CM 1
Présentation de l'EC ; Exemple d'expérience
Vocabulaire ; Nature d'une variable ; Centralité et dispersion

Christophe Genolini

24 septembre 2007

EC Statistique et Informatique

1 Présentation de l'EC

- Responsable : Genolini, Leborgne
- 2x7=14 h CM (Genolini)
- 1,5x12=18h TD (Genolini, Leborgne, Driss)
- Contrôle continu : 6 mini QCM, les premiers TD
- Contrôle terminal : 40 minutes sur machine
- Retard en CM => Premier rang
- Retard en TD => Pénalité sur la note
- Outil de travail : R et Excel

On reprend tout à zéro ; on utilise un nouveau logiciel

$\Downarrow$

Tout le monde peut réussir

2 Exemple

5 étudiants “plutot beau mecs” de Nicola Guegen demandent à des jeunes filles : “Bonjour Mademoiselle, je vous trouve vraiment ravissante mais là, il faut que j’aille travailler je n’ai vraiment pas le temps : accepteriez-vous de me donner votre numéro de téléphone pour que je puisse vous inviter à prendre un verre un peu plus tard ?”.
La moitié des demandes sont faites avec un léger touché de l’avant bras, dans l’autre, pas de touché.

Résultat (suspens...) :

[Id]	[Contact]	[T l phone]
P1	Sans contact	Non
P2	Sans contact	Non
P3	Avec contact	Oui
P4	Sans contact	Non
P5	Avec contact	Oui
P6	Avec contact	Non
P7	Sans contact	Non
P8	Sans contact	Non
P9	Sans contact	Non
P10	Sans contact	Non
P11	Sans contact	Non
P12	Avec contact	Non
P13	Sans contact	Non
P14	Avec contact	Non
P15	Sans contact	Non
P16	Avec contact	Oui
P17	Sans contact	Non
P18	Sans contact	Non
P19	Avec contact	Non
P20	Sans contact	Non
P21	Avec contact	Non
P22	Sans contact	Non
P23	Avec contact	Non
P24	Avec contact	Non
P25	Sans contact	Non
P26	Sans contact	Non
P27	Avec contact	Non
P28	Avec contact	Non
P29	Sans contact	Non
P30	Sans contact	Non
P31	Avec contact	Non
P32	Sans contact	Non
P33	Avec contact	Non
P34	Sans contact	Non
P35	Sans contact	Non
P36	Avec contact	Oui
P37	Sans contact	Non
P38	Sans contact	Non
P39	Sans contact	Non
P40	Avec contact	Non
P41	Sans contact	Non
P42	Avec contact	Non
P43	Sans contact	Non
P44	Sans contact	Non
P45	Avec contact	Non
P46	Avec contact	Non
P47	Sans contact	Non
P48	Sans contact	Non
P49	Avec contact	Non
P50	Sans contact	Non
P51	Sans contact	Non
P52	Avec contact	Non
P53	Avec contact	Non
P54	Sans contact	Non
P55	Avec contact	Non
P56	Avec contact	Non
P57	Sans contact	Non
P58	Avec contact	Non
P59	Sans contact	Non
P60	Sans contact	Oui

[Id]	[Contact]	[T l phone]
P61	Avec contact	Oui
P62	Sans contact	Non
P63	Avec contact	Non
P64	Sans contact	Non
P65	Avec contact	Non
P66	Sans contact	Non
P67	Sans contact	Non
P68	Avec contact	Oui
P69	Sans contact	Non
P70	Avec contact	Oui
P71	Avec contact	Non
P72	Sans contact	Non
P73	Sans contact	Non
P74	Avec contact	Non
P75	Avec contact	Non
P76	Sans contact	Oui
P77	Avec contact	Non
P78	Sans contact	Non
P79	Avec contact	Non
P80	Avec contact	Oui
P81	Sans contact	Non
P82	Avec contact	Non
P83	Avec contact	Non
P84	Avec contact	Oui
P85	Sans contact	Non
P86	Avec contact	Non
P87	Avec contact	Non
P88	Sans contact	Non
P89	Avec contact	Non
P90	Avec contact	Non
P91	Sans contact	Non
P92	Sans contact	Non
P93	Avec contact	Non
P94	Avec contact	Non
P95	Sans contact	Non
P96	Sans contact	Non
P97	Avec contact	Non
P98	Sans contact	Non
P99	Avec contact	Non
P100	Avec contact	Non
P101	Avec contact	Non
P102	Avec contact	Non
P103	Sans contact	Non
P104	Avec contact	Non
P105	Sans contact	Non
P106	Avec contact	Oui
P107	Avec contact	Oui
P108	Avec contact	Non
P109	Avec contact	Non
P110	Sans contact	Non
P111	Avec contact	Non
P112	Avec contact	Non
P113	Avec contact	Oui
P114	Avec contact	Non
P115	Sans contact	Non
P116	Sans contact	Non
P117	Avec contact	Non
P118	Avec contact	Non
P119	Sans contact	Non
P120	Sans contact	Non

[Id]	[Contact]	[T l phone]
P121	Avec contact	Non
P122	Sans contact	Non
P123	Avec contact	Non
P124	Avec contact	Non
P125	Sans contact	Non
P126	Sans contact	Non
P127	Avec contact	Oui
P128	Sans contact	Non
P129	Sans contact	Non
P130	Avec contact	Non
P131	Sans contact	Non
P132	Avec contact	Oui
P133	Sans contact	Non
P134	Avec contact	Non
P135	Avec contact	Oui
P136	Avec contact	Oui
P137	Avec contact	Oui
P138	Sans contact	Non
P139	Sans contact	Non
P140	Avec contact	Non
P141	Avec contact	Non
P142	Sans contact	Non
P143	Avec contact	Non
P144	Sans contact	Non
P145	Avec contact	Non
P146	Sans contact	Non
P147	Avec contact	Oui
P148	Avec contact	Non
P149	Avec contact	Non
P150	Sans contact	Non
P151	Sans contact	Non
P152	Avec contact	Non
P153	Sans contact	Non
P154	Sans contact	Non
P155	Avec contact	Non
P156	Sans contact	Oui
P157	Sans contact	Oui
P158	Sans contact	Non
P159	Avec contact	Non
P160	Avec contact	Non
P161	Avec contact	Non
P162	Sans contact	Non
P163	Sans contact	Non
P164	Sans contact	Non
P165	Sans contact	Non
P166	Sans contact	Non
P167	Avec contact	Non
P168	Sans contact	Non
P169	Avec contact	Non
P170	Avec contact	Non
P171	Avec contact	Non
P172	Sans contact	Non
P173	Avec contact	Non
P174	Sans contact	Non
P175	Avec contact	Non
P176	Sans contact	Non
P177	Avec contact	Non
P178	Avec contact	Non
P179	Sans contact	Non
P180	Sans contact	Non

[Id]	[Contact]	[T l phone]
P181	Avec contact	Non
P182	Avec contact	Non
P183	Avec contact	Non
P184	Sans contact	Non
P185	Sans contact	Non
P186	Avec contact	Non
P187	Avec contact	Non
P188	Avec contact	Non
P189	Sans contact	Oui
P190	Sans contact	Oui
P191	Avec contact	Non
P192	Avec contact	Oui
P193	Avec contact	Non
P194	Avec contact	Non
P195	Avec contact	Oui
P196	Avec contact	Non
P197	Avec contact	Non
P198	Avec contact	Non
P199	Sans contact	Non
P200	Sans contact	Non
P201	Sans contact	Non
P202	Sans contact	Non
P203	Sans contact	Non
P204	Avec contact	Non
P205	Avec contact	Non
P206	Avec contact	Oui
P207	Avec contact	Oui
P208	Avec contact	Non
P209	Sans contact	Oui
P210	Avec contact	Non
P211	Avec contact	Non
P212	Avec contact	Non
P213	Sans contact	Non
P214	Sans contact	Non
P215	Sans contact	Non
P216	Sans contact	Non
P217	Avec contact	Non
P218	Avec contact	Non
P219	Sans contact	Non
P220	Sans contact	Non
P221	Sans contact	Non
P222	Sans contact	Non
P223	Sans contact	Oui
P224	Sans contact	Non
P225	Sans contact	Oui
P226	Sans contact	Non
P227	Sans contact	Oui
P228	Avec contact	Non
P229	Sans contact	Oui
P230	Avec contact	Non
P231	Sans contact	Non
P232	Sans contact	Non
P233	Avec contact	Non
P234	Sans contact	Oui
P235	Avec contact	Non
P236	Sans contact	Non
P237	Sans contact	Non
P238	Avec contact	Oui
P239	Sans contact	Non
P240	Sans contact	Non

TAB. 1 – Gueguen 2007, données brutes (reconstituées)

Que peut-on en dire ? Rien sous cette forme. Par contre, un tableau croisé donne :

Groupe	**Nombre de personne**
Avec contact, acceptent	23
Avec contact, refusent	97
Sans contact, acceptent	12
Sans contact, refusent	120

TAB. 2 – Gueguen, données triées

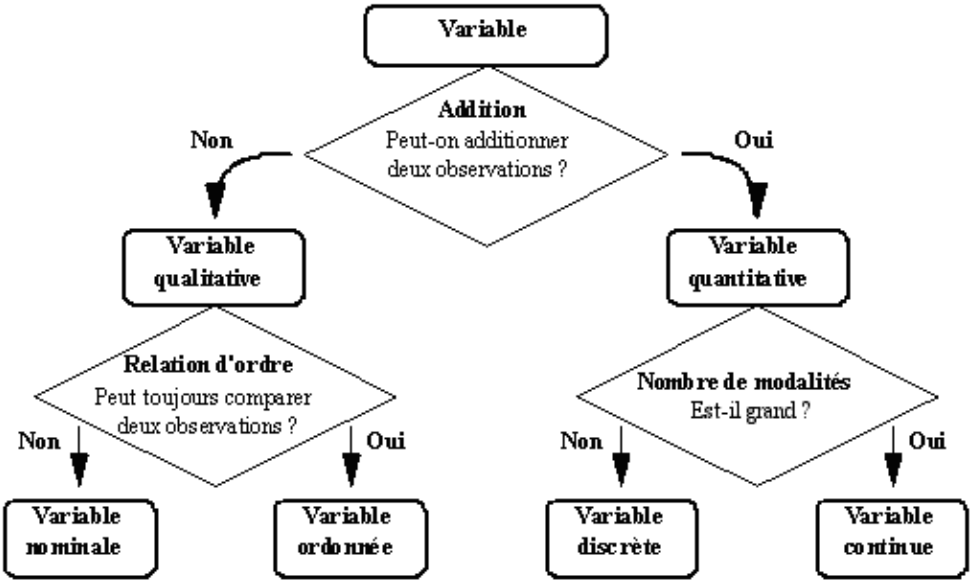
Conclusion ? Sans contact, 10% des demoiselles acceptent de donner leur numéro de téléphone. Avec contact, 19.2% acceptent...

3 Vocabulaire

Nom	Définition	Exemple
Individu	Objet étudié	un étudiant
Population	Ensemble des individus	tous les étudiants participant à l'expérience
Variables	Ce qui est étudié chez les individus (et qui varie d'un individu à l'autre)	[Age], [CigaretteJour], [NiveauSportif]
Les modalités d'une variable	Liste de toutes les valeurs possibles pour une variable	Modalité de age : de 0 ans a 120 ans
Observation	Valeurs prise par un individu particulier	l'individu 3 a 21 ans

4 Nature d'une variable

La nature d'une variable détermine le type d'outil statistique qu'on pourra utiliser sur la variable. Pour déterminer son type, il faut se demander ce qu'on peut faire avec ses observations.



5 Nomenclature

Le choix du nom des variables est très important.

1. ROWNAMES	2. QUANTAGE	3. QUALAGE	4. SEX	5. EDU
5. RACE	6. INCOME	7. PTID	8. CAT1	9. CAT2
11. SWANG1	12. DEATH	13. SADMDTE	14. DSCHDTE	15. DTHDTE
16. LSTCTDTE	17. CA	18. CARDIOHX	19. CHFHX	20. DEMENTHX
21. PSYCHHX	22. CHRPULHX	23. RENALHX	24. LIVERHX	25. IMMUNHX
26. GIBLEDHX	27. MALIGHX	28. AMIHX	29. RESP	30. CARD
31. NEURO	32. GASTR	33. RENAL	34. META	35. HEMA
36. SEPS	37. TRAUMA	38. ORTHO	39. APS1	40. SCOMA1
41. DAS2D3PC	42. ADLD3P	43. WTKILO1	44. TEMP1	45. HRT1
46. RESP1	47. MEANBP1	48. WBLC1	49. PAFI1	50. ALB1
51. HEMA1	52. BILI1	53. CREA1	54. SOD1	55. POT1
56. PACO21	57. PH1	58. URIN1	59. DNR1	60. TRANSHX

TAB. 3 – Connors, liste des variables initiales

Mauvaise nomenclature : Si, chez un patient, on remarque que `[44.TEMP1]=(33)`, `[2.QUANTAGE]=(19)` et `[28.AMIHX]=(Non)`, on ne peut rien en dire.

1. IdPatient	2. Age	3. AgeParClasse	4. Sexe	5. Education
6. Race	7. Salaire	8. AssuranceMed	9. Pathologie1	10. Pathologie2
11. Catheterisation	12. Mort	13. DateAdmission	14. DateDechargeHopital	15. DateDeces
16. DateDerniereNouvelle	17. Cancer	18. InsuffisanceCardiaque	19. Prb_Cardiaques	20. Prb_Neurologiques
21. Prb_Psycologique	22. Prb_Pulmonaire	23. Prb_Renaux	24. Prb_Hepatic	25. Prb_Immunitaire
26. HemoragieGastroSup	27. TumeurSolide	28. InfarctusMyocarde	29. DiagRespiratoire	30. DiagCardiaque
31. DiagNeurologique	32. DiagGastrologique	33. DiagRenal	34. DiagMetabolique	35. DiagHematologique
36. DiagSepsis	37. DiagTraumatique	38. DiagOrthopedique	39. APACHE	40. GLASGOW
41. IndependanceDASI	42. IndependanceADL	43. Poids	44. Temperature	45. FrepCardiaque
46. FreqRespiratoire	47. TensionArterielle	48. HematoGlobulesBlancs	49. PressionArtO2Sonde	50. HematoAlbumine
51. HematoHematocrit	52. HematoBilirubin	53. HematoCreatine	54. HematoSodium	55. HematoPotassium
56. PressionArtCO2	57. HematoPH	58. Urine	59. NonResurection	60. Transfer

TAB. 4 – Connors, liste des variables renommées

Bonne nomenclature : Si, chez un patient, `[44.Temp rature]=(33)`, clairement le patient a très froid. `[2.Age]=(19)`, il est tout jeune alors que `[28.InfarctusMyocarde]=(Non)` indique qu'il n'a pas eu d'infarctus du myocarde. Quant à savoir le traitement qui lui a été appliqué, il suffira probablement de regarder la variable `[11.Catheterisation]`

6 Compter les données

6.1 Effectifs

L'effectif d'une modalité m_i est le nombre de personnes dont la variable vaut m_i. La tableau des effectifs est la liste de toutes les modalités avec leur effectif.

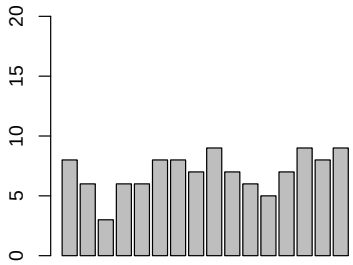
6.2 Valeurs aberrantes et données manquantes

Une valeur aberrante est une valeur qui est liée à une erreur de mesure. Une donnée manquante est une donnée qui n'est pas renseignée. Les valeurs manquantes ne rentrent pas dans l'analyse statistique (les individus sont supprimés). Par contre, il est très important de noter le nombre de données manquantes.
Cinq femmes courrent un 200m : 22.03s ; 23.56 ; 24,13 ; 248.2s. Y a-t-il une valeur aberrante ?

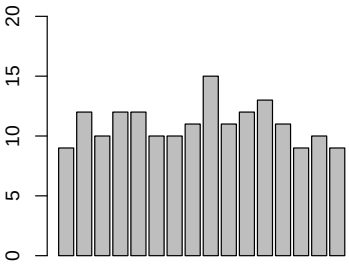
7 Centralité et dispersion

7.1 Intuitivement

Considérons deux groupes de TD (un groupe est noté chaque semaine) :



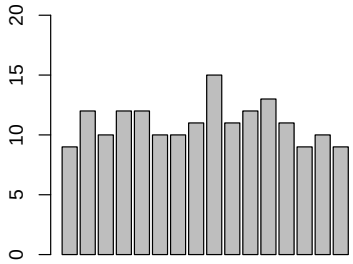
Note, semaine 1



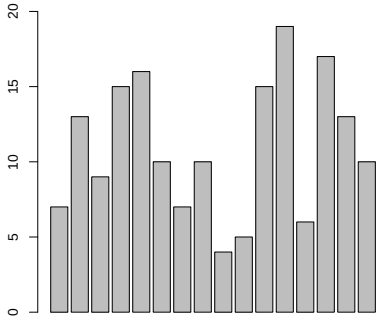
Note, semaine 2

On “résume” les notes par un indice, la moyenne. Une moyenne haute indique que globalement, les notes sont hautes et inversement.

Considérons les groupes 2 et 3 :



Note, semaine 2



La moyenne est la même. Pourtant, les groupes ne se ressemblent pas. Dans le groupe 2, les notes sont proches de la moyenne ; dans le groupe 3, elles sont éparpillées. On “résume” cette concentration ou éparpillement par les indices de centralités.
Le premier indice auquel on pense est l’écart absolue moyen, mais il n’est pas simple d’utilisation parce qu’il fait appelle à la fonction valeur absolue. Le deuxième indice est la variance, mais l’unité de la variance est celle de la moyenne au carré. Au final, on préfère donc l’écart type.

7.2 Calculs

- **Moyenne :** on additionne toutes les notes et on divise par le nombre de notes. $Moyenne(2, 10, 6, 8, 5) = \frac{2+10+6+8+5}{5}$
- **Médiane :** on ordonne toutes les notes, on prend celle du milieu. $Mdiane(2, 10, 6, 8, 5) = Milieu(2, 5, 6, 8, 10) = 6$
- **Mode :** on prend la modalité qui a le plus grand effectif.
- **Variance :** pour chaque note, on calcul l’écart au carré entre la note et la moyenne ; puis on fait la moyenne des écarts au carré : Pour calculer $Variance(2, 3, 5, 6)$:
  - $Moyenne(2, 3, 5, 6) = 4$
  - Ecarts : $2 - 4 = -2; 3 - 4 = -1; 5 - 4 = +1; 6 - 4 = +2$
  - Ecarts au carré : $-2^2 = 4; -1^2 = 1; +1^2 = 1; +2^2 = 4$

  - Moyenne des écarts au carré : $Variance(2,3,5,6) = \frac{4+2+2+4}{4} = 3$
- **Ecart type :** on prend la racine carré de la variance.
- **Quartiles :** on ordonne les notes puis on prend la plus basse pour $Q0$, la plus haute pour $Q4$, la médiane pour $Q2$; $Q1$ est alors la note à égale distance entre $Q0$ et $Q2$; $Q3$ est la note à égale distance entre $Q2$ et $Q4$.

8 Analyse univarié : Bilan

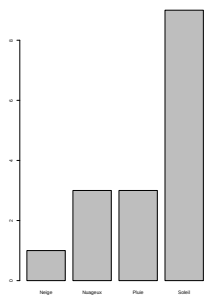
Pour les nominales et ordonnées, on donne :
- Les effectifs de chaque modalité, l'effectif total et le nombre de valeurs manquantes.
- Le mode (c'est à dire la modalité ayant le plus grand effectif).
- On ne donne pas d'indice de dispersions.
- Diagramme en batons.

Pour les discrètes, on donne :
- Les effectifs de chaque modalité, l'effectif total et le nombre de valeurs manquantes.
- Les indices de centralité et de dispersion.
- Diagramme en batons.
- Eventuellement la boite à moustache.

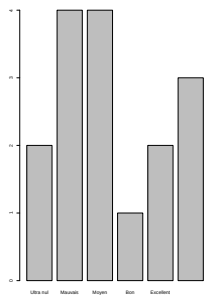
Pour les continues, on donne :
- L'effectif total et le nombre de valeurs manquantes
- Les indices de centralité et de dispersion.
- L'histogramme regroupé par intervalle.
- Une boite à moustache.

9 Exemple

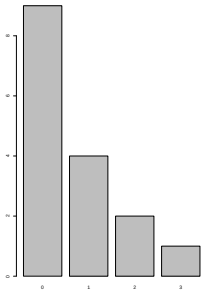
[Match]	[Resultat]	[Points]	[Carton]	[Meteo]	[Commentaire]
1	Perdu	0	0	Soleil	Mauvais
2	Bonus	32	17	Soleil	Excellent
3	Gagné	25	0	Pluie	Mauvais
4	Nul	11	0	Soleil	Moyen
5	Bonus	30	1	Soleil	
6	Bonus	42	0	Pluie	
7	Perdu	6	0	Nuageux	Ultra nul
8	Perdu	22	0	Soleil	Mauvais
9	Gagné	21	1	Nuageux	Mauvais
10	Bonus	31	0	Soleil	
11	Perdu	17	2	Soleil	Mauvais
12	Nul	0	3	Neige	Ultra nul
13	Perdu	15	2	Nuageux	Ultra nul
14	Gagné	27	0	Soleil	Moyen
15	Perdu	18	1	Pluie	Bon
16	Bonus	36	0	Soleil	Excellent

Effectif	**Diagramme en batons**
Neige : 1 (6.25%) Nuageux : 3 (18.8%) Pluie : 3 (18.8%) Soleil : 9 (56.2%)	
Effectif total : 16 ; Valeurs manquantes : 0 (0 %)	

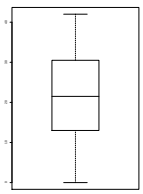
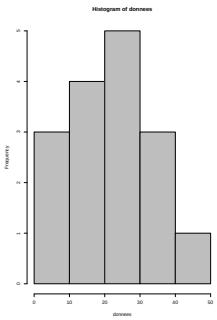
[Meteo] (variable nominale)

Effectif	**Diagramme en batons**
Ultra nul : 2 (12.5%) Mauvais : 4 (25%) Moyen : 4 (25%) Bon : 1 (6.25%) Excellent : 2 (12.5%)	
Effectif total : 16 ; Valeurs manquantes : 3 (18.75%)	

[Commentaires] (variable ordonnée)

Effectifs	**Indices**	**Diagramme en batons**
0 : 9 (56.2%) 1 : 4 (25%) 2 : 2 (12.5%) 3 : 1 (6.25%)	Moy. : 0.688 Var. : 0.896 Ecart : 0.946 Min. : 0 Q1 : 0 Méd. : 0 Q3 : 1 Max. : 3	
Effectif total : 16 ; Valeurs manquantes : 0 (0 %)		

[Carton] (variable dicrète)

Indices	**Box Plot**	**Histogramme**
Moy. : 20.8 Var. : 153 Ecart : 12.4 Min. : 0 Q1 : 14 Méd. : 21.5 Q3 : 30.2 Max. : 42		
Effectif total : 16 ; Valeurs manquantes : 0 (0 %)		

[Point] (variable continue)

10 Statistiques descriptives *VS* statistiques inférentielles

En statistiques descriptives, on dispose de toute la population, on veut la décrire. Exemple : les notes de licences, les accidents du concorde, les températures de l'été.
En statistique inférentielle, on ne connait qu'une partie de notre population, on voudrait connaitre toute la population. Exemple : intentions de vote, audimat, réaction devant un médicament.

11 Loi d'une variable

11.1 Définition

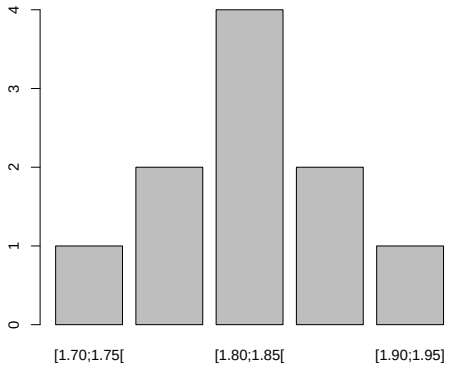
Pour une variable numérique, on regroupe les données par intervalle, puis on fait un histogramme :

Taille
1.72
1.75
1.79
1.80
1.81
1.81
1.84
1.85
1.89
1.91

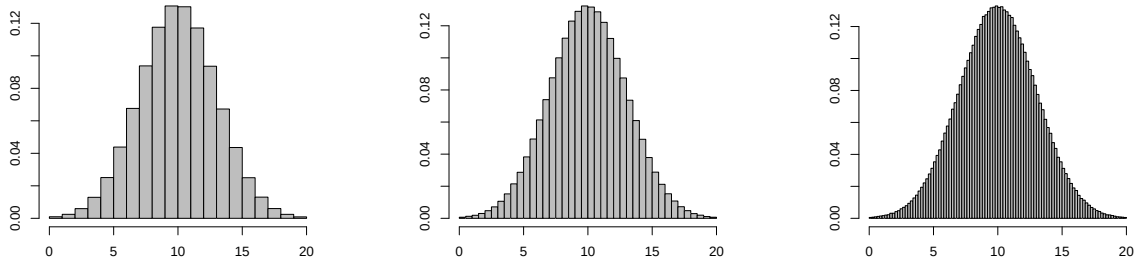
$\Longrightarrow$

Intervalle	Effectifs
[1.70; 1.75[	1
[1.75; 1.80[	2
[1.80; 1.85[	4
[1.85; 1.90[	2
[1.90; 1.95[	1

$\Longrightarrow$



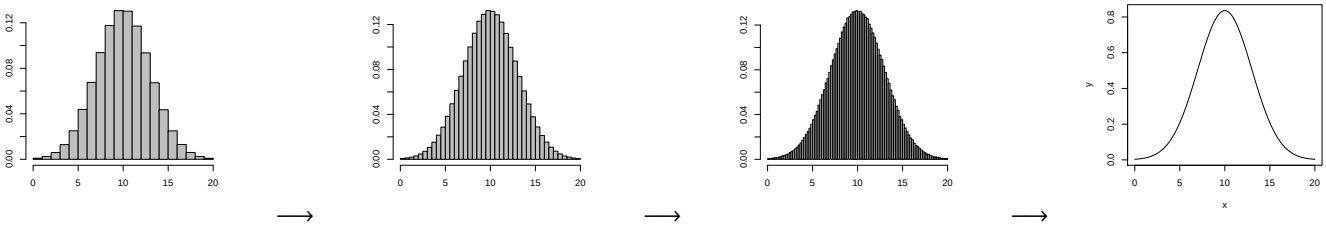
Si on a beaucoup de données, on peut faire des intervalles plus ou moins fin :



20 intervalles 40 intervalles 80 intervalles

FIG. 1 – Différents découpages d'un grand nombre de notes

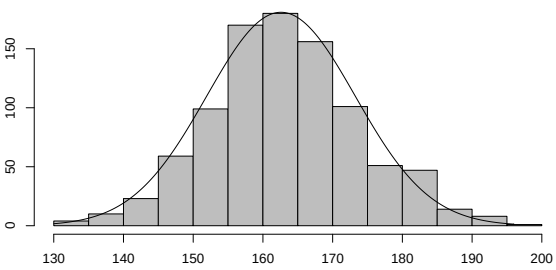
Si on avait une infinité de données, on pourrait faire une infinité d'intervalle tous de taille nulle. On obtiendrait une loi. Les mathématiciens l'ont fait pour nous :



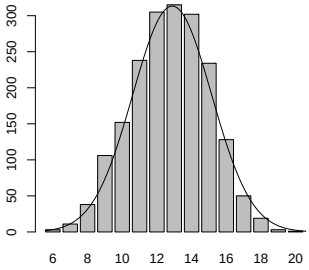
20 intervalles $\longrightarrow$ 40 intervalles $\longrightarrow$ 80 intervalles $\longrightarrow$ Loi normale

FIG. 2 – Évolution vers la loi normale

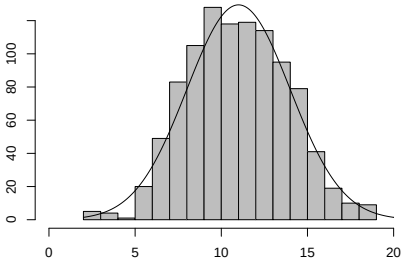
Une loi est une fonction qui ressemble beaucoup aux distributions qu'on trouve dans la nature.



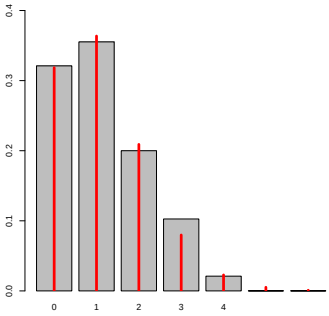
Taille des adolescents



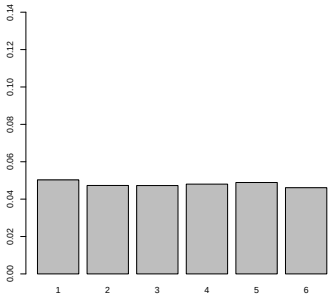
Loges Capsulaires



Note Mini QCM



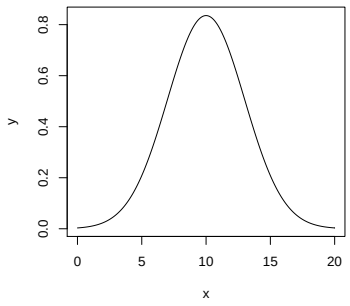
But en foot



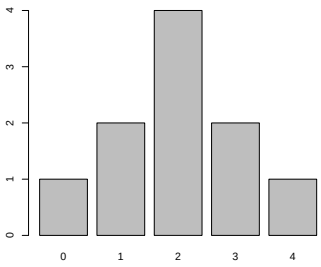
Lancés de dès

11.2 Moyenne, écart type, médiane d'une loi

Une loi provient d'un histogramme qui provient d'une population. Étant donnée une loi, on peut faire le chemin inverse :



$\Longrightarrow$



$\Longrightarrow$

Val.	Eff.
0	1
1	2
2	4
3	2
4	1

$\Longrightarrow$

Taille
1.72
1.75
1.79
1.80
1.81
1.81
1.84
1.85
1.89
1.91

Puis on peut calculer la moyenne, la variance et tous les autres indices de la loi SUR la population ainsi obtenue. Plus les intervalles seront fin, plus la précision sera grande.

11.3 Caractèristique de la loi normale

Il y a une loi qu'on retrouve partout dans la nature : la loi normale. Chaque loi normale à une moyenne et un écart type. La moyenne détermine la position de la loi sur l'axe des X, l'écart type détermine son applatissement.

– La loi normale est symétrique.
– Sa moyenne est l'abscisse du point le plus haut. C'est aussi sa médiane et son mode.
– A mi-pente, elle présente un point d'inflexion (point où la pente s'inverse). Ce point est à une distance de un écart type de la moyenne.

– Elle est infinie à gauche et à droite.
– Chaque loi est définie par sa moyenne et son écart type.

12 Intervalle de confiance pour la loi normale

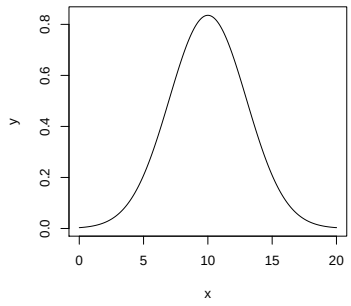
Il y a une loi qu'on retrouve partout dans la nature : la loi normale. Chaque loi normale à une moyenne et un écart type. La moyenne détermine la position de la loi sur l'axe des X, l'écart type détermine son applatissement.

12.1 Variabilité biologique et intervalle de confiance

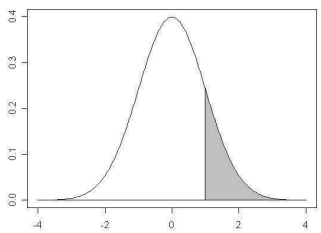
La variabilité biologique est le fait que deux évènements *apparaments* identiques conduisent à deux résultats différents.
Les choses varient, mais on cherche tout de même à deviner le résultat (la valeur que prend une variable). On va non pas donner un chiffre, mais donner une fourchette et une probabilité associée :
– Il y a 95% de chances que la taille d'un individu prit au hasard se situe dans l'intervalle [1.65 ; 1.85].
– Il y a 80% de chances que la croissance du PIB se situe entre 1.9 et 2.1
– Il y a 0.001% de chances que la variation des températures de l'été soit dans l'intervalle [+5 ; +∞]

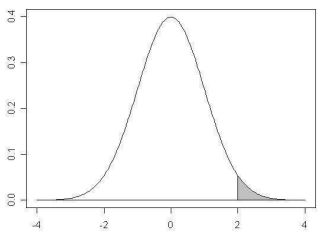
12.2 Loi normale centré réduite

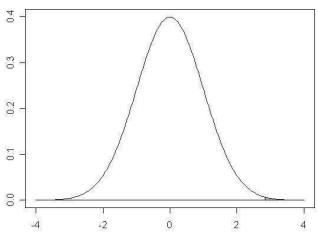
On cherche a deviner la valeur d'une variable. Si on n'a aucune information sur la variable, on ne peut pas deviner. Par contre, si on sait que la variable suit une loi normale, on peut trouver son intervalle de confiance.
Pour commencer, considérons une variable qui suit une loi normale de moyenne 0 et d'écart type 1, par exemple, les températures de décembre.



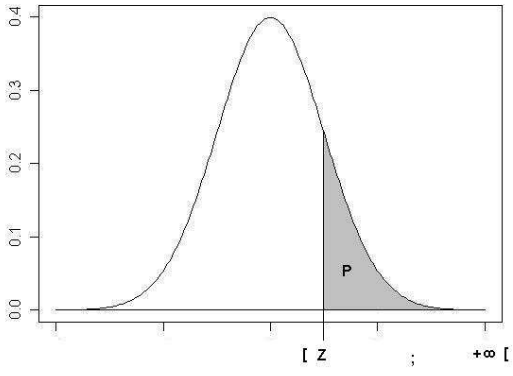
On sait que chaque point sous la courbe représente un individu "théorique" (ici, les individus sont des jours). En tout état de cause, si on choisit un individu au hasard, il y a plus de chance de tomber "au centre", pas trop loin de la moyenne, que proche des bords.
Les mathématiciens ont fait les calculs pour nous :
– l'intervalle [1 ; +∞[contient 0,16% de la population.
– l'intervalle [2 ; +∞[contient 0,025% de la population.
– l'intervalle [3 ; +∞[contient 0,003% de la population.







Pour les autres valeurs, on utilise une table : la table de la loi normale centrée réduite donne la probabilité qu'un individu pris au hasard soit dans l'intervalle $[z; +\infty[$.



Z	P
0,0	0,5%
0,1	0,4602%
0,2	0,4207%
0,3	0,3821%
0,4	0,3446%
0,5	0,3085%
0,6	0,2743%
0,7	0,2420%
0,8	0,2119%
0,9	0,1841%
1,0	0,1587%
1,1	0,1357%
1,2	0,1151%
1,3	0,0968%
1,4	0,0808%
1,5	0,0668%
1,6	0,0548%
1,7	0,0446%
1,8	0,0359%
1,9	0,0287%

Z	P
2,0	0,0228%
2,1	0,0179%
2,2	0,0139%
2,3	0,0107%
2,4	0,0082%
2,5	0,0062%
2,6	0,0047%
2,7	0,0035%
2,8	0,0026%
2,9	0,0019%
3,0	0,0013%
3,1	0,0010%
3,2	0,0007%
3,3	0,0005%
3,4	0,0003%
3,5	0,0002%
3,6	0,0002%
3,7	0,0001%
3,8	0,0001%
3,9	0,000 05%

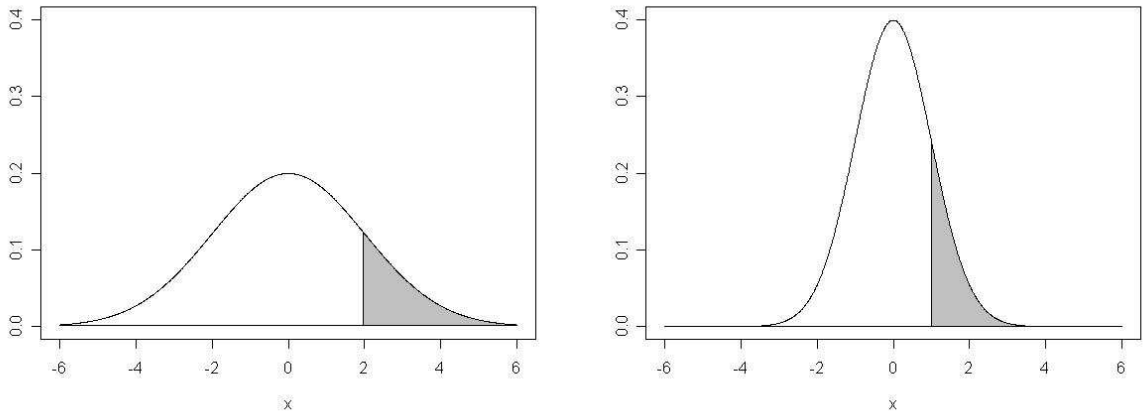
Z	P
4,0	0,000 03%
4,1	0,000 02%
4,2	0,000 01%
4,3	0,000 01%
4,4	0,000 01%
4,5	0,000 003%
4,6	0,000 002%
4,7	0,000 001%
4,8	0,000 001%
4,9	0,000 000 5%
5,0	0,000 000 3%
5,1	0,000 000 2%
5,2	0,000 000 1%
5,3	0,000 000 1%
5,4	0,000 000 03%
5,5	0,000 000 02%
5,6	0,000 000 01%
5,7	0,000 000 01%
5,8	0,000 000 003%
6,0	0,000 000 001%

12.3 Lois normales, autre

Si on ne travaille plus sur la loi de moyenne 0 et d'écart type 1 mais sur une loi normale de moyenne quelconque et d'écart type 1, la position du "centre" change mais le principe est le même. Les notes de licences suivent une loi normale de moyenne 12 et d'écart type 1, on sait que :
– l'intervalle [13 ; +∞[contient 0,16 (=16%) de la population.
– l'intervalle [14 ; +∞[contient 0,025% (=2,5%) de la population.
– l'intervalle [15 ; +∞[contient 0,003% (=0,3%) de la population.

Comment le sait-on ? On "déplace" la loi normale de moyenne 12 et d'écart type 1 vers la gauche jusqu'à la superposer avec la loi normale centrée réduite. Plus précisément, l'intervalle [13 ; +∞[sur la loi normale $\mathcal{N}(12, 1)$ contient le même pourcentage de la population que l'intervalle [1 ; +∞[sur la loi normale centré réduite $\mathcal{N}(0, 1)$.
De manière plus générale, l'intervalle [12+z ; +∞[sur la loi normale $\mathcal{N}(12, 1)$ contient le même pourcentage de la population que l'intervalle [z ; +∞[sur la loi normale centrée réduite.
De manière encore plus générale, l'intervalle [m+z ; +∞[sur la loi normale $\mathcal{N}(m, 1)$ contient le même pourcentage de la population que l'intervalle [z ; +∞[sur la loi normale centrée réduite.

Si on ne travaille plus sur une loi normale d'écart type 1 mais sur une loi d'écart type quelconque, l'applatissement de la loi normale change. Pour la faire ressembler à la loi normale centré réduite, il faut non seulement la déplacer, mais également "l'étirer" (si elle est étroite) ou la "rétrécir" (si elle est large).
Par exemple, l'intervalle [2 ; +∞] sur la loi normale $\mathcal{N}(0, 2)$ contient le même pourcentage de la population que l'intervalle [1 ; +∞] sur la loi normale $\mathcal{N}(0, 1)$.



Sur les deux lois, la partie grisée a la même surface.

Plus généralement, l'intervalle [s ; +∞[sur la loi normale $\mathcal{N}(0, s)$ contient le même pourcentage de la population que l'intervalle [1 ; +∞[sur la loi normale $\mathcal{N}(0, 1)$. Encore plus généralement, l'intervalle [zs ; +∞[sur la loi normale $\mathcal{N}(0, s)$ contient le même pourcentage de la population que l'intervalle [z ; +∞[sur la loi normale $\mathcal{N}(0, 1)$.

Au final, si on travaille avec une loi normale quelconque $\mathcal{N}(m, s)$:

L'intervalle $[m + zs; +\infty[$ **sur** $\mathcal{N}(m, s)$
contient le même pourcentage de la population que
l'intervalle $[z; +\infty[$ **sur** $\mathcal{N}(0, 1)$

Une autre manière d'écrire exactement la même chose et de poser $x = m + zs$ (et donc $z = \frac{x-m}{s}$). On a alors

L'intervalle $[x; +\infty[$ **sur** $\mathcal{N}(m, s)$
contient le même pourcentage de la population que
l'intervalle $[z; +\infty[$ **sur** $\mathcal{N}(0, 1)$

12.4 Exemples

1. La moyenne de l'amphi est 9, écart type 3. Les notes suivent une loi normale.
2. Un étudiant a 16,5 ; il demande une lettre de recommandation.
3. Formulation du statisticien : quel pourcentage de la population contient l'intervalle $[16, 5; +\infty[$?

Toute la difficulté est simplement de changer les notations :

1. m=9 ; s = 3
2. x=16,5
3. Solution : on calcule $z = \frac{16,5-9}{2} = 2, 5$, puis on cherche le P associé sur la table. $\boxed{Z = 2, 5 \Longrightarrow p = 0, 0062}$

Un étudiant prit au hasard a 0,62% de chance d'être dans l'intervalle [16.5 ;20]. On peut donc considerer un étudiant qui a 16.5 ou plus comme très bon.

Problème : la taille des francaises suit une loi normale de moyenne 162cm écart type 6cm. Quel pourcentage de la population mesure plus de 170cm ?
$z = \frac{170-162}{6} = 1.33$ donc $p = 0.0968$. Environ 10% des françaises mesurent 1,70m et plus.

12.5 Manipulation d'intervalles

On travaille sur la loi centré réduite.
– L'intervalle $]-\infty; +\infty[$ contient 100% de la population.
– Si z est positif, l'intervalle $[z; +\infty[$ contient p% de la population, on le trouve grace à la table de la loi centrée réduite.
– Si z est positif, l'intervalle $]-\infty; z]$ contient 100% - p% de la population.

- Si z est négatif, l'intervalle $]-\infty; z]$ contient le même pourcentage que l'intervalle $[-z; +\infty[$.
- Si z est négatif, l'intervalle $[z; +\infty[$ contient le même pourcentage que l'intervalle $]-\infty; -z]$.
- Au final, l'intervalle [z ; z'] contient le pourcentage de l'intervalle $[z; +\infty[$ moins le pourcentage de l'intervalle $[z'; +\infty[$.