

Utilisation de statistiques descriptives

Pierre-Yves Henry
henry@mnhn.fr



Quelle analyse ?

Y \ X	Var. catégorielle	Var. continue
	Var. catégorielle	Var. continue
Var. catégorielle	Analyse de fréquence	Faire des catégories Transformer en variable quantitative
Var. continue	Comparaison de moyenne ANOVA	Corrélation Régression



Pourquoi des statistiques descriptives ?

**Représentation synthétique des données utile pour
"voir" si:**

- **il semble y avoir l'effet recherché**
- **les hypothèses d'application de l'analyse sont respectées**
- **décrire de manière synthétique les données**

Distribution des données

Ex: distribution de longueur d'aile pour 1A et +1A d'ERIRUB

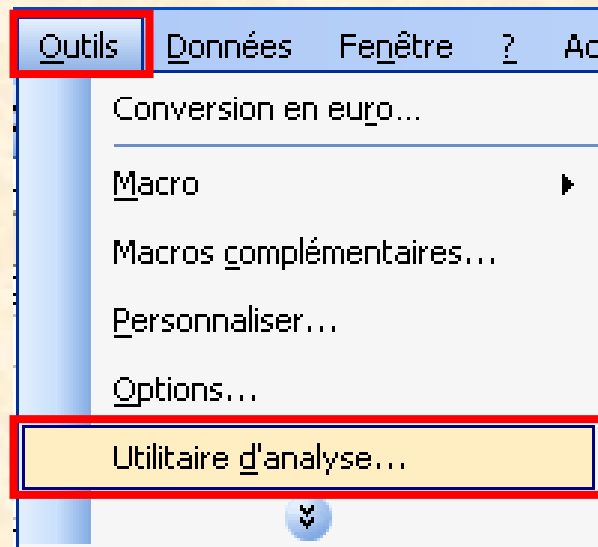
Etape 1: visualisation graphique de la distribution d'une variable

Utiliser le fichier "2 2a Statistiques descriptives et exploration graphique.xls"

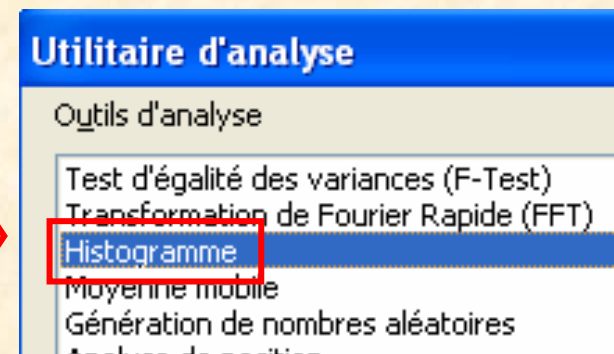
Coller les données dans la page "Données"

	A	B	C
1	Groupe	1	2
2	Nom de groupe	+1A	1A
3		74.0	73.0
4		73.0	71.0
5		70.5	72.5
6		71.0	70.5
7		72.5	70.5
8		71.0	71.0
9		70.5	72.5
10		73.5	71.5
11		72.0	72.5
12		72.5	68.5
13		75.5	74.0
14		73.0	70.0
15		70.0	70.5
16		68.5	71.0
17		72.5	72.0
18		73.5	73.0
19		71.0	68.5
20		74.0	75.0
21		73.0	72.5

Ouvrir l'Utilitaire d'analyse



Choisir
"Histogramme"



Distribution des données

Etape 1: visualisation graphique de la distribution d'une variable

Histogramme

Paramètres d'entrée

Plage d'entrée:

Plage des classes:

☐ Intitulé présent

Options de sortie

☐ Plage de sortie:

☒ Insérer une nouvelle feuille:

☐ Créer un nouveau classeur

☐ Pareto (Histogramme ordonné)

☐ Pourcentage cumulé

☒ Représentation graphique

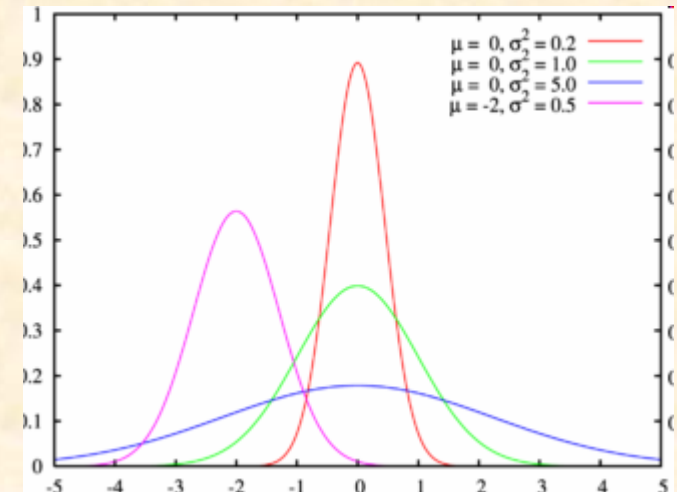
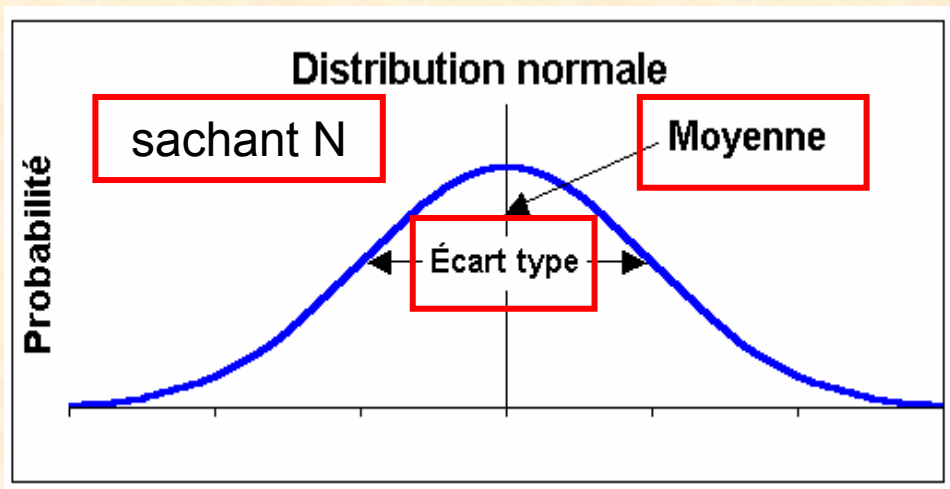
OK
Annuler
Aide

Définition des classes à utiliser (en indiquant la borne supérieure)

Distribution des données

Etape 2: interprétation de l'histogramme

Beaucoup d'analyses ont besoin de faire l'hypothèse que la distribution est Normale
car les test n'utilisent pas les données
mais trois statistiques résumant les données
moyenne + écart-type + N

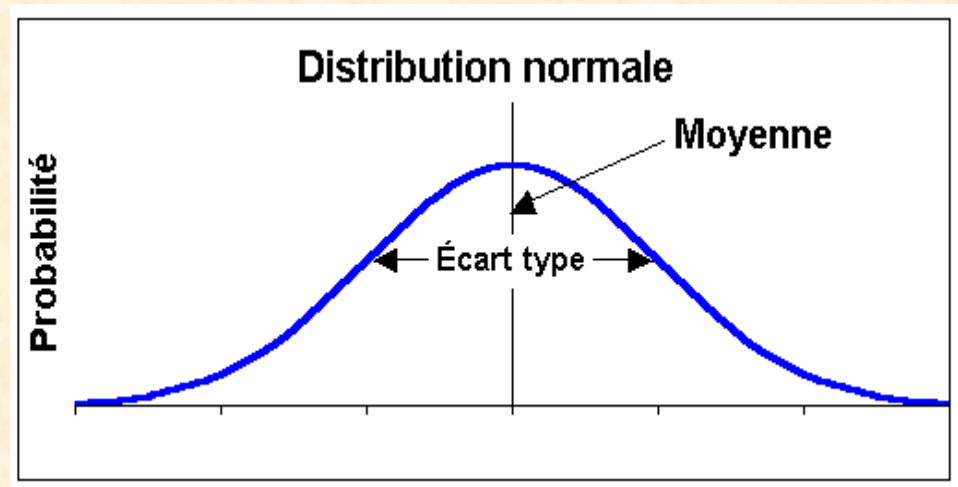
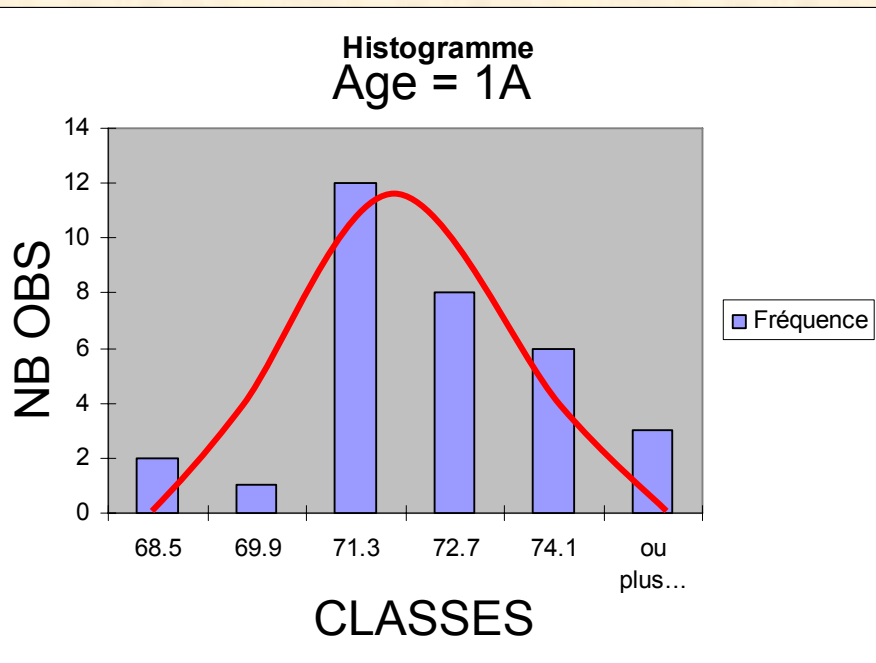


La question: la distribution est-elle à peu près normale ?

Distribution des données

Etape 2: interprétation de l'histogramme

La question: la distribution est-elle à peu près normale ?

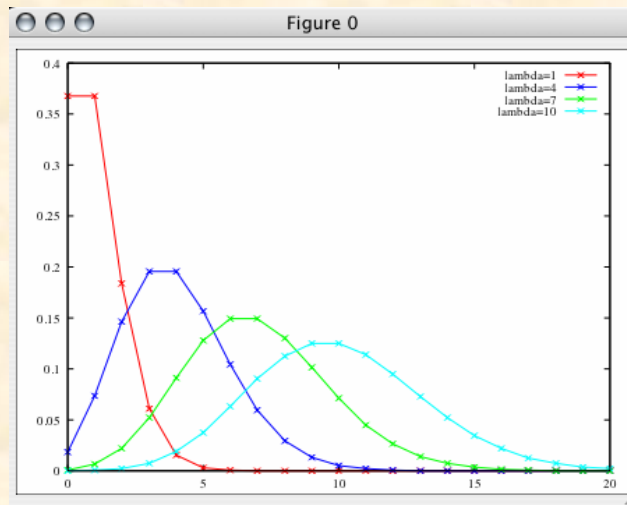


Distribution des données

Etape 2ter: **Et si la distribution des données n'est pas normale?**
Une solution: la transformation de la variable

Comptages

(nb d'individus, nb de contrôles, etc)



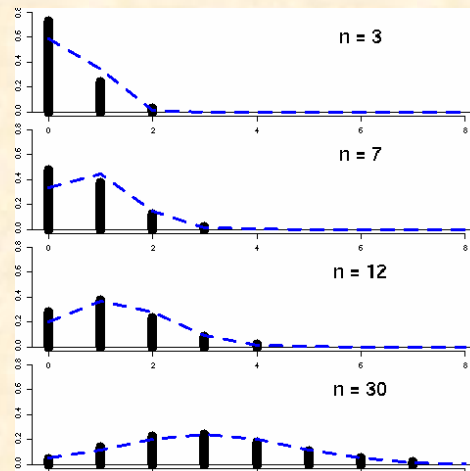
Transformation Log

$$\text{Si } Y > 0 \quad Y' = \text{LN}(Y)$$

$$\text{Si } Y \geq 0 \quad Y' = \text{LN}(Y+1)$$

Pourcentages

(taux de contrôles, âge-ratio)



Transformation Logit

$$\text{Si } Y > 0 \quad Y' = \text{LN}(Y/(1-Y))$$

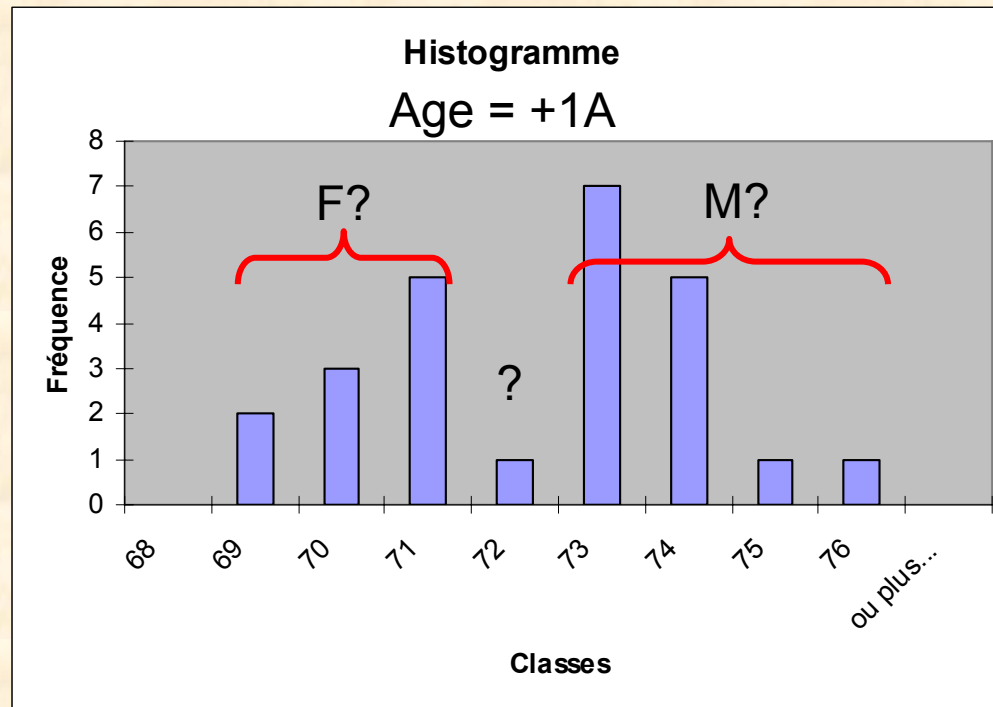
$$\text{Si } Y \geq 0 \quad Y' = \text{LN}(Y+0.001)$$

Distribution des données

Etape 2ter: **Cas plus compliqués de non-normalité**

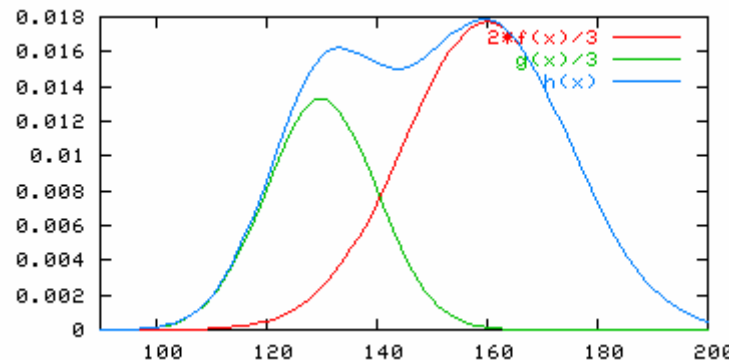
Mélange de groupes différents

(mélange de mâles et femelles)



Pas de solution simple

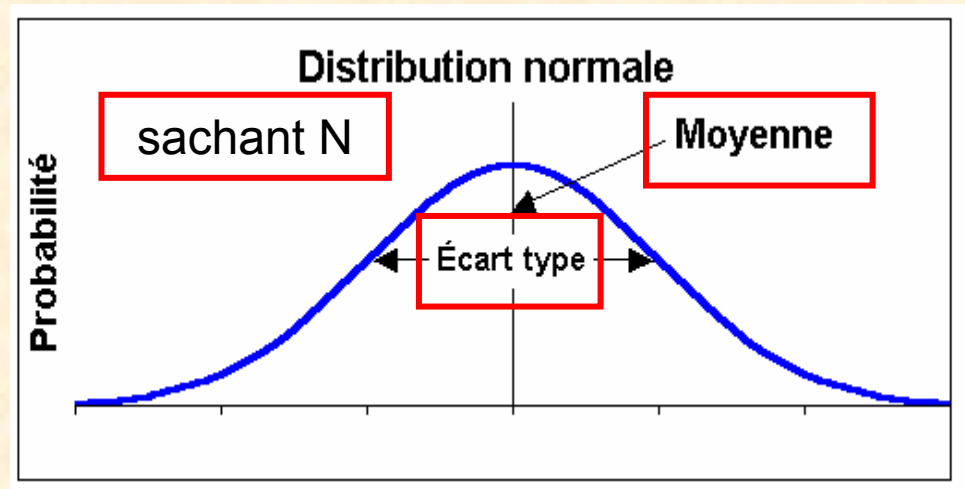
Identifier *a priori* les catégories à prendre en compte dans l'analyse



Distribution des données

Etape 3: **Si la distribution est normale**, on peut la résumer par

Groupe	1	2
Nom	+1A	1A
N	25	32
Moyenne	72.1	71.8
Ecart-type	1.8	1.8
Erreur-standard	0.4	0.3
1.96*ES	0.7	0.6
Borne inf IC 95%	71.4	71.1
Borne sup IC 95%	72.8	72.4



Une difficulté: l'écart-type augmente avec le nombre de mesures
-> une estimation de la variabilité réelle: l'**erreur-standard** = écart-type / \sqrt{N}

L'**intervalle de confiance au seuil de significativité 5%**: identifie la gamme de valeurs contenant 95% des mesures dans la population sous l'hypothèse de normalité

