

La statistique dans mon projet de recherche ?



Christine Detrembleur

UCL – pôle de recherche NMSK

INTRODUCTION

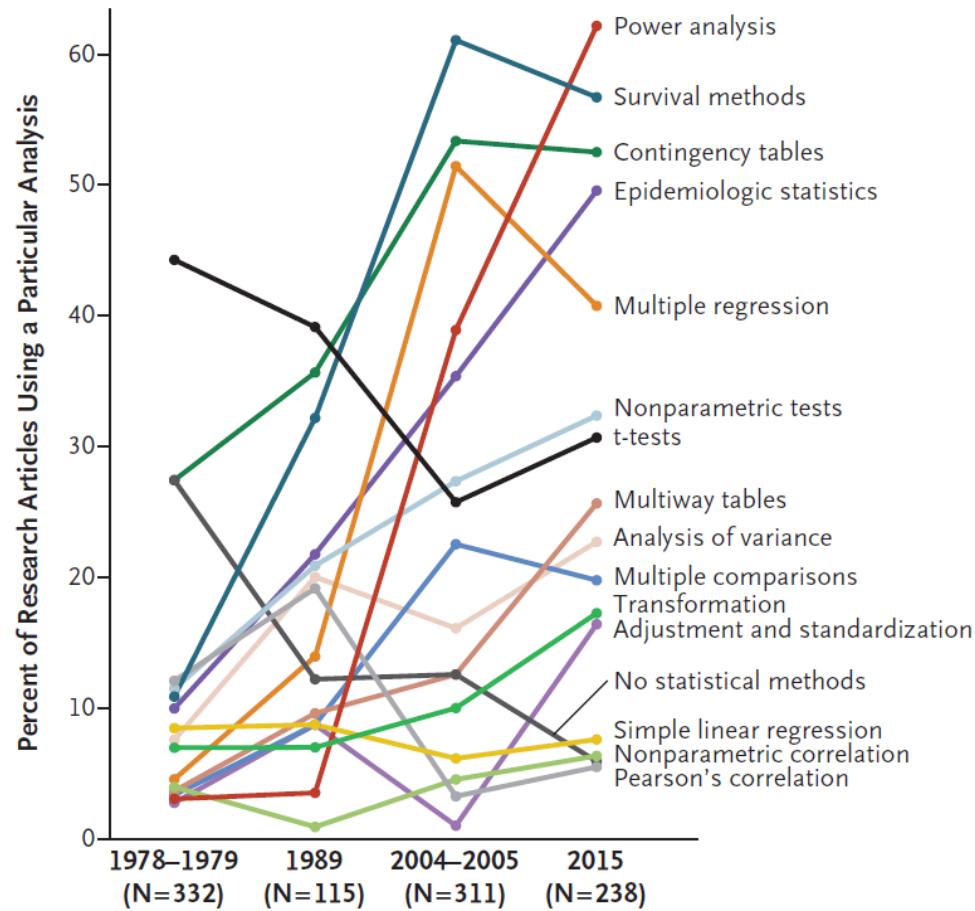


Figure 1. Percentage of Studies Using Particular Types of Statistical Analysis during Four Periods between 1978 and 2015.

1. Type des données

1.1. Variable qualitative

1.2. Variable quantitative

1. Type des données

1.1. Variable qualitative

1.2. Variable quantitative

1. Type des données

1.1. Variable qualitative

Variable dont les modalités sont des **mots**.

- Variable qualitative **nominale** (ne peut être ordonnée)
- Variable qualitative **ordinale** (peut suivre un ordre)

1. Type des données

1.1. Variable qualitative nominale vs ordinale

Variables	Modalités	Types
Etat civil	Célibataire, veuf, marié, conjoint de fait, divorcé	Qualitative Nominale
Degré de satisfaction face à son traitement	Très satisfait, satisfait, peu satisfait, insatisfait	Qualitative Ordinale

1. Type des données

1.1. Variable qualitative

1.2. Variable quantitative

1. Type des données

1.2. Variable quantitative

Fait référence à des valeurs, des **nombre**s

- Une variable quantitative **discrète**: nombre précis, score d'échelle
- Une variable quantitative **continue**: si elle peut prendre toutes les valeurs comprises entre 2 nombres.

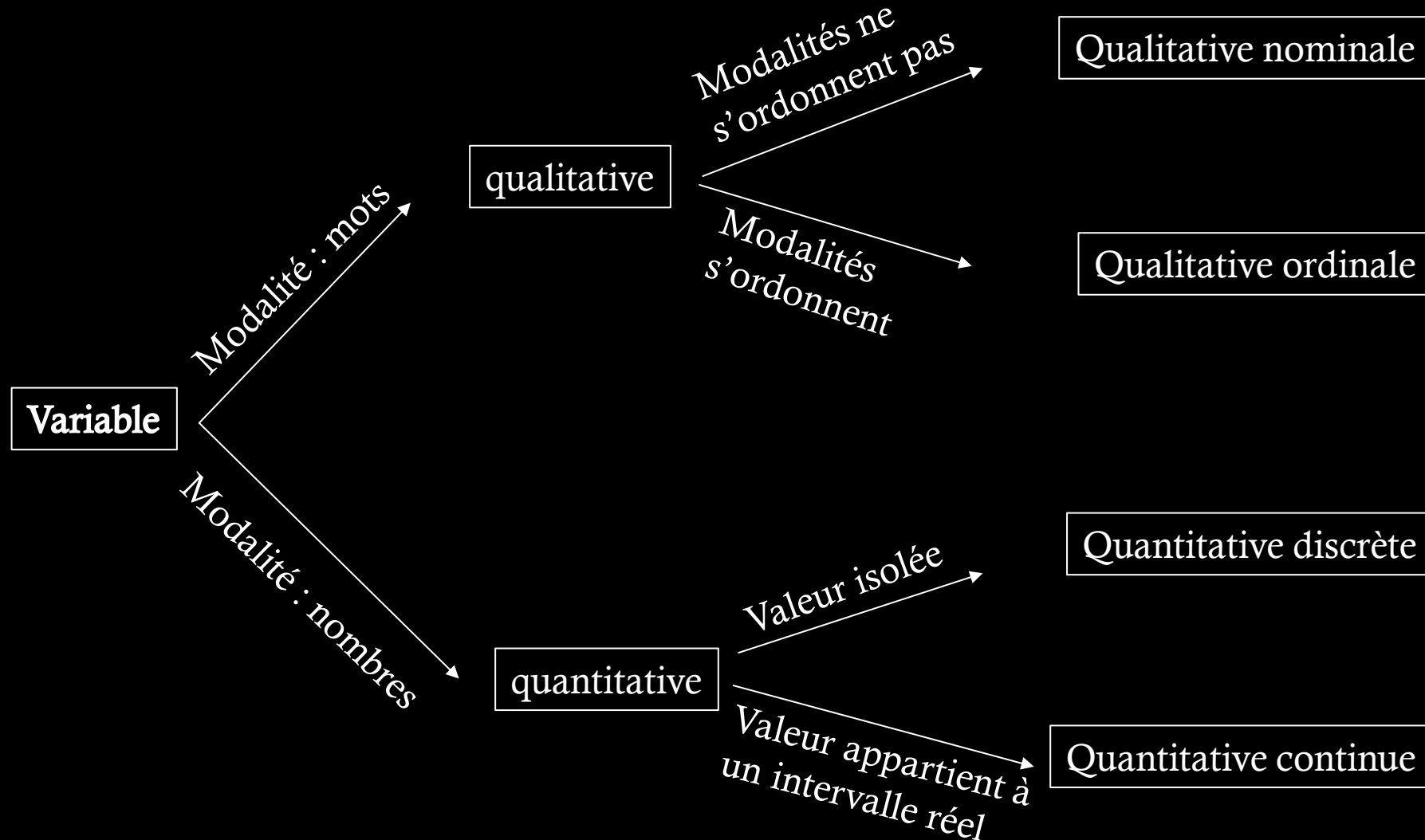
1. Type des données

1.2. Variable quantitative discrète vs continue

Variables	Valeurs	Types
Score échelle (KOOS, AFOS, DASH...)	1, 2 , 3 , 4 ou 10, 20, 30	Quantitative discrète
Poids	25.6 kg, 33.15 kg...	Quantitative continue

1. Type des données

1.3. Résumé



2. Statistique

2.1 statistique descriptive

2.2 Inférence statistique

2. Statistique

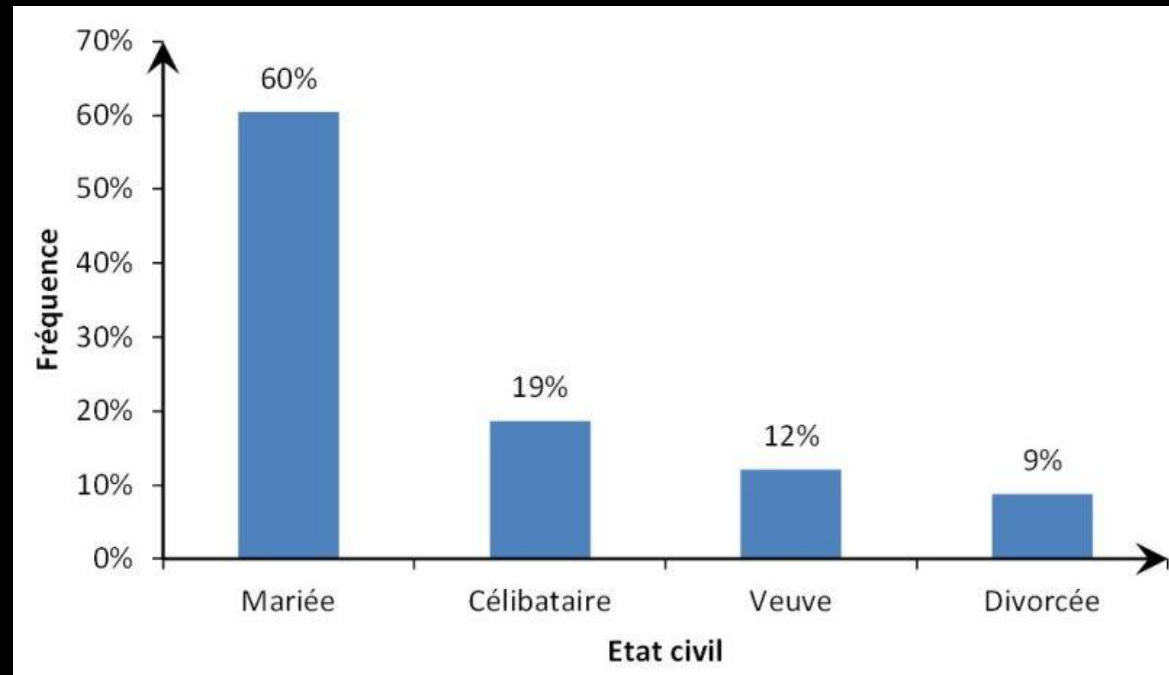
2.1 statistique descriptive

2.2 Inférence statistique

2. Statistique

2.1 statistique descriptive : variables qualitatives

Table de fréquence ex : Etat Civil



2. Statistique

2.1 statistique descriptive : Variable quantitative discrète

Médiane

Range (1^{er} et 3^e quartiles)

Percentiles

La médiane est la valeur qui sépare une série de données ordonnées par ordre croissant en deux groupes de même effectif.

Ex nombre de frères et sœurs chez 5 sujets

0,1,**1**,2,2

Médiane = 1

2. Statistique

2.1 statistique descriptive: Variable quantitative continue

Moyenne

Ecart Type

Erreur Standard

Intervalle de confiance

La moyenne = somme des effectifs / nombre d'effectifs.

Ex nombre de frères et sœurs chez 5 sujets

0,1,1,2,2

moyenne=[0+1+1+2+2]/5=1.2

2. Statistique

2.1 statistique descriptive

2.2 Inférence statistique

2. Statistique

2.2. Inférence statistique : variables qualitatives

Tester la liaison entre 2 variables pour voir s'il existe une relation

- 2 groupes d'individus distincts : table de contingence chi-carré

2. Statistique

2.2. Inférence statistique : variables qualitatives

Tester la liaison entre 2 variables pour voir s'il existe une relation

- 2 groupes d'individus distincts : table de contingence chi-carré
- 3 groupes ou plus d'individus distincts : table de contingence chi-carré

2. Statistique

2.2. Inférence statistique : variables qualitatives

Tester la liaison entre 2 variables pour voir s'il existe une relation

- 2 groupes d'individus distincts: table de contingence chi-carré
- 3 groupes ou plus d'individus distincts: table de contingence chi-carré
- 2 groupes pairés : Avant et après traitement : Table de McNemar

2. Statistique

2.2. Inférence statistique : variables qualitatives

Tester la liaison entre 2 variables pour voir s'il existe une relation

- 2 groupes d'individus distincts: table de contingence chi-carré
- 3 groupes ou plus d'individus distincts: table de contingence chi-carré
- 2 groupes pairés : Avant et après traitement : Table de McNemar

Si nombre d'observations $< 5 \Rightarrow$ utiliser Fisher test au lieu chi-carré

2. Statistique

2.2. Inférence statistique : variables qualitatives

Vérifier d'abord la sensibilité et la spécificité de vos variables

	Cancer bronchique	Non Cancer bronchique
Fumeur	1350 (A vrai positif)	1296 (B faux négatif)
Non Fumeur	7 (C faux positif)	61 (D vrai négatif)

Sensibilité = proportion de malade parmi les patients à dépister : $A / (A+C)$ $1350 / (1350+7) = 0.99$

Spécificité = proportion de vrais négatifs parmi les non malades : $D / (B+D)$ $61 / (1296 + 61) = 0.044$

2. Statistique

2.2. Inférence statistique : variables qualitatives

Vérifier d'abord la sensibilité et la spécificité de vos variables

Sensibilité = 0.99 ou 99%

Spécificité = 0.044 ou 4.5%

- Facile d'avoir une bonne sensibilité: déclarer tout le monde positif... mais peu spécifique
- Un bon test doit être sensible et spécifique

Ici sensible mais non spécifique

2. Statistique

2.2. Inférence statistique : variables qualitatives

Exemple : table de contingence : 2 * 2

	Cancer bronchique	Non Cancer bronchique
Fumeur	1350	1296
Non Fumeur	7	61

Chi-square= 19.8 with 2 degrees of freedom. (P < 0.001)

2. Statistique

2.2. Inférence statistique : variables qualitatives

Odd ratio : Rapport éventuel entre facteurs de risque et morbidité ou mortalité

	Cancer bronchique	Non Cancer bronchique	
Fumeur	1350	1296	2646
Non Fumeur	7	61	68
	1357	1357	2714

Probabilité d'être atteint du cancer bronchique chez les fumeurs $1350 / 2646$ (ratio : 0.51)

Probabilité d'être atteint du cancer bronchique chez les non-fumeurs $7/68$ (ratio : 0.1)

2. Statistique

2.2. Inférence statistique : variables qualitatives

Odd ratio : exemple

	Cancer bronchique	Non Cancer bronchique	
Fumeur	1350	1296	2646
Non Fumeur	7	61	68
	1357	1357	2714

1350 / 2646 (ratio : 0.51)

7/68 (ratio : 0.1)

$$P_1 = [0,51 / (1-0,51)] = 1,04$$

$$P_0 = [0,10 / (1-0,1)] = 0,11$$

$$\text{Odd Ratio} = P_1 / P_0 = 1,04 / 0,11 = 9,45$$

2. Statistique

2.2. Inférence statistique : variables qualitatives

Odd ratio : exemple

Interprétation :

Odds ratio ≥ 1 , ce qui signifie que le facteur de risque (tabac) augmente la proportion de cancers bronchiques.

OR ≤ 1 , le «facteur de risque» étudié représenterait plutôt un facteur préventif (un traitement) efficace qui diminuerait la proportion de malades.

2. Statistique

2.2. Inférence statistique : variables qualitatives

Régression logistique multiple

Ex : Probabilité d'avoir une infection après clou centro - médullaire? Facteurs prédictifs?

Logit P infection = $-8.393 + (1.501 * \text{Diabète}) + (4.048 * \text{Fracture ouverte}) + (0.276 * \text{Fixateur}) + (3.057 * \text{Antibiotique}) - (0.771 * \text{Schéma AB}) + (1.934 * \text{Foyer O F}) + (3.491 * \text{nombre complications})$

2. Statistique

2.2. Inférence statistique : variables quantitatives

Préambule : Vérifier si distribution normale et égalité des variances respectées

Test de Shapiro $n \leq 5000$

Test Kolmogorov $n > 5000$

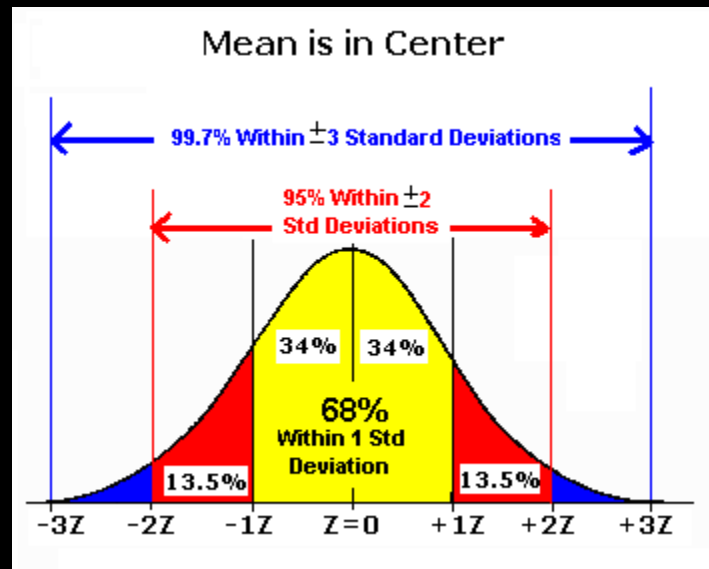
Test F ou Test de Bartlett

2. Statistique

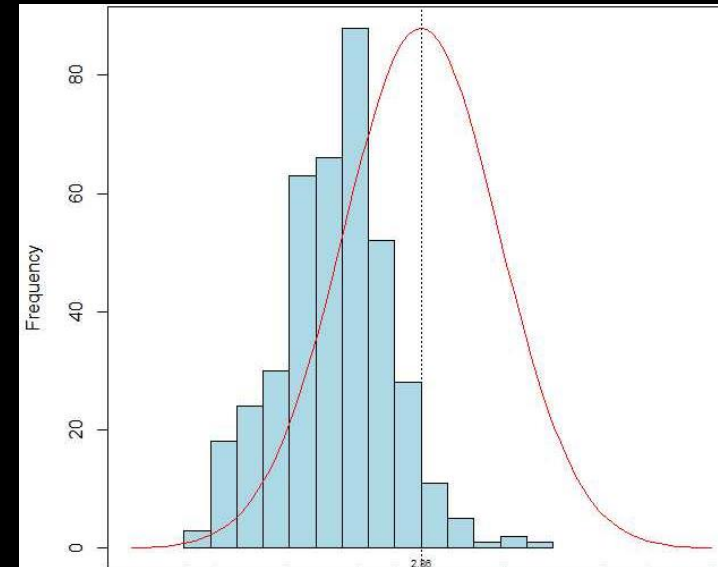
2.2. Inférence statistique : variables quantitatives

Préambule : Vérifier si distribution normale ??

Distribution normale: Stat paramétrique



Non normale: Stat NON paramétrique



2. Statistique

2.2. Inférence statistique : variables quantitatives discrètes

Ou quantitatives CONTINUES si normalité et/ou égalité des variances NON respectées

Test de comparaison d'échantillons

- 2 groupes d'individus distincts: test des rangs de Mann-Whitney

2. Statistique

2.2. Inférence statistique : variables quantitatives discrètes

Ou quantitatives CONTINUES si normalité et/ou égalité des variances NON respectées

Test de comparaison d'échantillons

- 2 groupes d'individus distincts : test des rangs de Mann-Whitney
- 3 groupes ou plus d'individus distincts : Analyse de variance sur les rangs de Kruskal-Wallis

2. Statistique

2.2. Inférence statistique : variables quantitatives discrètes

Ou quantitatives CONTINUES si normalité et/ou égalité des variances NON respectées

Test de comparaison d'échantillons

- 2 groupes d'individus distincts : test des rangs de Mann-Whitney
- 3 groupes ou plus d'individus distincts : Analyse de variance sur les rangs de Kruskal-Wallis
- 2 groupes pairés : Avant et après traitement : test sur les rangs de Wilcoxon

2. Statistique

2.2. Inférence statistique : variables quantitatives discrètes

Ou quantitatives CONTINUES si normalité et/ou égalité des variances NON respectées

Test de comparaison d'échantillons

- 2 groupes d'individus distincts : test des rangs de Mann-Whitney
- 3 groupes ou plus d'individus distincts : Analyse de variance sur les rangs de Kruskal-Wallis
- 2 groupes pairés : Avant et après traitement : test sur les rangs de Wilcoxon
- 3 groupes pairés : Anova à mesures répétées sur les rangs à un facteurs de Friedman

2. Statistique

2.2. Inférence statistique : variables quantitatives discrètes

Ou quantitatives CONTINUES si normalité et/ou égalité des variances NON respectées

Trouver une liaison entre les variables

- 2 groupes d'individus distincts : test des rangs de Mann-Whitney
- 3 groupes ou plus d'individus distincts : Analyse de variance sur les rangs de Kruskal-Wallis
- 2 groupes pairés : Avant et après traitement : test sur les rangs de Wilcoxon
- 3 groupes pairés : Anova à mesures répétées sur les rangs à un facteurs de Friedman
- Trouver une association entre deux variables : coefficient de corrélation de Spearman

2. Statistique

2.2. Inférence statistique : variables quantitatives continues

Test de comparaison d'échantillons

- 2 groupes d'individus distincts: test-t de student

2. Statistique

2.2. Inférence statistique : variables quantitatives continues

Test de comparaison d'échantillons

- 2 groupes d'individus distincts: test-t de student
- 3 groupes ou plus d'individus distincts : Analyse de variance à un ou plusieurs facteurs

2. Statistique

2.2. Inférence statistique : variables quantitatives continues

Test de comparaison d'échantillons

- 2 groupes d'individus distincts: test-t de student
- 3 groupes ou plus d'individus distincts : Analyse de variance à un ou plusieurs facteurs
- 2 groupes pairés: Avant et après traitement : test t-pairé

2. Statistique

2.2. Inférence statistique : variables quantitatives continues

Test de comparaison d'échantillons

- 2 groupes d'individus distincts: test-t de student
- 3 groupes ou plus d'individus distincts : Analyse de variance à un ou plusieurs facteurs
- 2 groupes pairés: Avant et après traitement : test t-pairé
- 3 groupes pairés: Anova à mesures répétées à un, deux ou trois facteurs

2. Statistique

2.2. Inférence statistique : variables quantitatives continues

Prédire une variable

- 2 groupes d'individus distincts: test-t de student
- 3 groupes ou plus d'individus distincts : Analyse de variance à un ou plusieurs facteurs
- 2 groupes pairés: Avant et après traitement : test t-pairé
- 3 groupes pairés: Anova à mesures répétées à un, deux ou trois facteurs
- Prédire une variable : Régression linéaire, quadratique, polynomiale.....

2. Statistique

2.2. Inférence statistique : variables quantitatives continues

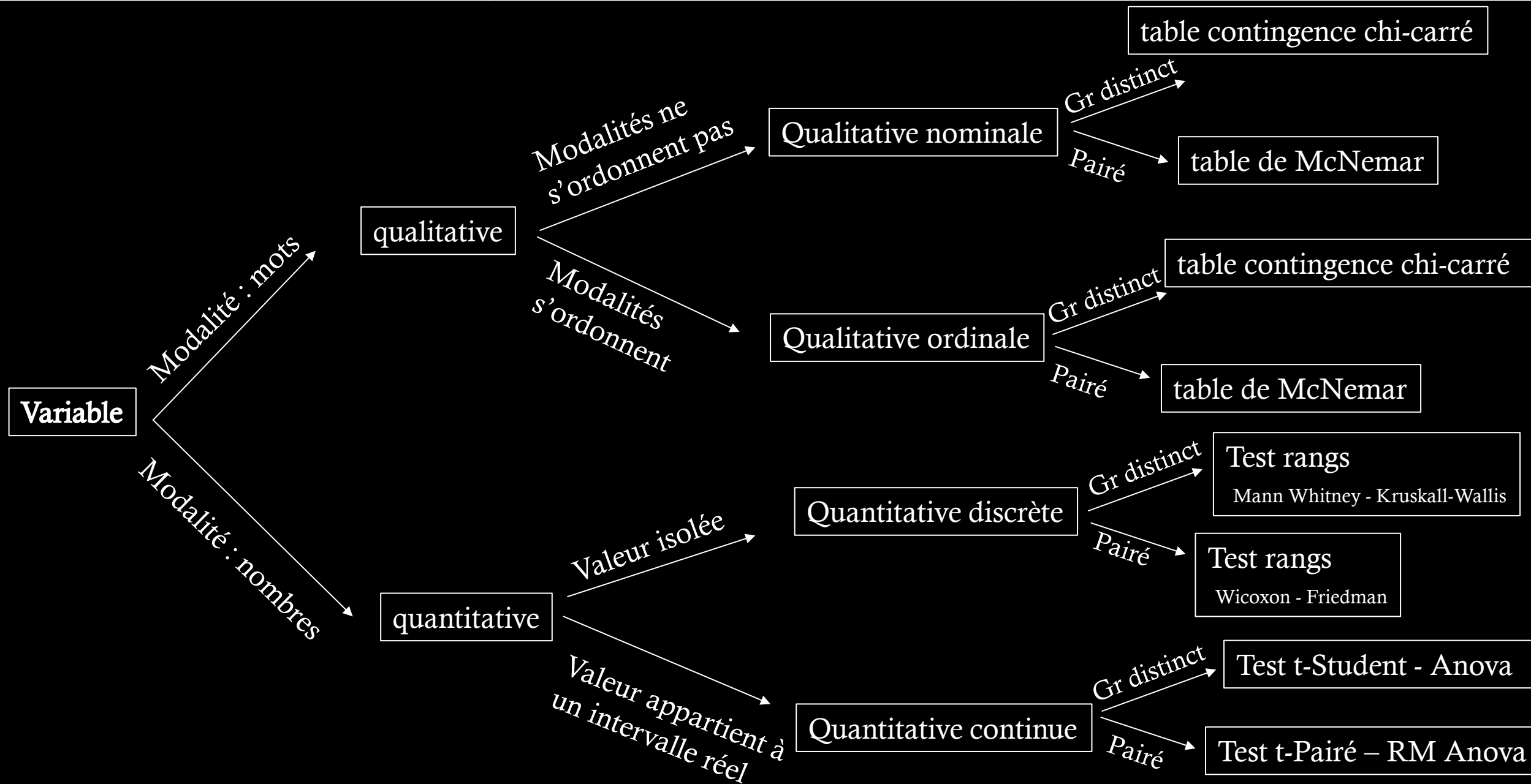
Trouver une association entre deux variables

- 2 groupes d'individus distincts: test-t de student
- 3 groupes ou plus d'individus distincts : Analyse de variance à un ou plusieurs facteurs
- 2 groupes pairés: Avant et après traitement : test t-pairé
- 3 groupes pairés: Anova à mesures répétées à un, deux ou trois facteurs
- Prédire une variable : Régression
- Trouver une association entre deux variables : coefficient de corrélation de Pearson

1. Type de données

2. Tests statistiques

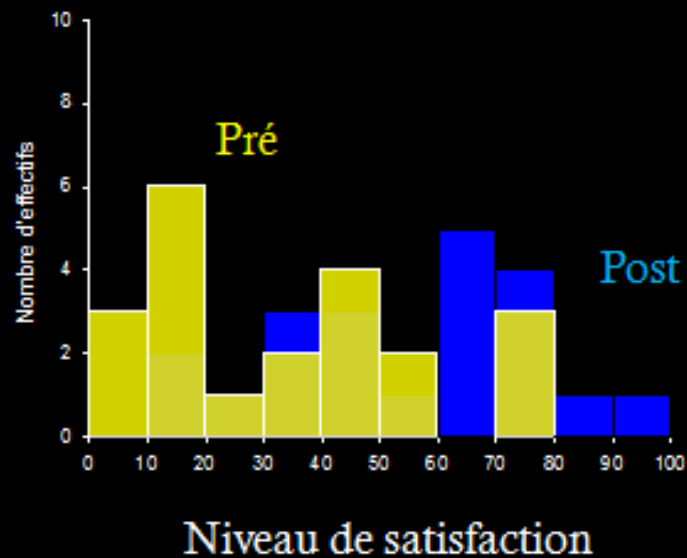
3. Graphes



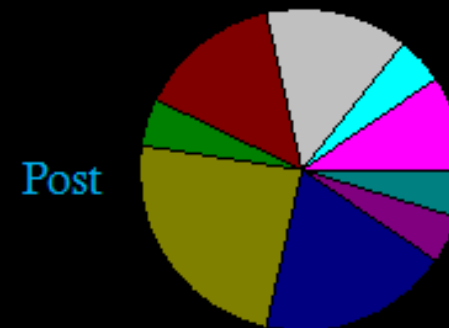
3. Graphes

3.1. Graphes variables qualitatives

Diagramme de fréquence



Pie

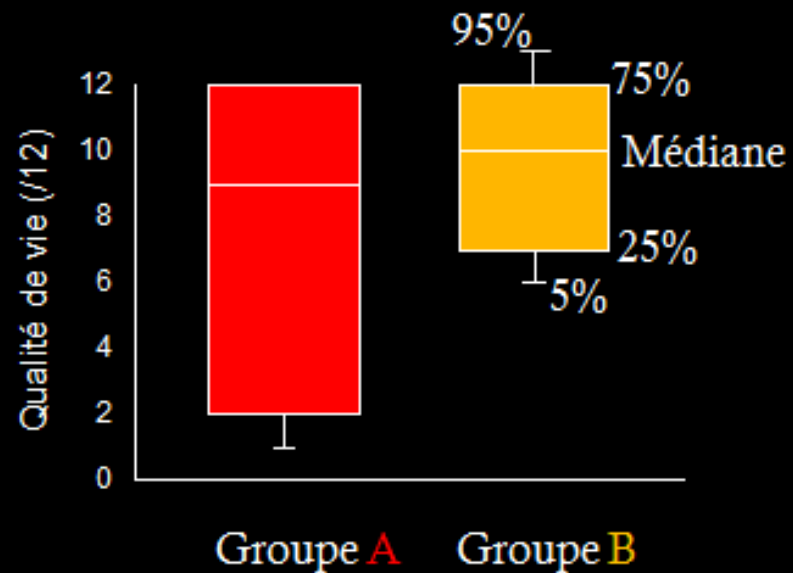


Niveau de satisfaction

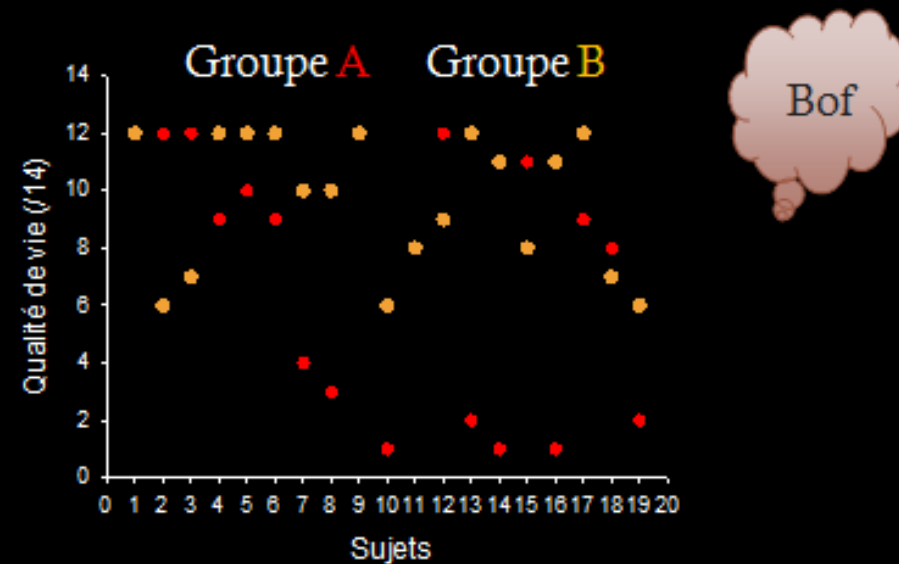
3. Graphes

3.1. Graphes variables quantitatives discrètes

Box plot



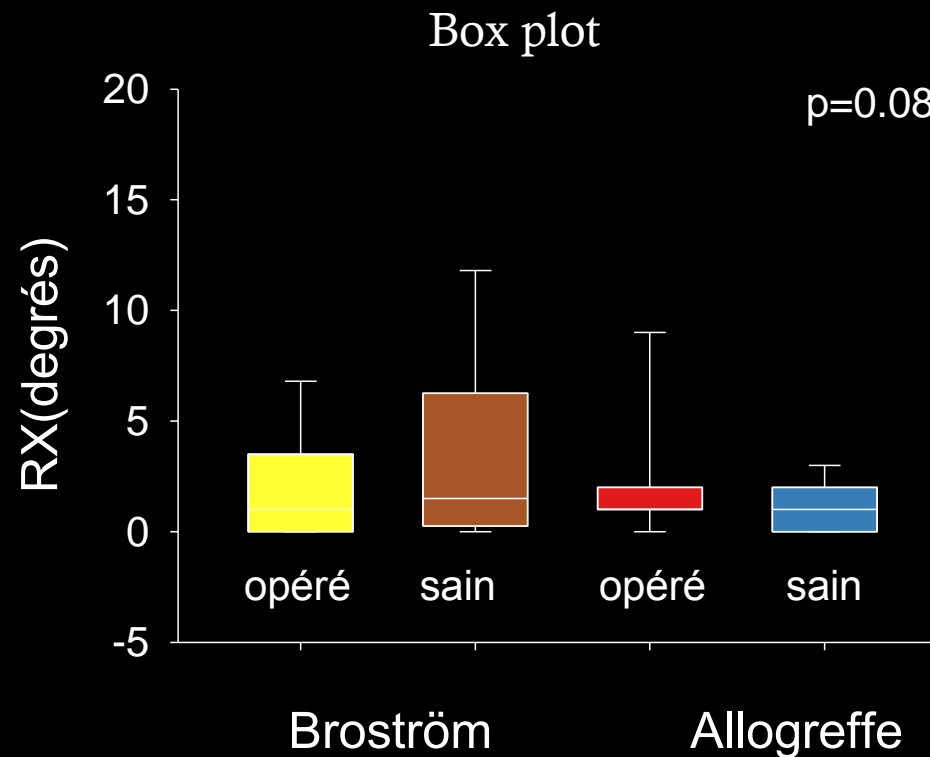
Plot point



3. Graphes

3.1. Graphes variables quantitatives continues

Mais qui ne respectent pas le test de Shapiro



3. Graphes

3.1. Graphes variables quantitatives continues

Plot XY, graphe 3D.....beaucoup de possibilités

