VARIABLES ALÉATOIRES

1

Structure du cours

Définition
Types de variables aléatoires
Lois de distribution
Distributions théorétiques:
- Modèle, espérance (moyenne), variance, applications, exemples
- discrets: Bernoulli, Binomiale, Poisson
- continues: Normale, Student (t), Khi carré, Fisher

2

L’utilite des variables aleatoires

A l’aide de la **statistique descriptive** on peut **décrire** des **phénomènes variables**. A ces phénomènes **on associe des variables aléatoires**.

Pour extrapoler les résultats obtenus **sur un échantillon** extrait d’un population **a la population entière**, on a **besoin** de savoir certaines **lois des probabilités** (les probabilités des résultats possibles des phénomènes)

Ces probabilités vont nous aider a:

- Identifier un intervalle des valeurs qui va contenir un résultat avec une certaine probabilité (95%)(voir le cours sur les intervalles de confiance)

- Identifier quelle est la probabilité de dire que la différence entre les « moyennes » des deux populations sont différentes si en réalité ils ne sont pas, puis vérifier si cette probabilité est < 5%. (voir les cours des tests statistiques)

Ces situations ont besoin de savoir la probabilité d’avoir des valeurs dans certaines intervalles des valeurs.

Ex: pour des intervalles de confiance: on a besoin de la probabilité cumulée d’avoir des valeurs inferieure a 1,96. Ou, plus pratique: on a besoin de savoir pour quelle valeur critique la probabilité d’avoir des valeurs inferieure a cette valeur critique est une demie de 95% (47,5%)

Ex: pour des test statistiques: on a besoin de la probabilité cumulée d’avoir des valeurs supérieure a 1,96. Ou, plus pratique: on a besoin de savoir pour quelle valeur critique la probabilité d’avoir des valeurs supérieure a cette valeur critique est une demie de 5% (2,5%)

3

Variables aléatoires

Une **expérience (épreuve) peut être aléatoire** ou non.

O peut considérer un expérience aléatoire si son résultat est imprédictible.

Même si on répète l’expérience dans des conditions identiques on peut observer des résultats différents.

Exemples des **expériences**:

- le jeux de dé, mesurer la température d’un individu, mesurer le taux du cholestérol d’un patient, savoir si le patient a du diabète ou non, s’il est fumeur ou pas

Le **résultat d’une expérience** est un événement élémentaire.

- ex. Pour le jeux de dé: la face 1, 2, …; pour la température: 36,5 ou 38,2;
- pour la présence du diabète : oui/non, …

L’ensemble des toutes événements élémentaires possibles d’une expérience aléatoire représente **l’espace fondamental.**

- ex. Pour le jeux de dé: l’ensemble {1, 2, 3, 4, 5, 6},

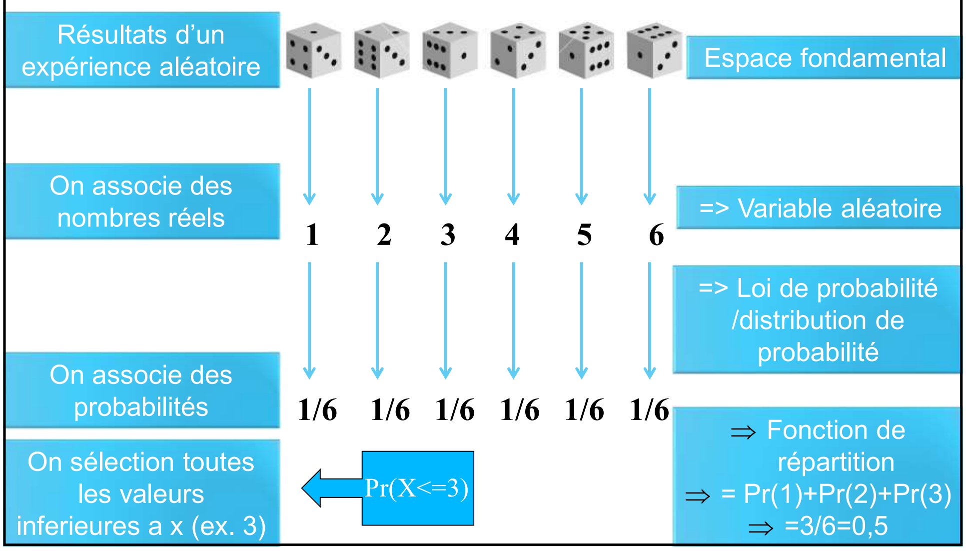
4

Variables aléatoires

- Etant donnée E – **espace fondamental** qui correspond a un **expérience aléatoire**.
- **Définition**:
  - Une **variable aléatoire** (v.a.) dans un espace fondamental E, noté X est un fonction définit sur E qui prend des valeurs dans l'ensemble des nombres réels
  - Une variable aléatoire X est une variable numérique qui prend différents valeurs avec des probabilités spécifiées
- A une variable aléatoire X on peut associées des probabilités avec lesquelles cette variable aléatoire peut prendre certaines valeurs – ca représente la lois de distribution (la distribution des probabilités)
  - Pr(X = a) – la probabilité que X va prendre la valeur a
  - Pr(a <= X <= b) – la probabilité que X prends des valeur dans l'intervalle [a; b] – Ex. [- 1,96; 1,96]
  - Pr(X <= a) – la probabilité que X prends des valeur <= a. Ex. <= 1,96

5

Synthèse des notions



6

Pour expériences avec valeurs binaires

Résultats d'un expérience aléatoire
Réponse a une question: fumeur?

oui → 1 "succes" → 0,25

non → 0 "echec" → 0,75

Espace fondamental

On associe des nombres réels

=> Variable aléatoire

On associe des probabilités

=> Loi de probabilité /distribution de probabilité

7

Pour expériences avec valeurs quantitatives discrètes

Résultats d'un expérience aléatoire
Une mesure: pouls?

... 64   65   66 …

Espace fondamental

On associe des nombres réels

... 64   65   66 …

=> Variable aléatoire

On associe des probabilités

... 0,07   0,06   0,06 …

=> Loi de probabilité /distribution de probabilité

On sélection toutes les valeurs inferieures a x (ex. 90)

Pr(X<=90)

⇒ Fonction de répartition
⇒ = 0,8

https://upload.wikimedia.org/wikipedia/commons/d/d2/FingertipPulseOximeter300C2.jpg

8

Pour expériences avec valeurs quantitatives continues

Résultats d'un expérience aléatoire
Mesure du tension artérielle systolique

0,…,100,…,120,…,200,…

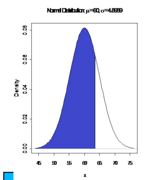
Espace fondamental

=> Variable aléatoire

On associe des nombres réels

0,…,100,…,120,…,200,…

On associe des probabilités pour des intervalles des valeurs



=> Loi de probabilité /distribution de probabilité

On sélection toutes les valeurs inferieures a x

Pr(X<=120)

⇒ Fonction de répartition

Les Types des variables aléatoires

Discrète

- pour laquelle il y a un ensemble discret de valeurs
  - avec des probabilités spécifiées
- la variable prends un nombre finit
  - ou au moins dénombrable des valeurs

Ex. Nombre des sujets qui sont végétariens, ont arthrose du genoux, ont du fièvre.

Les Types des variables aléatoires

Continue

- pour laquelle il y a un ensemble continu de valeurs,
  - les domaines de valeurs ont des probabilités spécifiées
- elle varie dans un mode continu dans un intervalle
- peut prendre un ensemble non dénombrable des valeurs
  - Ex:
    - concentration d'un substance dans le sang,
    - la température corporelle
    - l'allongement chirurgicale du mandibule/tibia

11

11

Variables aléatoires discrètes
Loi de probabilité finie

DEFINITION:

Loi de probabilité finie (distribution de probabilité finie)

- règle par laquelle on assignent a chaque valeur possible r d'une variable aléatoire (v.a.) discrète X la probabilité Pr(X = r),

Exemple 1:

- La probabilité d'apparition d'un **face d'un dés** est 1/6
- X: 1 2 3 4 5 6

X:	1	2	3	4	5	6
Pr	1/6	1/6	1/6	1/6	1/6	1/6

Exemple 2:

La distribution de probabilité d'apparition d'un **résultat positif d'un traitement** qui a été donné a 4 sujets (0 sur 4, 1 sujet sur 4, 2 sujets sur 4 … :

X:	0	1	2	3	4
Pr	0,01	0,08	0,26	0,42	0,23

12

Variables aléatoires discrètes
Loi de probabilité finie

- **Propriétés**
- $0 < \Pr(X = r) \leq 1$ ex: 0,01 0,08 0,26 0,42 0,23
- $\sum \Pr(X = r) = 1$ ex: 0,01+0,08+0,26+0,42+0,23 = 1
- La **fonction de répartition** pour un v.a. X $F : R \to R$
- Est un probabilité cumulé
- Repesent la somme des toutes probabilités pour les valeurs y du v.a. X inferieures ou égale a x

$$\Pr(X \leq x) = \sum_{y \leq x} p(y)$$

$$F(x) = \Pr(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

13

Variables aléatoires discrètes
Esperance, variance et écart type

- **Esperance mathématique** ou la **moyenne** théorique d'une v.a. X

$$E(X) = \sum_{i=1}^{n} x_i \Pr(X = x_i) \sim \boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

Exemple: avec le traitement

- X: 0 1 2 3 4

Pr 0,01 0,08 0,26 0,42 0,23

E(X) = 0 x 0,01 + 1 x 0,08 +2 x 0,26 + 3 x 0,42 + 4 x 0,23 = 2,78

- **Variance** de X:

$$V(X) = \sum_{i=1}^{n} \left[x_i - E(X)\right]^2 * \Pr(x_i) \sim \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

- **Ecart-type** (deviation standard) de X

$$\sigma(X) = \sqrt{V(X)} = V(X)^{1/2}$$

14

Variables aléatoires
Lois des distribution

En général on ne connait pas les loi des distributions des variables
dans le domaine médical/médical dentaire/biologique, …
On cherche d'encadrer ces distributions dans des distributions théorétiques.
Ces distributions théorétiques sont des modèles avec lesquelles on va travailler pour estimer ce que se passe dans les populations

Distributions théorétiques:
discrets:
- Bernoulli
- Binomiale
- Poisson

continues:
- Normale
- Student
- Khi carré
- Fisher

15

Variables aléatoires discrètes
Loi Bernoulli

- **B (1, p)**
  - L'expérience consiste dans **1** épreuve
  - La épreuve a seulement les résultats 1 ("succès") et 0 ("échec")
  - La probabilité de "succès" est **p** et "d' échec" est 1 – p = **q**
  - La variable aléatoire X = {0, 1}

  Pr(succès) = p, Pr(échec) = q = 1 -p
  - Exemple:
    - avec/sans obésité, fumeur/non fumeur, avec/sans ulcère gastrique
  - C'est un cas spécial du distribution Binomiale avec n = 1

16

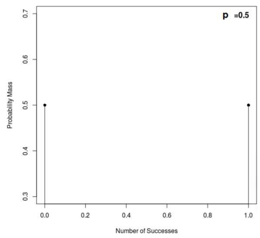
Variables aléatoires discrètes
Loi Bernoulli

Distribution de X:

$$\Pr(X=1)=1-\Pr(X=0)=p=1-q$$

l'espérance mathématique: E(X) = p

variance: Var (X) = p * q = p * (1 – p)



écart type (déviation standard)

$$\sigma\ (X)=\sqrt{Var(X)}=\sqrt{p\times q}=\sqrt{p\times(1-p)}$$

17

Variables aléatoires discrètes
Loi Binomiale

- **Bi(n, p)**
  - L'expérience consiste dans n épreuves indépendantes
  - Une épreuve a seulement les résultats "succès" et "échec"
  - La probabilité de "succès" est **p** et "d' échec" est 1 – p = **q**

  Pr(succès) = p, Pr(échec) = q = 1 - p
  - La variable aléatoire X = le nombre des "sucées" = {0, 1, 2, … , n}
  - Exemple:
    - Le nombre de personnes Avec obésité, fumeur, avec diabète, avec hypertension, avec caries parmi 100 sujets

- Application:
  - test binomial pour un proportion
  - intervalle de confiance pour un proportion

18

Loi Binomiale

Application:

- Intervalle de confiance pour un proportion
  - Exemple: **on prend un échantillon** d'une population des personnes âgées .
    - On **compte combien** des **sujets** ont une prothèse vasculaire **dans** un **échantillon.**
    - On **trouve** le **pourcentage** (4%).
    - **Notre question** est: quelle est **le vrai pourcentage dans la population**, et on s'intéresse de l'**intervalle dans lequel on va trouver la vrai pourcentage avec un probabilité de 95%** (par convention).
    - Pour trouver cet intervalle **on a besoin** de **savoir quelle est la lois de probabilité de avoir des prothèses (succès) parmi n sujets (épreuves).**
    - Dans ce cas on peut **observer** qu'il **ressemble** a un **expérimente binomial.**
    - Ici le **succès** est la **présence de la prothèse**, et l'échec est l'absence.

19

Variables aléatoires discrètes
Loi Binomiale

On peut calculer la probabilité qu'un variable aléatoire est égale a un certain nombre des "succès" (k) dans un certain nombre des expérimentes (n)

$$\mathbf{C_n^k = \frac{n!}{k!(n-k)!}}$$

ex: $3! = 3 * 2 * 1 = 6$

Distribution de X:

$$\mathbf{Pr(X = k) = C_n^k p^k q^{n-k}}$$

n – nombre des expériences, k nombre des résultats «succès»

l'espérance mathématique: E(X) = n * p

variance: Var (X) = n * p * q = n * p * (1 – p)

écart type (déviation standard)

$$\sigma\ (X) = \sqrt{Var(X)} = \sqrt{n \times p \times q} = \sqrt{n \times p \times (1-p)}$$

20

Variables aléatoires discrètes
Loi Binomiale

Exemple:

Quelle est la probabilité d'avoir 2 garçons dans une famille avec 5 enfants.

On connait que:

- a chaque naissance la probabilité d'avoir un garçon est 0.51
- et les sexes d'enfants consécutifs - des événements indépendantes

Solution:

En utilisant une distribution binomiale avec le paramètres

n = 5 et p = 0.51 pour k = 2, nous avons:

$$\Pr(X = 2) = C_5^2 (0.51)^2 (0.49)^3 = \frac{5!}{2! \times 3!} (0.51)^2 (0.49)^3 = 0.306$$

21

Variables aléatoires discrètes Loi Binomiale

M = 20 * 0,9 = 18

V = 20 * 0,9 * 0,1 = 1,8

σ = 1,34

M = 100 * 0,9 = 90

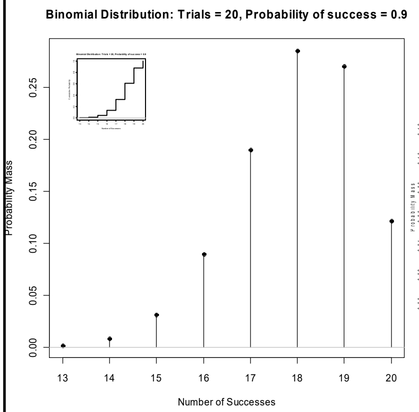
V = 100*0,9*0,1 = 9

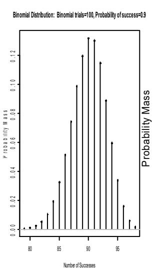
σ = 3

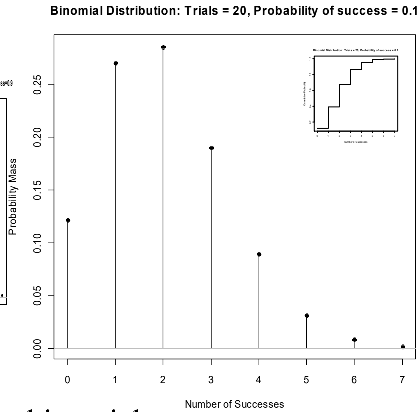
M = 20 * 0,1 = 2

V = 20 * 0,1 * 0,9 = 1,8

σ = 1,34







Voici deux simulations de 20 expérimentes binomiales avec un probabilité de succès p = 90% a la gauche et p = 10% a la droite. 22
La variance est identique, mais la moyenne est différente.

22

Variables aléatoires discrètes Loi Binomiale

M = 20 * 0,6= 12

V = 20 * 0,6 * 0,4 = 4,8

σ = 2,19

M = 100 * 0,6 = 60

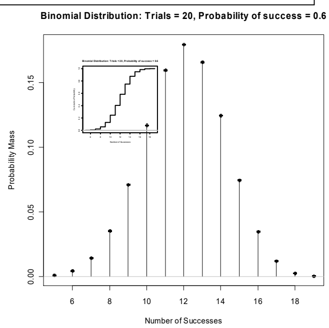
V = 100*0,6*0,4 = 24

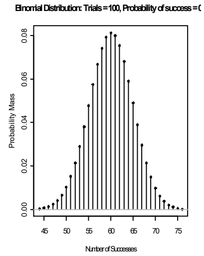
σ = 4,9

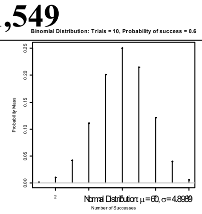
M = 10 * 0,6 = 6

V = 10 * 0,6 * 0,4 = 2,4

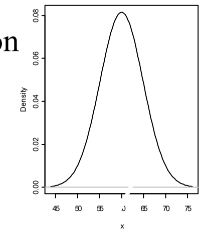
σ = 1,549







distribution Normale



Si np >= 10 et nq >= 10 la distribution Binomiale tend a un distribution Normale N(np,√npq)

Plus p est proche a les extrêmes (0 ou 1), plus la variance ou l'écart type est petit

Plus p est proche a 0,5 plus la variance ou l'écart type est grande

23

Variables aléatoires discrètes Loi de Poisson

- La variable aléatoire Poisson:
  - variable discret qui a un infinité dénombrable des valeurs 0, 1, …
- Représente le **nombre des réalisations dans un intervalle de temps ou d'espace**
  - nombre des entrées dans un cabinet/nombre interventions/nombre extractions,…
  - nombre des bactéries dans un millilitre d'eau, nombre des cellules blanc dans 1 ml de sang…
- On connais le nombre **attendu** λ d'événements dans une intervalle unité
- La variable aléatoire X = le nombre d'événements ayant lieu (**observe**) dans une intervalle unité
- **Application**:
  - Prédiction des phénomènes
  - Régression Poisson
    - Prédire le nombre des malades dans une région, dans une unité de temps

24

Variables aléatoires discrètes
Loi de Poisson

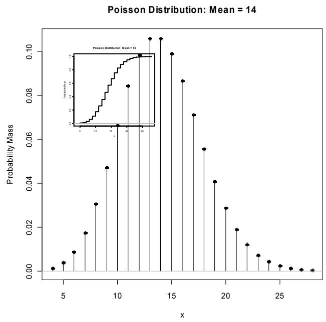
Distribution:

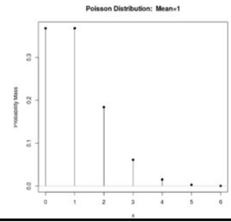
Pr(X=k) = $e^{-\mu}\mu^k/k!$

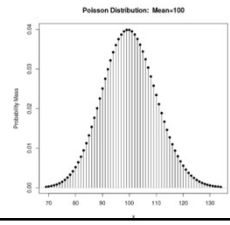
$\mu = \lambda$ t, e = 2,71828

l'espérance mathématique: E(X) = μ

variance: Var (X) = μ







25

Variables aléatoires discrètes
Loi de Poisson

Le nombre des naissances dans un maternité d'un hôpitaux dans un année est 1000. La moyenne des naissances par jour est 1000/365, donc 2,74.

On peut calculer la probabilité d'avoir 0, 1, 2, … naissances par jour, avec la distribution du Poisson. Ex:

P(0) = $e^{-2,74}$ $2,74^0$/ 0! = 0,065

P(1) = $e^{-2,74}$ $2,74^1$/ 1! = 0,177

P(2) = $e^{-2,74}$ $2,74^2$/ 2! = 0,242

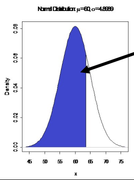
P(3) = $e^{-2,74}$ $2,74^3$/ 3! = 0,221 (22,1%)

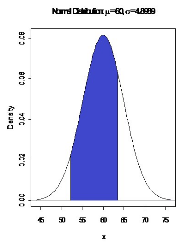
26

Variables aléatoires continues
Densité de Probabilité

- La **fonction densité de Probabilité** f(x) d'une v.a. X
- est une fonction f positive telle que:
  - Pr(a < X <= b) = la surface entre le graphique f et les droites x = a et x =
  - La surface totale en dessous du graphique $\mathbb{P}(a < X \leq b) = \int_a^b f(u)\, du.$
- La **fonction de répartition F** associe a la v.a. X
  - F(x) = Pr(X ≤ x)

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)du.$$





Variables aléatoires continues Esperance mathématique et Variance

- Esperance mathématique :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$$E(X) = \sum_{i=1}^{k} x_i \Pr(X = x_i)$$

- Variance:
  - Var(X)= σ^2 = $V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx$

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

$$V(X) = \sum_{i=1}^{n}[x_i - E(X)]^2 \times \Pr(x_i)$$

- Ecart type

$$\sigma(X) = \sqrt{Var(X)} = Var(X)^{1/2}$$

Variables aléatoires continues
Loi Normale (GAUSS)

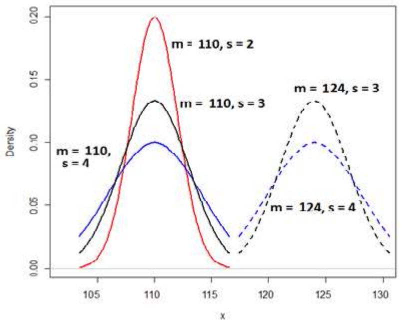
Model: **N(μ, σ^2),** dépend de la moyenne et la variance/déviation standard

la concentration des substances dans le sang (glicemie, trigilicerides…

l'erreur de mesure d'un certain objet, quantité (hauteur…), etc.

Fonction densité de probabilité:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



$E(X) = \mu$

$V(X) = \sigma^2$

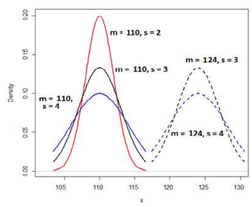
Applications:

le test Z pour comparaison des moyennes/fréquences

intervalles de confiance pour les moyennes/fréquences

29

Variables aléatoires continues
Loi Normale centrée réduite



Pour simplifier l'utilisation de la lois Normale

On fait une transformation $Z = \frac{X - \mu}{\sigma}$

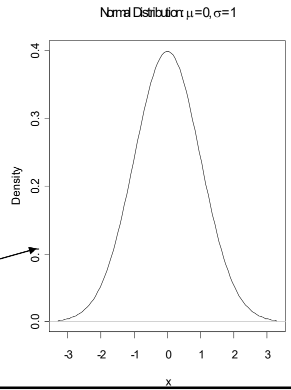
Donc pour la variable aléatoire Z:

$\mu = 0$

$\sigma = 1$

sans unité de mesure

La fonction densité de probabilité:



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \rightarrow f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

30

Variables aléatoires continues
Loi Normale centrée réduite

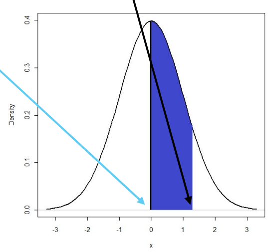
Associée a la lois normale centrée réduite avec la densité de probabilité… il y a un nombre des tableaux qui permet un utilisation pratique de ce lois. Elle sont calculées depuis la fonction de densité de probabilite

Ce tableau donne l'aire sous la courbe normale centrée réduite, entre 0 et un valeur Z

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	[illegible]	[illegible]	[illegible]	[illegible]	0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	[illegible]	[illegible]	[illegible]	[illegible]	0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	[illegible]	[illegible]	[illegible]	[illegible]	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	[illegible]	[illegible]	[illegible]	[illegible]	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	[illegible]	[illegible]	[illegible]	[illegible]	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	[illegible]	[illegible]	[illegible]	[illegible]	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916

deuxième décimale

Première décimale



1DS-0,34, 2*1DS-0,68
1,96DS – 0,475
2*1,96DS – 0,95

Variables aléatoires continues
Loi Normale centrée réduite

Propriétés de la distribution Normale (Gauss)

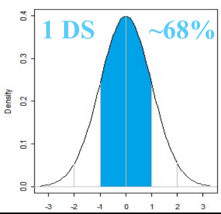
Symétrique, unimodale, en forme de cloche

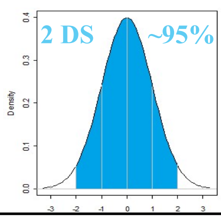
Dans l'intervalle m +/- **1** DS on trouve **~68**,3 **%** de la population.

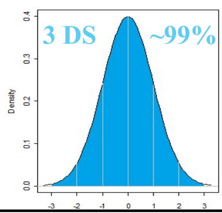
Dans l'intervalle m +/- **2** DS on trouve **~95**,5 **%** de la population.

Dans l'intervalle m +/- **3** DS on trouve **~99**,7 **%** de la population.

(m – moyenne, DS – déviation standard)







Variables aléatoires continues
Loi Normale centrée réduite

Application de la lois Normale:

La description des données

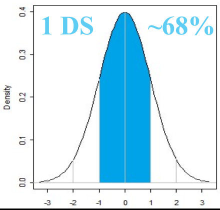
Si vous lisez un article scientifique beaucoup des statistiques sont écrites comme moyennes et déviations standard.

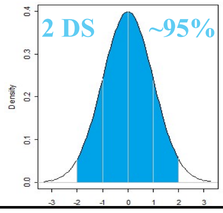
Exemple:

- le poids des malades a été 78,6 kg en moyenne, avec un écart type de 8,3.

Ca signifie que

- entre 78,6 (la moyenne) plus ou moins 8,3 (1 écart type) on trouve 68% des poids des sujets dans l' étude.
- entre 78,6 plus ou moins 8,3*1,96 (approximatif 2 écart type) on trouve 95% des poids des sujets dans l' étude.





33

Loi Normale (GAUSS)

Applications:

- Intervalle de confiance pour un moyenne

  Exemple: on prend un **échantillon** d'une population des personnes âgées .

  On **mesure** le taux du **cholestérol**. Notre **question** est: quelle est la **vrai moyenne dans la population**, et on s'intéresse a calculer un **intervalle** dans lequel on va trouver la vrai moyenne avec un probabilité de **95%** (par convention).

  Pour trouver cet intervalle on a besoin de savoir **quelle est la lois de probabilité dans la population**.

  On observe qu'il y a une variable continue, et on pense a la lois Normale**. On vérifie la normalité** des données. **Si** le données sont **proche** a la distribution normale (histogramme symetrique, …) on considère qu'**on peut utiliser cette lois**.

  On **calcule** la **moyenne** (ex. 184,3) et la **déviation standard dans l'échantillon**

  Avec la moyenne et l' écart type du cholestérol (si on suppose que le taux du cholestérol est normale distribuée), en utilisant la lois normale, **on peut trouver que la vrai moyenne est compris entre** 164,3 et 204,3 **avec une probabilité de 95%.**

34

Loi Normale (GAUSS)

Applications:

- **Intervalle de confiance pour un fréquence** (voir la lois Binomiale) –

c'est identique. D'habitude si le nombre des sujets (expérimentes) est réduit on utilise la loi Binomiale mais s'il est grand on utilise la lois Normale.

- **Le test Z pour comparaison des moyennes**

Exemple: On donne a l'un des **deux échantillons** un traitement hypocholestérolémiant. On **mesure** le taux du **cholestérol**.

On **calcule** les **moyennes** des deux échantillons (ex. 174,1 pour le group avec le traitement, et 194,4 pour le group sans traitement).

Notre **question** est: est qu'**il y a une différence réelle** entre les deux groups? Ou mieux: est qu'il y a une différence réelle **entre les deux les populations?** Ou en réalité les deux populations (traitée et non traitée) sont identiques?

35

Loi Normale (GAUSS)

Applications:

On peut calculer quelle est la **probabilité** d'avoir **une différence identique a celle qu'on a observe** 194,4 – 174,1, ou plus grande qu'elle, **si dans la réalité dans les populations depuis lesquelles on a extrait nos deux échantillons** (traitée et non traitée) **la différence est zero** (0). Pour trouver cet probabilité on a besoin de savoir quelle est la lois de probabilité.

On observe qu'il y a une variable continue, et on pense a la lois Normale. **On vérifie la normalité des données**. **Si** le données sont **proche** a la distribution normale on considère qu'**on peut utiliser cet lois**.

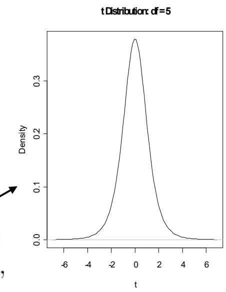
On peut trouver que cette probabilité est **0.03 donc 3%.**

Par convention cette probabilité **est petite (inferieure a 0.05)**, et **on peut conclure que en fait il y a une différence statistiquement signifiante entre les deux populations**.

36

La loi de Student t_n

- La variable aléatoire Student t_n
  - variable aléatoire continue – pour estimer la moyenne d’un echantillon
  - qui prend des valeurs dans l’intervalle $(-\infty, +\infty)$
  - dépend d’un paramètre n appelée degrés de liberté (d.d.l).
- Soit $X_0, X_1, \ldots, X_n$ n variables aléatoires indépendantes
  - qui suive la lois normale centrée réduite.
- La variable aléatoire T_n suive
  - un lois de probabilité Student avec n dégrée de liberté



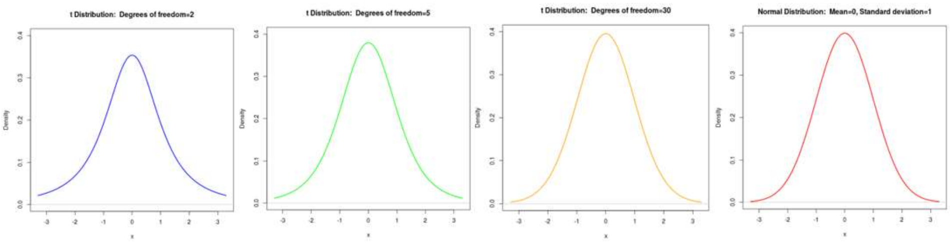
- Ou ν c’est le nombre d.d.l et Γ c’est la fonction gamma.
- Fonction densité de probabilité:
- $P[\, t_k < -x \,] = P[\, t_k > x],$

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2},$$

37

La loi de Student t_n

- **$t_n \to N(0, 1)$, quand $n \to \infty$,**
- **Si n > 30 la lois Normale et la lois Student sont très proches**
- Student (n = 2) Student (n = 5) Student (n = 30) Normale



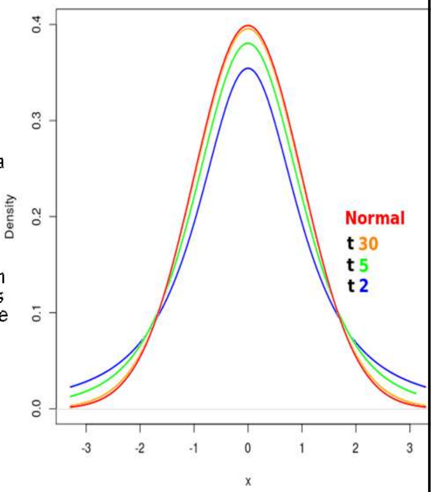
On peut observer si le nombre des degrés de liberté (n) est égal ou plus grand que 30, la loi de Student tend vers un lois Normale.
Donc on peut utiliser le test Student pour des échantillons réduits, et puis on peut approximer avec le test Z (qui utilise la lois Normale).
Mais on va pas faire ca parce que le test Z a besoin de savoir la déviation standard dans la population (qui d’habitude n’est pas connue)
C’est plus simple d’utiliser le test Student pour toutes les types d’ échantillons (petits ou grands)

38

La loi de Student t_n

Propriétés de la distribution t
symétrique
unimodale
en forme de cloche
les queues sont plus longues que les queues de la distribution normale

Donc elle est plus indiquée pour des situations quand on a une petit échantillon. Quand on a données limitées, on n'a pas confiance que les résultats sont surs. Donc quand on cherche a deviner qu'est que se passe dans la population, si l' échantillon est petit, on a besoin des intervalles de confiance plus larges que ceux offert par la lois normale, pour indiquer la manque de confiance dans les données en quantité limitée. La lois student offre des intervalles de plus en plus larges, plus le nombre des sujets (et les dégrées de liberté) sont réduits



39

La loi de Student t_n

- **Applications:**
- Intervalle de confiance
  - d'une moyenne pour un échantillon de taille n
    - avec n-1 dégrées de liberté (voir cours)
  - d'une différence entre deux moyennes pour deux échantillons indépendants
    - avec $n_1 + n_2 - 2$ dégrées de liberté
- Le test de Student pour la comparaison de moyennes,
  - Test t pour un echantillon (voir cours)
  - Test t pour échantillons appariés, (voir cours)
    - (on comparer le même échantillon (n sujets) avant et après un intervention
    - avec n - 1 dégrées de liberté
  - Test t pour des échantillons indépendants (avec n_1, et n_2 sujets) (voir cours)
    - avec $n_1 + n_2 - 2$ dégrées de liberté
- Le test Student pour le coefficient de corrélation (voir cours)

40

40

La loi de Student t_n

Applications:

- Intervalle de confiance pour un moyenne

  Exemple: on prend un échantillon de 25 sujets d'une population des personnes âgées .

  On mesure le taux des triglycérides. On calcule la moyenne (ex. 162,8).

  Notre question est: quelle est le vrai moyenne dans la population.

  On peut calculer un intervalle dans lequel on va trouver la vrai moyenne avec un probabilité de 95% (par convention).

  On observe qu'il y a une variable continue, et on pense a la lois Normale. On vérifie la normalité des données. Si le données sont proche a la distribution normale on considère qu'on peut utiliser cet lois. Mais on n'a pas la moyenne et l'écart type dans la population. En plus l' échantillon est petit < 30. On va utiliser la lois t du Student.

  Avec la moyenne et l' écart type du cholestérol (si on suppose que le taux du cholestérol est normale distribuée), en utilisant la lois t, on peut trouver que la vrai moyenne est compris entre 152,8 et 172,8 avec une probabilité de 95%.

41

41

La loi de Student t_n

Applications:

- Le test Student pour comparaison des moyennes des échantillons indépendants

  Exemple: On donne a l'un des deux échantillons (de 12 sujets) un traitement et un placebo pour baiser le taux des triglycérides. On mesure le taux du triglycérides.

  On calcule les moyennes des deux échantillons (ex. 145,1 pour le group avec le traitement, et 187,9 pour le group sans traitement).

  Notre question est: il y a une différence réelle entre les deux groups? Ou mieux il y a une différence réelle entre les deux les populations? Ou en réalité les deux populations (traitée et non traitée) sont identiques?

42

42

La loi de Student t_n

Applications:

- Le test Student pour comparaison des moyennes des échantillons indépendants

  On peut calculer quelle est la probabilité d’avoir une différence identique a celle qu'on a observe 187,9 – 145,1, ou plus grande qu’elle, si dans la réalité dans les populations depuis lesquelles on a extrait nos deux échantillons (traitée et non traitée) la différence est en réalité nulle (0).

  On observe qu’il y a une variable continue, et on pense a la lois Normale. On vérifie la normalité des données. Si le données sont proche a la distribution normale on considère qu’on peut utiliser cet lois. Mais on n’a pas la moyenne et l’écart type dans la population. En plus l’ échantillon est petit < 30. On va utiliser la lois t du Student.

  Avec la moyenne et l’ écart type du cholestérol (si on suppose que le taux du cholestérol est normale distribuée), en utilisant la lois t, on peut trouver que cette probabilité est 0.01 donc 1%.

  Par convention cette probabilité est petite (inferieure a 0.05), et on peut conclure que en fait il y a une différence statistiquement signifiante entre les deux populations.

43

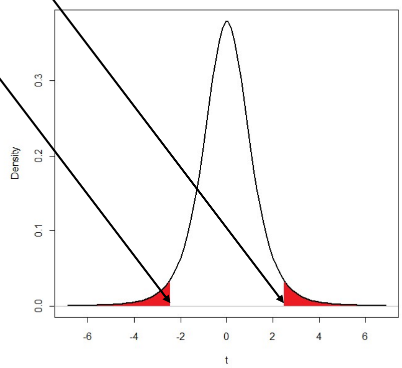
43

La loi de Student t_n

$Pr(|t_n>x|)=P$ $Pr(t_n<-x$ ou $t_n>x)=P$

Ou n – d.d.l., P une probabilité, t_n – valeur du v.a. Student

	P						
ν	0.50	0.40	0.30	0.20	0.10	0.05	0.01
1	1.000	1.376	1.963	3.078	6.314	12.706	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.921



44

Loi du χ^2 (KHi-2)

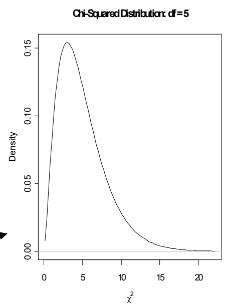
La Distribution χ^2 de Pearson
décrit le comportement statistique
d'une somme des carres des variables aléatoires normale distribuées, centrées, réduites, indépendantes:

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

(ou X_i sont de type N(0,1)).

Fonction densité de probabilité:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} I_{\{x\geq 0\}},$$



45

Loi du χ^2 (KHi-2)

- La loi χ^2 est définit sur [0, + ∞) et a d.d.l. = n degrés de liberté,
  - $E(\chi^2) = n$,
  - $Var(\chi^2) = 2\,n$,
- **Applications**:
  - Le test χ^2 d'independence (voir cours)
    - avec (L-1)*(C-1) degrés de liberté

$$\chi^2 = \sum_{i=1}^{LC} \frac{(f_i^o - f_i^t)^2}{f_i^t}$$

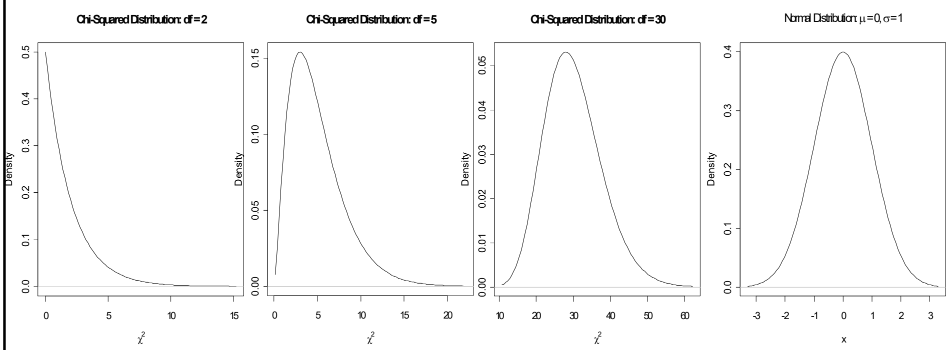
  - Le test McNemar (voir cours):
    - suivre la loi de Chi-deux avec 1 degré de liberté

$$\chi^2_{1ddl} = \frac{(|b-c|-0,5)^2}{b+c}$$

46

Loi du χ^2 (KHi-2)

- $\chi^2 \rightarrow N(n, \sqrt{2n})$, quand $n \rightarrow \infty$,
- **Si n > 30 la lois χ^2 et la lois Student sont proches**
- χ^2 (n=2) χ^2 (n=5) χ^2 (n=30) Normale

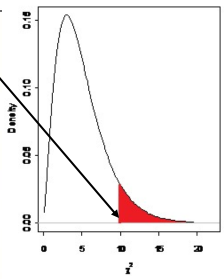


47

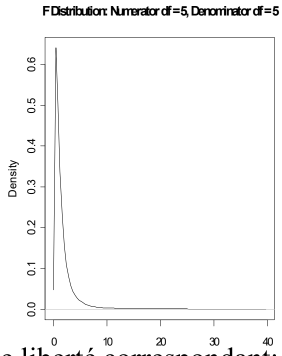
Loi du χ^2 (KHi-2)

Ce tableau donne la valeur pour une probabilité alpha tel que $X^2 > X^2_{n\alpha}$, ou n = d.d.l

	La probabilité α									
v	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0025	0,001
1	1,3233	1,6424	2,0722	2,7055	3,8415	5,0239	6,6349	7,8794	9,1404	10,8274
2	2,7726	3,2189	3,7942	4,6052	5,9915	7,3778	9,2104	10,5965	11,9827	13,8150
3	4,1083	4,6416	5,3170	6,2514	7,8147	9,3484	11,3449	12,8381	14,3202	16,2660
4	5,3853	5,9886	6,7449	7,7794	9,4877	11,1433	13,2767	14,8602	16,4238	18,4662
5	6,6257	7,2893	8,1152	9,2363	11,0705	12,8325	15,0863	16,7496	18,3854	20,5147
6	7,8408	8,5581	9,4461	10,6446	12,5916	14,4494	16,8119	18,5475	20,2491	22,4575
7	9,0371	9,8032	10,7479	12,0170	14,0671	16,0128	18,4753	20,2777	22,0402	24,3213
8	10,2189	11,0301	12,0271	13,3616	15,5073	17,5345	20,0902	21,9549	23,7742	26,1239
9	11,3887	12,2421	13,2880	14,6837	16,9190	19,0228	21,6660	23,5893	25,4625	27,8767
10	12,5489	13,4420	14,5339	15,9872	18,3070	20,4832	23,2093	25,1881	27,1119	29,5879
11	13,7007	14,6314	15,7671	17,2750	19,6752	21,9200	24,7250	26,7569	28,7291	31,2635
12	14,8454	15,8120	16,9893	18,5493	21,0261	23,3367	26,2170	28,2997	30,3182	32,9092
13	15,9839	16,9848	18,2020	19,8119	22,3620	24,7356	27,6882	29,8193	31,8830	34,5274
14	17,1169	18,1508	19,4062	21,0641	23,6848	26,1189	29,1412	31,3194	33,4262	36,1239
15	18,2451	19,3107	20,6030	22,3071	24,9958	27,4884	30,5780	32,8015	34,9494	37,6978
16	19,3689	20,4651	21,7931	23,5418	26,2962	28,8453	31,9999	34,2671	36,4555	39,2518
17	20,4887	21,6146	22,9770	24,7690	27,5871	30,1910	33,4087	35,7184	37,9462	40,7911
18	21,6049	22,7595	24,1555	25,9894	28,8693	31,5264	34,8052	37,1564	39,4220	42,3119
19	22,7178	23,9004	25,3289	27,2036	30,1435	32,8523	36,1908	38,5821	40,8847	43,8194
20	23,8277	25,0375	26,4976	28,4120	31,4104	34,1696	37,5663	39,9969	42,3358	45,3142



La loi de Fisher



- La loi F
  - est défini sur [0, + ∞)
  - décrit le comportement
    - d'une division entre deux variables aléatoires
    - qui suivent un lois χ^2
    - chaque d'eux divisée par le nombre des dégrées de liberté correspondant:

$F_{n1,\, n2} = \chi^2(n_1)/\chi^2(n_2)$

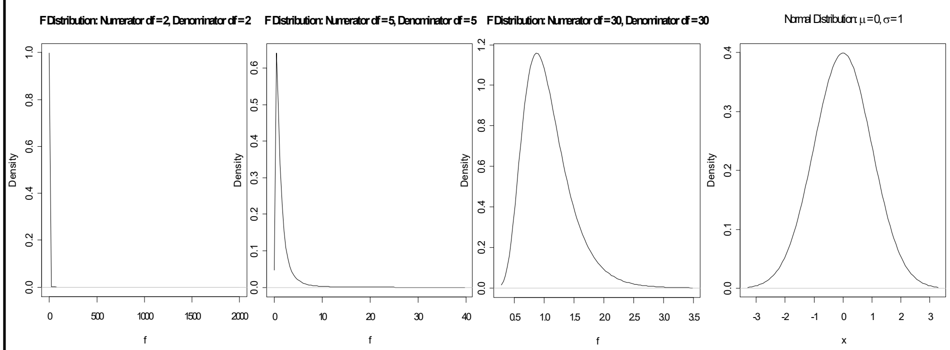
- La loi de Fisher a deux dégrées de liberté n_1 et n_2
- **Applications**:
  - Le test Fisher pour comparer deux variances
  - Le test ANOVA pour comparer des moyennes de >=3 échantillons (voir cours)

49

Loi du Fisher

Fisher →N(), quand n_1 et $n_2 \to \infty$,

- Fisher ($n_1,$ n_2=2) Fisher ($n_1,$ n_2=5) Fisher ($n_1,$ n_2=30) Normale



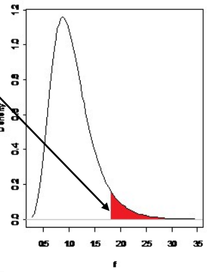
50

50

La Loi du Fisher

Ce tableau donne la valeur telle que pour une probabilité alpha de 0,05, $F>F_{n1n2\alpha}$ ou n_1, n_2=d.d.l

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.882	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255



51

Les lois de prob. et leurs applications

Lois	Tests statistiques
Binomiale	▪ Intervalle de confiance pour un proportion ▪ test binomial pour un proportion
Poisson	Régression Poisson – prédiction des événements dans le temps / espace
Normale	• intervalles de confiance pour les moyennes/fréquences • le test Z pour comparaison des moyennes/fréquences

Student (t)	▪ Intervalle de confiance
	  ▪ d'une moyenne pour un échantillon
	  ▪ d'une différence entre deux moyennes pour deux échantillons indépendants
	▪ Le test de Student pour la comparaison de moyennes
	  ▪ Test t pour échantillons appariés (eg. on compare le même échantillon avant et après un intervention
	  ▪ Test t pour des échantillons indépendants (avec variances égales ou inégales)
	  ▪ Le test Student pour le coefficient de corrélation
	  ▪ Le test Wald pour les coefficients de la régression linéaire

χ^2	Le test χ^2 d'independence Le test McNemar
Fisher	Le test Fisher pour comparer deux variances Le test ANOVA pour comparer des moyennes de >=3 échantillons

52

Les lois de prob. et leurs applications

On observe que au lieu d'utiliser la lois normale centrée réduite on utilise la lois t (student) parce qu'elle est plus pratique, elle nous aide aussi quand les échantillons sont petits, mais aussi quand les échantillons sont grands (et similaire a la lois normale).

On effet il n'y a pas une seule bonne variante d' utiliser les lois de probabilité. Ca dépend de préférence, d'habitude.

Mais dans notre cours on a fait ces chois pour pour avoir une uniformité pour l' évaluation pendant l'examen.

53

Exemples dans des articles

Observez l'utilisation du: moyenne et la déviation standard dans les résultats des articles scientifiques:

Characteristics	Treatment Group (N=257)	Control Group (N=257)
Age, mean (sd), years	56.7 (10.5)	57.9 (9.6)
Female, No. (%)	114 (44.4)	123 (47.9)
Smoking history, No. (%)		
Never smoked	129 (50.2)	144 (56.0)
Former	89 (34.6)	86 (33.5)
Current	39 (15.2)	27 (10.5)
Diabetes factors, mean (SD)		
HbA1c, %	7.8 (0.65)	7.8 (0.60)
Fasting glucose, mg/dL , median	150	147
IQR (quartile 1 - quartile 3)	(125 – 174)	(122 – 172)
Duration of diabetes, years	12.3 (8.2)	11.3 (8.4)
Anthropometrics, mean (SD)		
Weight, kg	99.5 (24.3)	97.5 (21.7)
BMI, kg/m^2	34.7 (7.5)	34.2 (6.7)
Blood pressure[c], mean(sd), mm Hg		
Systolic	133.1 (20.7)	135.1 (20.4)
Diastolic	78.8 (12.3)	78.8 (10.9)

Pour le poids (weight) – le fait qu'ils montre la moyenne et la déviation standard signifie que les données sont normale distribuées (on suppose q'ils ont vérifiée ca mais ils ne montre pas cette étape).

La moyenne de 99,5 nous **indique** ou est le centre des valeurs du poids dans le group avec traitement

La **déviation standard (SD – standard deviation)** de 24,3 nous indique que dans l' intervalle moyenne de 99,5 moins 24,3 (1DS) et moyenne plus un déviation standard, il y a approximative 68% des poids des sujets

Depuis: Engebretson SP, Hyman LG, Michalowicz BS, Schoenfeld ER, Gelato MC, Hou W, Seaquist ER, Reddy MS, Lewis CE, Oates TW, Tripathy D, Katancik JA, Orlander PR, Paquette DW, Hanson NQ, Tsai MY. **The effect of nonsurgical periodontal therapy on hemoglobin A1c levels in persons with type 2 diabetes and chronic periodontitis: a randomized clinical trial.** JAMA. 2013 Dec 18;310(23):2523-32.

54

Exemples des questions pour l'examen

X	0	1	2	3
Pr(x)	0.4	0.3	0.2	0.1

1) *La distribution de probabilité d'apparition d'un allergies a administration d'un traitement qui a été donné a 3 sujets est en haut. Quelle est l'Esperance mathématique de la variable aléatoire associée a le phénomène décrit antérieur?

a) 0,3

b) 0,4

c) 1

d) 1,4

e) 0,7

Réponse: a

Explication = 0*0.4 + 1*0.3 + 2*0.2 + 3*0.1

Réponse: c (**variable aléatoire discrète**)

2) Lesquels des suivantes applications de les lois des probabilités sont justes:

a) lois binomiale – test Ficher exact

b) lois student – test t pour des échantillons indépendants avec variances égales

c) lois t– test student pour des échantillons indépendants avec variances inégales

d) lois Fisher – test Fisher exact

e) lois Fisher – le test Fisher pour comparer deux variances

Réponse: b, c, e

55

Exemples des questions pour l'examen

3) Regardes le table suivant depuis une article scientifique médical. Il étudie des patients avec insuffisance rénale. Répondez aux questions sur le diapositive suivante

Variable	Global	eGFR ≥60	eGFR 59–45	eGFR 44–30	eGFR <30	p-value[a]
Female, n (%)	12,520 (44.2)	10,363 (42.8)	1,478(51.2)	568 (53.9)	111 (52.6)	<0.001
Hypertension, n (%)	22,971 (81.0)	19,001 (78.5)	2,737 (94.8)	1,026 (97.4)	207 (98.1)	<0.001
Any cardiovascular disease[b], n (%)	3,969 (14.0)	3,026 (12.5)	590 (20.4)	286 (27.2)	67 (31.8)	<0.001
Coronary Heart disease, n (%)	3,048 (10.8)	2,326 (9.6)	446 (15.5)	221 (21.0)	55 (26.1)	<0.001
Stroke, n (%)	1,148 (4.1)	854 (3.5)	186 (6.4)	90 (8.5)	18 (8.5)	<0.001
Insulin treatment, n (%)	4,737 (16.7)	3,781 (15,6)	567 (19.6)	302 (28.7)	87 (41.2)	<0.001
Age T2DM (years), mean (SD)	58.8(10.5)	57.5(10.2)	65.6 (9.0)	67.2 (9.7)	65.0(10.1)	<0.001
Age at retinography (years), mean (SD)	65.7(10.7)	64.3(10.4)	73.7(7.9)	76.1 (7.8)	74.6(8.9)	<0.001
Diabetes duration (years), mean (SD)	7.0 (5.2)	6.8 (5.0)	8.2 (5.6)	9.0 (6.0)	9.7(6.1)	<0.001
A1C (%), mean (SD)	7.4 (1.4)	7.4(1.4)	7.2 (1.2)	7.3 (1.4)	7.2 (1.3)	<0.001
Non-HDL cholesterol (mg/dL), mean (SD)	137.3(35.2)	137.8(35.2)	135.4(34.6)	132.5(35.2)	132.5(38.2)	<0.001
Hemoglobin (g/dL), mean (SD)	14.0 (1.5)	14.1 (1.4)	13.3 (1.6)	12.7 (1.6)	12.0 (1.5)	<0.001
SBP (mmHg), mean (SD)	134.9(12.6)	134.8(12.5)	136.1(13.1)	135.8(13.5)	133.9(14.0)	<0.001
DBP (mmHg), mean (SD)	76.4 (8.3)	76.9 (8.2)	74.2 (8.2)	71.9 (8.3)	70.4 (8.3)	<0.001
Heart Rate (bpm), mean (SD)	76.1 (12.1)	76.3 (12.0)	75 (12.5)	74.2 (12.5)	72.5 (12.0)	<0.001

[a]P value for comparison of groups by glomerular filtration rate with Pearson's chi-square test for qualitative variables and t-test for quantitative ones.

[b]Any cardiovascular disease includes coronary heart disease and/or stroke

A1C, glycated hemoglobin; bpm, beats per minute; DBP, Diastolic Blood Pressure; DR: diabetic retinopathy.

doi:10.1371/journal.pone.0149448.t001

Rodríguez-Poncelas A, Mundet-Tudurí X, Miravet-Jiménez S, Casellas A, Barrot-De la Puente JF, Franch-Nadal J, Coll-de Tuero G. Chronic Kidney Disease and Diabetic Retinopathy in Patients with Type 2 Diabetes. PLoS One. 2016 Feb 17;11(2):e0149448.

https://doi.org/10.1371/journal.pone.0149448

56

Exemples des questions pour l'examen

3) Lesquelles des réponses suivantes sont correctes :

a) Pour la caractéristique des sujets - Haemoglobin (g/dL) (hémoglobine), on peut faire des mesures (essais) sur un nombre des individues, et puis leurs résultats peuvent être associée a des valeurs numériques, et ca représente une variable aléatoire continue

-> parce que les valeurs de la caractéristique (hémoglobine en g/dL) sont des nombres, elle est quantitative; parce que on peut mesurer les valeurs avec une précision qui nous donne des valeurs avec décimales, (même si on n'écrive pas les décimales, seulement s'ils sont possibles, en théorie), la caractéristique est continue; le fait qu'on associe a des résultats des essais des valeurs numériques – ca représente une variable aléatoire. au final est une variable aléatoire continue. => vraie

Une indice pour le fait que la caractéristique est quantitative, est le fait qu'ils on calcule la moyenne - Haemoglobin (g/dL), mean

b) Pour la caractéristique des sujets - Heart rate (bpm) (fréquence cardiaque – en battements par minute, une seule mesure par sujet)), on peut faire des mesures (essais) sur un nombre des individues, et puis leurs résultats (on compte combiens des battements par minute) peuvent être associée a des valeurs numériques, et ca représente une variable aléatoire discrète

-> parce que les valeurs de la caractéristique (fréquence cardiaque – en battements par minute) sont des nombres, elle est quantitative; C'est difficile de dire s'il est continue ou discrète. Ca dépend des chercheurs. Un étude peut mesurer un seul fois la fréquence cardiaque – dans ce cas, le nombre des battements sera quantitative discrète. Si on les chercheurs ont mesure trois fois la fréquence cardiaque a 5 minute intervalle, et ils ont fait la moyenne, la variable sera quantitative continue, parce que on peut avoir des décimales. Ici c'est précisé - une seule mesure par sujet, donc la caractéristique est quantitative discrète => vraie

Une indice pour le fait que la variable est quantitative, est le fait qu'ils on calcule la moyenne - Heart rate (bpm), mean

c) Pour la caractéristique des sujets – Hypertension n (%), on peut faire des mesures (essais) sur un nombre des individues, et puis leurs résultats (on compte combiens sujets ont l'hypertension (« succès » de l'essai), parmi le total des essais – le nombre total des sujets) peuvent être associée a des valeurs numériques, et ca représente une variable aléatoire discrète – et on peut étudier les probabilités associées a les résultats des essais avec la lois Binomiale

-> parce que les valeurs de la caractéristique (hypertension) sont du texte (avec/sans hypertension), elle est qualitative dichotomique (parce que il y a seulement deux valeurs possibles); on peut associer a ce gens des valeurs le concept de succès de l'essai, et l' échec; la variable aléatoire qui associée a un nombre des « succès » dans un nombre des essaies, une valeur numérique, est une variable aléatoire discrète parce que le nombre des succès sont seulement des valeurs entières; on peut étudier les probabilités associées a les résultats des essais de ce gens avec la lois Binomiale => vraie

Réponse: a, b, c

57

Les difficultés pour l'évaluation de la normalité des données

Pour savoir quelle est la lois de probabilité la plus proche a une série des données collectées sur un échantillon, on peut évaluer la distribution a l'aide des graphiques (histogramme, statistiques: coefficient d' asymétrie, coefficient d' aplatissement) – voir le cours 5.

Quand les **échantillons** sont avec **beaucoup** des **sujets**, on est **plus sur** que la **distribution** des probabilités **observe** sur **l' échantillon**, est **similaire** a la **distribution** des probabilités **de la population** depuis laquelle on a extrait l'échantillon.

Quand les **échantillons** sont avec **un nombre réduit** des **sujets**, on est **moins sur** que la **distribution** des probabilités **observe** sur **l' échantillon**, est **similaire** a la **distribution** des probabilités **de la population** depuis laquelle on a extrait l'échantillon.

Voir des **simulations** des extractions des échantillons depuis la même distribution normale d'une population de 1 000 000 sujets. Plus les échantillons sont petits ils ne semble pas normale distribuées (même si on les a extrait depuis une population avec distribution normale). Donc ces simulations sont une suggestion de la **difficulté de deviner la vraie distribution des données dans la population, en regardant une échantillon**

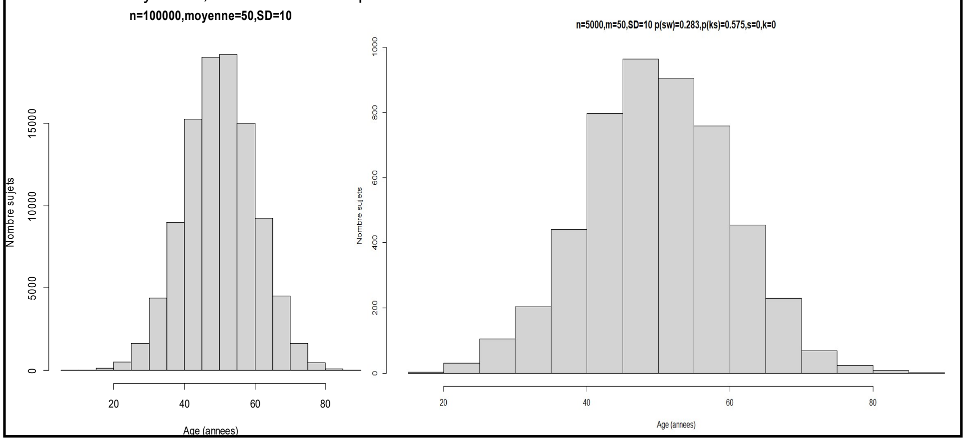
58

On observe que les échantillons de grande taille (avec beaucoup des observations) extraites depuis une population normale sont très proche a la distribution normale.

Histogrammes des échantillons extraits depuis une population normale distribuée (taille de **1 000 000** sujets) de l'âge avec une Moyenne de 50, déviation standard de 10.

Ici les échantillons ont la taille de **5000 sujets**.

Abréviations: n – nombre sujets de l'échantillon, m = Moyenne de la population (le logiciel n'accepte pas μ, la notation classique), SD = déviation standard de la population, p(sw) = valeur du p du test de normalité Shapiro-Wilk, p(ks) = valeur du p du test de normalité Kolmogorov Smirnov, s = coefficient d'asymétrie, k = coefficient d'aplatissement



59

On observe que plus la taille de l'échantillons extraites depuis une population normale, diminue, elles peuvent dévier parfois de la normalité.

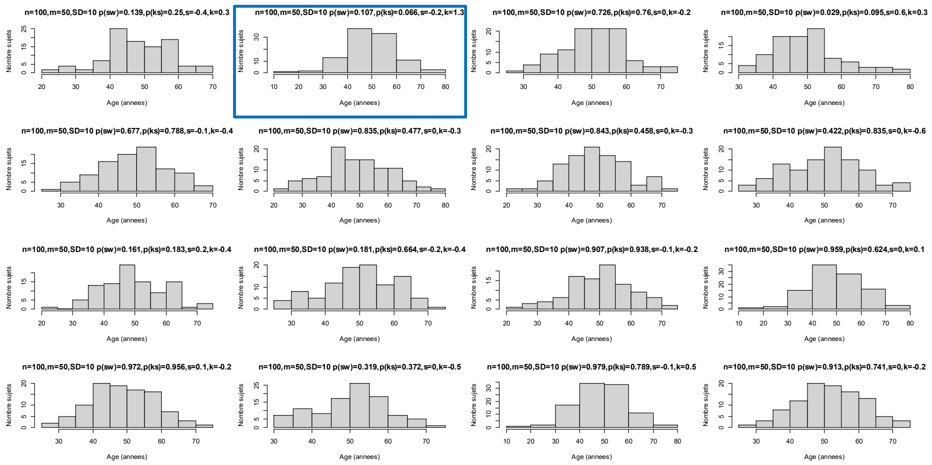
Histogrammes des échantillons extraits depuis une population normale distribuée de l'âge avec une Moyenne de 50, déviation standard de 10.

Ici les échantillons ont la taille de **100 sujets**.

Abréviations: n – nombre sujets de l'échantillon, m = Moyenne de la population, SD = déviation standard de la population, p(sw) = valeur du p du test de normalité Shapiro-Wilk, p(ks) = valeur du p du test de normalité Kolmogorov Smirnov, s = coefficient d'asymétrie, k = coefficient d'aplatissement

Avec bleu – le coefficient d' aplatissement ou asymétrie sont en dehors de (-1; 1)

La forme de l'histogramme peut sembler parfois que les données ne sont pas normale distribuées (même si la population est normale …)



60

On observe que plus la taille de l'échantillons extraites depuis une population normale, diminue, elles peuvent dévier parfois de la normalité.

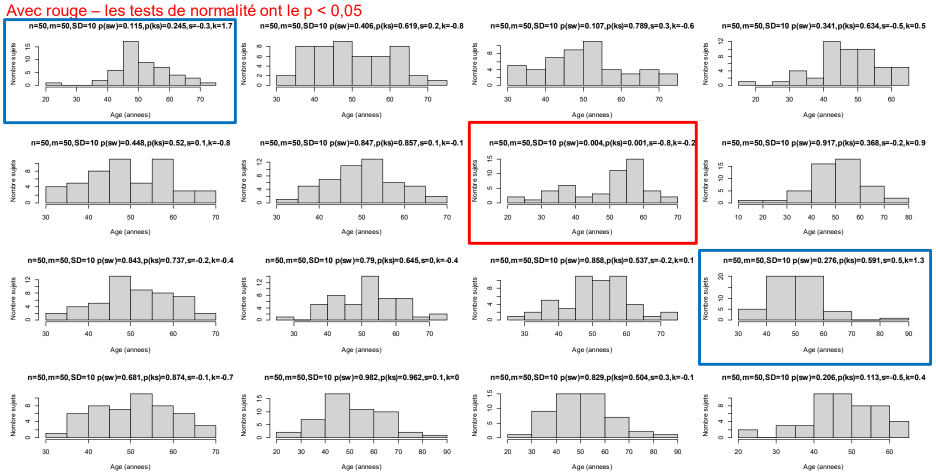
Histogrammes des échantillons extraits depuis une population normale distribuée de l'âge avec une Moyenne de 50, déviation standard de 10.

Ici les échantillons ont la taille de **50 sujets**.

Abréviations: n – nombre sujets de l'échantillon, m = Moyenne de la population, SD = déviation standard de la population, p(sw) = valeur du p du test de normalité Shapiro-Wilk, p(ks) = valeur du p du test de normalité Kolmogorov Smirnov, s = coefficient d'asymétrie, k = coefficient d'aplatissement

Avec bleu – le coefficient d' aplatissement ou asymétrie sont en dehors de (-1; 1)

Avec rouge – les tests de normalité ont le p < 0,05



61

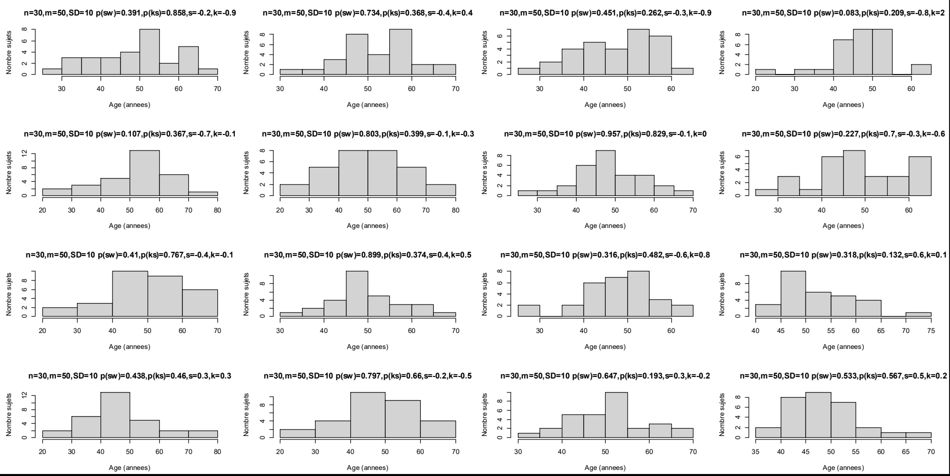
On observe que plus la taille de l'échantillons extraites depuis une population normale, diminue, elles peuvent dévier parfois de la normalité.

Histogrammes des échantillons extraits depuis une population normale distribuée de l'âge avec une Moyenne de 50, déviation standard de 10.

Ici les échantillons ont la taille de **30 sujets**.

Abréviations: n – nombre sujets de l'échantillon, m = Moyenne de la population, SD = déviation standard de la population, p(sw) = valeur du p du test de normalité Shapiro-Wilk, p(ks) = valeur du p du test de normalité Kolmogorov Smirnov, s = coefficient d'asymétrie, k = coefficient d'aplatissement

La forme de l'histogramme peut sembler parfois que les données ne sont pas normale distribuées (même si la population est normale …)



62

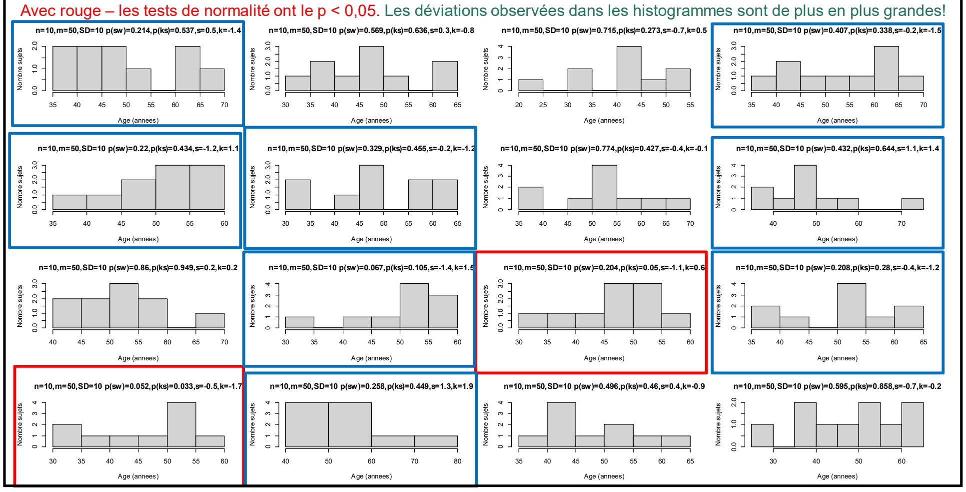
On observe que plus la taille de l'échantillons extraites depuis une population normale, diminue, elles peuvent dévier parfois de la normalité.

Histogrammes des échantillons extraits depuis une population normale distribuée de l'âge avec une Moyenne de 50, déviation standard de 10.

Ici les échantillons ont la taille de **10 sujets**. Abréviations: n – nombre sujets de l'échantillon, m = Moyenne de la population, SD = déviation standard de la population, p(sw) = valeur du p du test de normalité Shapiro-Wilk, p(ks) = valeur du p du test de normalité Kolmogorov Smirnov, s = coefficient d'asymétrie, k = coefficient d'aplatissement

Avec bleu – le coefficient d' aplatissement ou asymétrie sont en dehors de (-1; 1)

Avec rouge – les tests de normalité ont le p < 0,05. Les déviations observées dans les histogrammes sont de plus en plus grandes!



63

FIN….

Graphique réalisé par Daniel Leucuța avec:
R software envinronment for statistical computing and graphics
Ou avec Teaching demos – package
http://www.r-project.org/

64