Loi de Poisson

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$





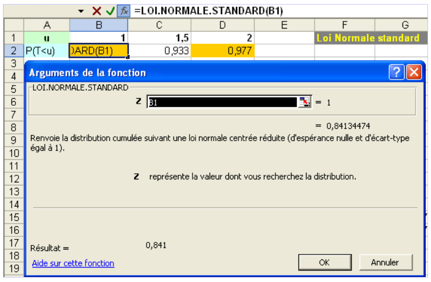
Excel

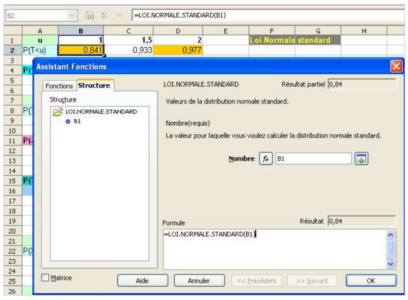
LibreOffice.org

LOI.POISSON()

168

Loi Normale centrée réduite ("Standard")



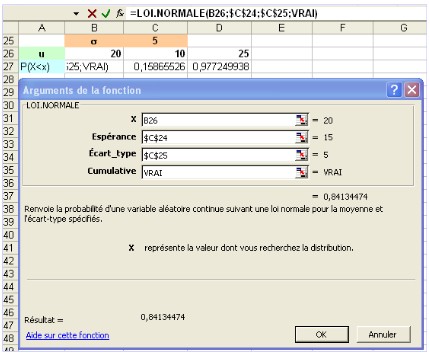


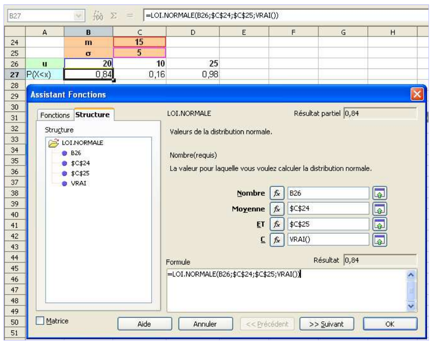
Excel **LibreOffice.org**

LOI.NORMALE.STANDARD()

169

Loi Normale quelconque



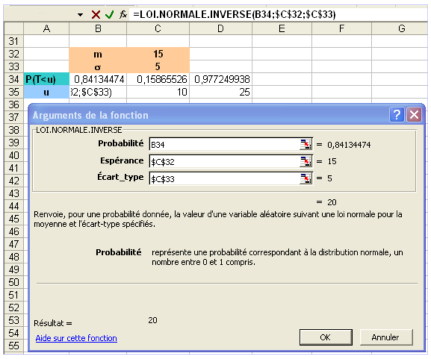


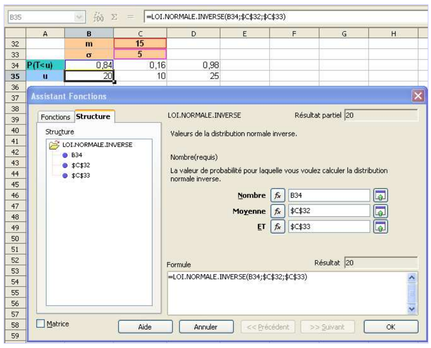
Excel **LibreOffice.org**

LOI.NORMALE.()

170

Loi Normale Inverse (quelconque)





Excel **OpenOffice.org**

LOI.NORMALE.INVERSE() et
LOI.NORMALE.STANDARD.INVERSE(B15)

Fin de la partie 2

Partie 3: Jugement sur échantillon

Plan

Sondages? Elections US 2004.

- Veille scrutin: 4 sondages ne peuvent départager les candidats, 6 donnent un infime avantage à Bush (environ 1%).
- La marge d'erreur de tous ces sondages est comprise entre 2 et 4 points.
- Bush 49% + ou – 4%.
- Kerry 47% + ou – 4%
- Résultat – Bush: 50,73%
- *La marge d'erreur explique l'incertitude et le résultat*.

Sondages? France 2002.

Date de publication	Institut	 Jacques Chirac	 Jean-Marie Le Pen	 Lionel Jospin
10 au 12 avril 2002	CSA	**21 %**	12 %	19 %
10 au 13 avril 2002	BVA	**18,5 %**	14 %	18 %
11 au 12 avril 2002	Ifop	**19 %**	11,5 %	17 %
13 avril 2002	Ifop	**20 %**	13 %	18 %
13 au 15 avril 2002	*Nouvel Observateur*/Sofres	**20 %**	13 %	18 %
17 au 18 avril 2002	CSA	**19,5 %**	14 %	18 %
17 au 18 avril 2002	Ipsos	**20 %**	14 %	18 %
17 au 18 avril 2002	LCI/Sofres	**19,5 %**	13,5 %	17 %
21 avril 2002	Sondage confidentiel	**18 %**	14,5 %	17 %
Résultats du premier tour		**19,88 %**	**16,86 %**	**16,18 %**

La marge d'erreur (+ ou – 3%) explique l'incertitude et le résultat.

Sondages? Grèce 2015.

- Sondages – Avant l'annonce du référendum

Dates	Source	Éditeur	✓ Oui	✗ Non	Ne sait pas
5 juillet	Metron Analysis		46,0 %	**49,0 %**	5,0 %
4-5 juillet	GPO		46 %	**48,5 %**	5,5 %
3-4 juillet	PAMAK		42,5 %	**43,0 %**	13,5 %
2–3 juillet	Metron Analysis		46,0 %	**47,0 %**	7,0 %
1er–3 juillet	GPO		**44,1 %**	43,7 %	12,2 %
1er–3 juillet	Alco		**41,7 %**	41,1 %	17,2 %
30 juin-3 juillet	Ipsos		**44,0 %**	43,0 %	13,0 %
2 juillet	PAMAK		42,5 %	**43,0 %**	14,5 %
30 juin–1er juillet	Alco		**44,8 %**	43,4 %	11,8 %
30 juin[N 3]	Palmos Analysis		37,0 %	**51,5 %**	11,5 %
29–30 juin[N 3]	ProRata	*efsyn.gr, le Journal des Journalistes*	37,0 %	**46,0 %**	17,0 %
28 juin[N 4]			30,0 %	**57,0 %**	13,0 %

Sondages? Grèce 2015.

- Sondages – Après l'annonce du référendum

En faveur d'un accord ?

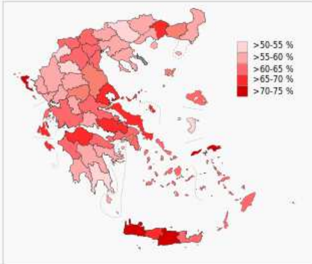
Date	Source	✓ Oui	✗ Non	Ne sait pas
27 juin	PAMAK	26,5 %	**52,0 %**	21,5 %
24–26 juin	Kapa Research	**47,2 %**	33,0 %	19,8 %
24–26 juin	Alco	**57,0 %**	29,0 %	14,0 %

Sondages? Grèce 2015.

Vrais sondages?

- Résultat final

« Approuvez-vous le plan proposé par la Commission européenne, la Banque centrale européenne et le Fonds monétaire international ? »		
Choix	**Votes**	**%**
✗ Non	**3 558 450**	**61.31**
Oui	2 245 537	38.69
Votes valides	5 803 987	94.20
Non valides ou blancs	357 153	5,80
Total des votes	**6 161 140**	**100**
Votants inscrits et participation	9 858 508	62,50
Source : (el) + (en) Ministère de l'intérieur		



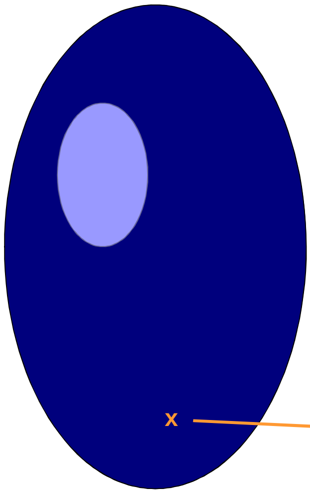
Carte des résultats

Quelques définitions de base

- Individu: Unité statistique de base (Ex: un étudiant TC).
- Population: ensemble fini faisant l'objet d'une observation et dont les éléments répondent à une ou plusieurs caractéristiques communes (Tous les étudiants TC).
- Echantillon: sous-ensemble, partie de la population étudiée (un groupe de 2ème année).

Etude de la population pas toujours possible ni judicieuse => Cf Sondages

Quelques notations de base:



Population:
Taille ou effectif total: N
Moyenne: m. Ecart-type: σ
Fréquence: p

Echantillon:
Taille ou effectif total: n
Moyenne: $\bar{x}$ **Ecart-type:** $\sigma_{\bar{x}}$
Fréquence: f

Individu statistique

Pourquoi travailler sur des échantillons?

- COUT: Une population est souvent composée de très nombreux individus. Les interroger tous suppose des moyens importants. (Recensement Insee)
- TEMPS: Les traitements sur une population sont plus longs.
- Population impossible à recenser de manière exhaustive (Clients potentiels).
- PRECISION: Le gain de précision est négligeable voire dérisoire entre des traitements sur un échantillon et sur une population.
- => Suppose donc l'existence d'une marge d'erreur.

Etude de la population pas toujours possible ni judicieuse => *Cf* Sondages

Comment prélever un échantillon?

- Objectif: échantillon qui possède certaines caractéristiques de la population => « Représentatif ».
- Choix de la méthode la mieux adaptée à la situation.
- Eliminer la subjectivité.

Méthodes basées autour des bases suivantes:

- **Choix raisonné**
- **Prélèvement totalement aléatoire.** → **Supposé être le cas par la suite**
- **Quotas**

A propos des quotas

- Prélèvement des individus composant l'échantillon se fait en fonction de critères précis fixés à priori dictés par la connaissance que l'on a de la population. Souvent l'âge, le sexe, les CSP…

Répartition de la population de la ville de X selon des facteurs socio-économiques. (225 000 habitants)

Sexe		Age		CSP	
Féminin	46%	16-24	14%	Patrons	14%
Masculin	54%	25-44	36%	cadres Sup	6%
		45-64	36%	Cadres, Empl.	16%
		>65	14%	Ouvriers	40%
				Inactifs	24%
	100%		100%		100%

A propos des quotas

Constitution d'un échantillon de 1125 personnes (taux de sondage de 1/200).

Sexe		Age		CSP	
Féminin	518	16-24	158	Patrons	158
Masculin	608	25-44	405	cadres Sup	68
		45-64	405	Cadres, Empl.	180
		>65	158	Ouvriers	450
				Inactifs	270
	1125		1125		1125

Structure pour un enquêteur devant interroger 50 personnes.

Sexe		Age		CSP	
Féminin	23	16-24	7	Patrons	7
Masculin	27	25-44	18	cadres Sup	3
		45-64	18	Cadres, Empl.	8
		>65	7	Ouvriers	20
				Inactifs	12
	50		50		50

185

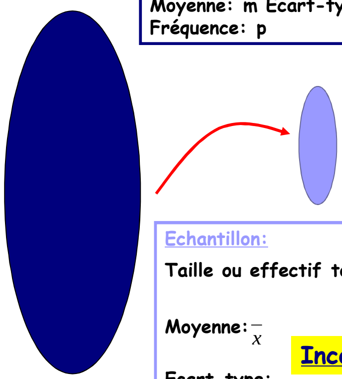
Plan

186

Objectif 1: Echantillonnage (Induction).

Population: **Connue**

Taille ou effectif total: N
Moyenne: m Ecart-type: σ
Fréquence: p



Echantillon:

Taille ou effectif total: n **Connu**

Moyenne: $\bar{x}$

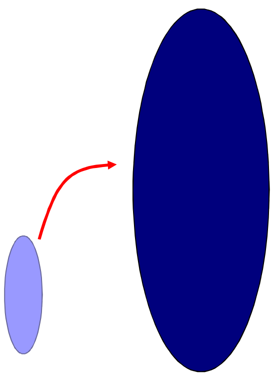
Ecart-type: $\sigma_{\bar{x}}$
Fréquence: f

Inconnus => Moyenne à estimer.

Problème: peut-on juger de la représentativité d'un échantillon en fonction d'un paramètre quantitatif de la population?

187

Objectif 2: Estimation.



Population:

Taille ou effectif total: N
Moyenne: m Ecart-type: σ
Fréquence: p

Inconnus => Moyenne à estimer.

Problème: peut-on généraliser à l'ensemble de la population les résultats observés sur un échantillon?

Echantillon: **Connu**

Taille ou effectif total: n
Moyenne: $\bar{x}$ Ecart-type: $\sigma_{\bar{x}}$
Fréquence: f

188

Relations population-échantillons

Soit une population de 5 étudiants ayant passé une épreuve de statistiques.

La moyenne de cette promo est: m = 12,2

Noms	Notes
A	15
B	8
C	11
D	9
E	18
Total	61

On s'intéresse à présent à des échantillons de taille 3 prélevés dans cette population. Combien peut-on constituer de tels échantillons?

$$C_5^3 = \frac{5!}{3!\times(5-3)!} = \frac{5\times4\times3\times2\times1}{(3\times2\times1)\times(2\times1)} = 10$$

	Echantillons	Notes			Moyenne des échantillons
1	A B C	15	8	11	11.33
2	A B D	15	8	9	10.67
3	A B E	15	8	18	13.67
4	A C D	15	11	9	11.67
5	A C E	15	11	18	14.67
6	A D E	15	9	18	14.00
7	B C D	8	11	9	9.33
8	B C E	8	11	18	12.33
9	B D E	8	9	18	11.67
10	C D E	11	9	18	12.67

Ce résultat est toujours vrai.

La moyenne des moyennes des échantillons est: 12,2

Relations population-échantillons

Soit une population de 5 étudiants ayant passé un examen.

La proportion moyenne de cette promo est: p=2/5 = 0.4.

Noms	Admis
A	oui
B	non
C	non
D	oui
E	non
Total	2

On s'intéresse à présent à des échantillons de taille 3 prélevés dans cette population. Combien peut-on constituer de tels échantillons?

	Echantillons	Résultat			% Moyenne des échantillons
1	A B C	oui	non	non	0.33
2	A B D	oui	non	oui	0.67
3	A B E	oui	non	non	0.33
4	A C D	oui	non	oui	0.67
5	A C E	oui	non	non	0.33
6	A D E	oui	oui	non	0.67
7	B C D	non	non	oui	0.33
8	B C E	non	non	non	0.00
9	B D E	non	oui	non	0.33
10	C D E	non	oui	non	0.33

Ce résultat est toujours vrai.

La proportion moyenne des moyennes des échantillons est: 0,4

Relations population-échantillons

Imaginons une population de taille N=1500 personnes. Combien d'échantillons de taille n=120 peut-on prélever?

$$C_{1500}^{120} = 1,5175 \times 10^{180}$$

Autant dire une infinité.

- Prélever 120 personnes parmi 1500 est équivalent à choisir 120 boules dans une urne qui en contient 1500…
- Le choix d'un échantillon fait donc toujours intervenir des grands nombres.
- Le choix d'un échantillon est une épreuve aléatoire (qui dépend du hasard).
- Nous supposerons par la suite que le prélèvement d'un échantillon est TOTALEMENT aléatoire.
- L'utilisation de résultats qui découlent des propriétés d'une loi de probabilité (la loi Normale) est dont logique.
- Maîtriser l'utilisation de la table de probabilité d'une loi Normale est donc préférable…

Relations population-échantillons

Impossibilité dans les faits de disposer de la distribution d'échantillonnage des moyennes

	Echantillons	Notes			Moyenne des échantillons
1	A B C	15	8	11	11.33
2	A B D	15	8	9	10.67
3	A B E	15	8	18	13.67
4	A C D	15	11	9	11.67
5	A C E	15	11	18	14.67
6	A D E	15	9	18	14.00
7	B C D	8	11	9	9.33
8	B C E	8	11	18	12.33
9	B D E	8	9	18	11.67
10	C D E	11	9	18	12.67

- Pouvoir estimer la moyenne d'une population, ou se montrer vigilant sur celle d'un échantillon suppose que l'on connaisse et maîtrise les propriétés des moyennes des échantillons et de la moyenne de la population.
- Ces propriétés reposent sur une loi de probabilité incontournable

Plan

Propriétés de la distribution d'échantillonnage.

Jamais disponible dans les faits

	…illons	Notes			Moyenne des échantillons
1	A B C	15	8	11	11.33
2	A B D	15	8	9	10.67
3	A B E	15	8	18	13.67
4	A C D	15	11	9	11.67
5	A C E	15	11	18	14.67
6	A D E	15	9	18	14.00
7	B C D	8	11	9	9.33
8	B C E	8	11	18	12.33
9	B D E	8	9	18	11.67
10	C D E	11	9	18	12.67

$\longrightarrow$

N° Echantillon	Composition	Moyenne
1		$\overline{x}_1$
2		$\overline{x}_2$
….. ….	……. ………	
xxx		$\overline{x}_n$

- **La distribution d'échantillonnage est une série statistique avec un effectif très important…dont on ne connaît pas la composition.**
- **Il est donc impossible de disposer des moyennes de tous les échantillons.**
- **Par contre les propriétés de cette distribution sont bien connues.**

Propriétés de la distribution d'échantillonnage des moyennes

- **Si n > 30 => Distribution d'échantillonnage des moyennes suit une loi Normale (X).**
  - **N>=20n, X suit:** $N(m, \frac{\sigma}{\sqrt{n}})$
  - **N<20n, X suit:** $N(m, \frac{\sigma}{\sqrt{n}} \times \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\text{Coefficient d'exhaustivité}})$

Avec:
N: taille population
n: taille échantillon
m: moyenne population
σ: écart type population

- **Si n < 30 => Distribution d'échantillonnage des moyennes ne suit pas une loi Normale, mais une loi de Student (X).**

Application Pratique: moyennes

- Une entreprise décide de mener une enquête auprès de ses 5 000 clients. Le principal critère de représentativité de ceux-ci est le CA moyen réalisé au cours de la dernière année.
- La dépense moyenne des clients est de 1500€ avec un écart type de 300€.
- On souhaite disposer d'un échantillon de 100 personnes ayant 95% de chances d'avoir sa moyenne « proche » de 1500€.
- Quelles valeurs pour la moyenne d'un échantillon peut-on accepter?

Nous disposons de:
N=5 000, et n=100, donc N > 20n
m=1500, et σ=300
Probabilité de succès: 95%

La distribution d'échantillonnage des moyennes suit une loi normale (X) dont les caractéristiques sont les suivantes:

$$N(m, \frac{\sigma}{\sqrt{n}}) = N(1500; \frac{300}{\sqrt{100}}) = N(1500;30)$$

Il nous faut donc déterminer un intervalle (dit de confiance) tel que tout échantillon ayant sa moyenne dans cet intervalle aura 95% de chances d'être satisfaisant.

Application Pratique: Moyennes

- Exprimons que la moyenne d'un échan[illegible]llon quelconque a 95% de chances d'appartenir à un intervalle ce[illegible]tré autour de la moyenne de la population (m).
- P(m-a<X<m+a)=0,95, avec a précisi[illegible] de l'intervalle recherché.
- Donc ici P(1500-a<X<1500+a)=0,95. [illegible] suit une loi normale de paramètre N(1500,30). En vertu du ch[illegible]gement de variable à présent connu: $T = \frac{X - 1500}{30}$

Table de la loi normale centrée réduite

υ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	[illegible]	0.50798	0.51197	0.51595	0.51994	[illegible]	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	[illegible]	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	[illegible]	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	[illegible]	0.62930	0.63307	[illegible]	[illegible]	[illegible]	0.64803	0.65173
0.4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.68439	0.68793
0.5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	[illegible]	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	[illegible]	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	[illegible]	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	[illegible]	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	[illegible]	0.85083	0.85314	[illegible]	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	[illegible]	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	[illegible]	0.89796	0.89973	0.90147
1.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	[illegible]	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	[illegible]	0.93699	[illegible]	[illegible]	[illegible]	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	[illegible]	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	[illegible]	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	[illegible]	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	[illegible]	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	[illegible]	[illegible]	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

- P(-a/30<T<a/30)=0.95
- Donc P(T<a/30)-P(T<-a/30)=0,95
- Et 2P(T<a/30)-1=0,95
- Soit P(T<a/30)<1,95/2 => P(T<a/30)=0,975

- La lecture dans la table N(0;1) donne: a/30=1,96
- Finalement a=30 x 1,96 = 58.8
- Et notre intervalle: [1500–58,8 ; 1500+58,8], soit [1441,2; 1558,8]
- Conclusion.

Application Pratique: Cas des proportions

- Une entreprise décide de mener une enquête auprès de ses 3 000 clients sur une carte de fidélité. Le principal critère de représentativité est la possession de la carte que seuls 750 d'entre eux ont.
- La proportion observée est donc p=750/3000=0,25.
- On souhaite disposer d'un échantillon de 144 personnes ayant 95% de chances d'avoir un taux de possession de la carte « voisin » de 0,25
- Quelles valeurs pour la proportion d'un échantillon peut-on accepter?

Les propriétés de la distribution d'échantillonnage des proportions sont proches de celle des moyennes.

Propriétés de la distribution d'échantillonnage des proportions

- **Si n > 30 => Distribution d'échantillonnage des proportions suit une loi Normale (X).**
  - **N>=20n, X suit:** $N(p, \sqrt{\frac{pq}{n}})$
  - **N<20n, X suit:** $N(p, \sqrt{\frac{pq}{n}} \times \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\text{Coefficient d'exhaustivité}})$

Avec:
N: taille population
n: taille échantillon
p: proportion population
q=1-p
pq: variance population

- **Si n < 30 => Distribution d'échantillonnage des proportions ne suit pas une loi Normale, mais une loi de Student (X).**

199

Application Pratique: Cas des proportions

- Une entreprise décide de mener une enquête auprès de ses 3 000 clients sur une carte de fidélité. Le principal critère de représentativité est la possession de la carte que seuls 750 d'entre eux ont.
- La proportion observée est donc p=750/3000=0,25.
- On souhaite disposer d'un échantillon de 144 personnes ayant 95% de chances d'avoir un taux de possession de la carte « voisin » de 0,25
- Quelles valeurs pour la proportion d'un échantillon peut-on accepter?

Nous disposons de:
N=3 000, et n=144, donc N > 20n
p=0,25, et σ=(0,25x0,75)²
Probabilité de succès: 95%

La distribution d'échantillonnage des proportions suit une loi normale (X) dont les caractéristiques sont les suivantes:

$$N(p, \sqrt{\frac{pq}{n}}) = N(0{,}25; \sqrt{\frac{0{,}25 \times 0{,}75}{144}}) = N(0{,}25; 0{,}036)$$

Il nous faut donc déterminer un intervalle (dit de confiance) tel que tout échantillon ayant sa proportion dans cet intervalle aura 95% de chances d'être satisfaisant.

200

Application Pratique: Cas des proportions

- **Exprimons que la proportion d'un échantillon quelconque a 95% de chances d'appartenir à un intervalle centré autour de la proportion de la population (p).**
- **P(p-a<X<p+a)=0,95, avec a précision de l'intervalle recherché.**
- **Donc ici P(0,25-a<X<0,25+a)=0,95. X suit une loi normale de paramètre N(0,25;0,036). En vertu du changement de variable à présent connu:** $T = \frac{X - 0{,}25}{0{,}036}$

- **P(-a/0,036<T<a/0,036)=0.95**
- **Donc P(T<a/0,036)-P(T<-a/0,036)=0,95**
- **Et 2P(T<a/,0,36)-1=0,95**
- **Soit P(T<a/0,036)<1,95/2 => P(T<a/0,036)=0,975**

- **La lecture dans la table N(0;1) donne: a/0,036=1,96**
- **Finalement a=0,036 x 1,96 = 0,061**
- **Et notre intervalle: [0,25-0,061 ; 0,25+0,061], soit [0,19 ; 0,31]**
- **Conclusion.**

Plan

Estimation

- Données relatives à la population inconnues: Moyenne, Ecart type (le plus souvent)
- Echantillon prélevé de manière totalement aléatoire: Taille, Moyenne et Ecart type calculés => Connus.
- Comment généraliser les résultats de l'échantillon à la population?
- Objectif: disposer d'une estimation de m moyenne de la population.

Estimation: Modification de la distribution d'échantillonnage.

1er Cas: σ est connu.

- **Si n > 30 => Distribution d'échantillonnage des moyennes suit une loi Normale (X).**
  - **N>=20n, X suit:** $N(\bar{x}, \frac{\sigma}{\sqrt{n}})$
  - **N<20n, X suit:** $N(\bar{x}, \frac{\sigma}{\sqrt{n}} \times \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\text{Coefficient d'exhaustivité}})$

Avec:
N: taille population
n: taille échantillon
m: moyenne population
σ: écart type population

- **Si n < 30 => Distribution d'échantillonnage des moyennes ne suit pas une loi Normale, mais une loi de Student (X).**

Estimation: Modification de la distribution d'échantillonnage.

2ème Cas: σ est inconnu.

$$S = \sqrt{\frac{n}{n-1}} \times \sigma_{\bar{x}}$$ *Estimateur de σ*

- **Si n > 30 => Distribution d'échantillonnage des moyennes suit une loi Normale (X).**
  - N>=20n, X suit: $N(\bar{x}, \frac{S}{\sqrt{n}})$
  - N<20n, X suit: $N(\bar{x}, \frac{S}{\sqrt{n}} \times \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\text{Coefficient d'exhaustivité}})$

Avec:
N: taille population
n: taille échantillon
m: moyenne population
σ: écart type population

- **Si n < 30 => Distribution d'échantillonnage des moyennes ne suit pas une loi Normale, mais une loi de Student (X).**

205

Application Pratique: estimation de la moyenne

- Une entreprise d'estimer la dépense moyenne de ses 1900 clients.
- Sur un échantillon de 81 personnes on obtient 235€ de moyenne et 99€ d'écart-type.
- Peut-on disposer d'un intervalle de confiance à 5% pour la moyenne?

Nous disposons de:
N=1 900, et n=81, donc N > 20n
σ inconnu
Probabilité de succès: 95%
Moyenne et écart type de l'échantillon connus.

La distribution d'échantillonnage des moyennes suit une loi normale (X) dont les caractéristiques sont les suivantes:

$$S = \sqrt{\frac{n}{n-1}}\sigma_{\bar{x}} = \sqrt{\frac{81}{80}} \times 99 = 99{,}62$$

$$N(\bar{x}, \frac{S}{\sqrt{n}}) = N(235; \frac{99{,}62}{\sqrt{81}}) = N(235; 11{,}07)$$

Il nous faut donc déterminer un intervalle (dit de confiance) ayant 95% de chances de contenir la moyenne de la population.

206

Application Pratique: Moyennes

- **Exprimons que la moyenne de la population a 95% de chances d'appartenir à un intervalle centré autour de la moyenne de l'échantillon**
- **$P(235-a < X < 235+a)=0,95$, avec a précision de l'intervalle recherché.**
- **Donc ici X suit une loi normale de paramètre N(235 ; 11,07). En vertu du changement de variable à présent connu:**

$$T = \frac{X-235}{11,97}$$

- **$P(-a/11,07 < T < a/11,07)=0.95$**
- **Donc $P(T < a/11,07)-P(T < -a/11,07)=0,95$**
- **Et $2P(T < a/11,07)-1=0,95$**
- **Soit $P(T < a/11,07) < 1,95/2 \Rightarrow P(T < a/11,07)=0,975$**

- **La lecture dans la table N(0;1) donne: a/11,07=1,96**
- **Finalement a=11,07 x 1,96 = 21,7**
- **Et notre intervalle: [235-21,6 ; 235+21,7], soit [213,31; 256,7]**
- **Conclusion.**

207

Fin de la partie 3

208