

# Analyses statistiques : que réaliser, que présenter dans la section résultats ?



Voir l'ensemble des ressources disponibles et ce fichier mis à jour sur <http://objectifthese.org>



# En théorie, tout serait possible. En pratique, présentation normée.

- En théorie, une foule d'indicateur calculables
- En pratique, normes internationales

<https://www.equator-network.org/>



- Nous listerons ici ce qu'il faut calculer et présenter.
- Généralement très succinct !

 Reporting guidelines for main study types		
<a href="#">Randomised trials</a>	<a href="#">CONSORT</a>	<a href="#">Extensions</a>
<a href="#">Observational studies</a>	<a href="#">STROBE</a>	<a href="#">Extensions</a>
<a href="#">Systematic reviews</a>	<a href="#">PRISMA</a>	<a href="#">Extensions</a>
<a href="#">Study protocols</a>	<a href="#">SPIRIT</a>	<a href="#">PRISMA-P</a>
<a href="#">Diagnostic/prognostic studies</a>	<a href="#">STARD</a>	<a href="#">TRIPOD</a>
<a href="#">Case reports</a>	<a href="#">CARE</a>	<a href="#">Extensions</a>
<a href="#">Clinical practice guidelines</a>	<a href="#">AGREE</a>	<a href="#">RIGHT</a>
<a href="#">Qualitative research</a>	<a href="#">SRQR</a>	<a href="#">COREQ</a>
<a href="#">Animal pre-clinical studies</a>	<a href="#">ARRIVE</a>	
<a href="#">Quality improvement studies</a>	<a href="#">SQUIRE</a>	
<a href="#">Economic evaluations</a>	<a href="#">CHEERS</a>	
See all <a href="#">432 reporting guidelines</a>		



# Trois volets

- Analyses univariées
  - Descriptives : description de la distribution d'une seule variable à la fois
  - Inférentielles : intervalles de confiance des estimateurs (généralement moyenne, proportion)
- Analyses bivariées
  - Généralement tester l'indépendance entre deux variables
- Analyses multivariées (ou multivariables)
  - Généralement expliquer (modéliser) une variable par un ensemble d'autres variables

Le champ de la statistique est plus large, mais généralement une thèse se limite à cela.

# Rappels typographiques

## En Français

- Nombres :
  - 123 456,78
  - 123456,78
  - (éviter 123.456,78)
  - 1,23<sup>E6</sup>
- Intervalles :
  - [123 ; 456]
  - (jamais de tiret !!!)

## En Anglais

- Nombres :
  - 123,456.78
  - (123456.78)
  - 1.23<sup>E6</sup>
- Intervalles :
  - [123; 456]
  - (jamais de tiret !!!)



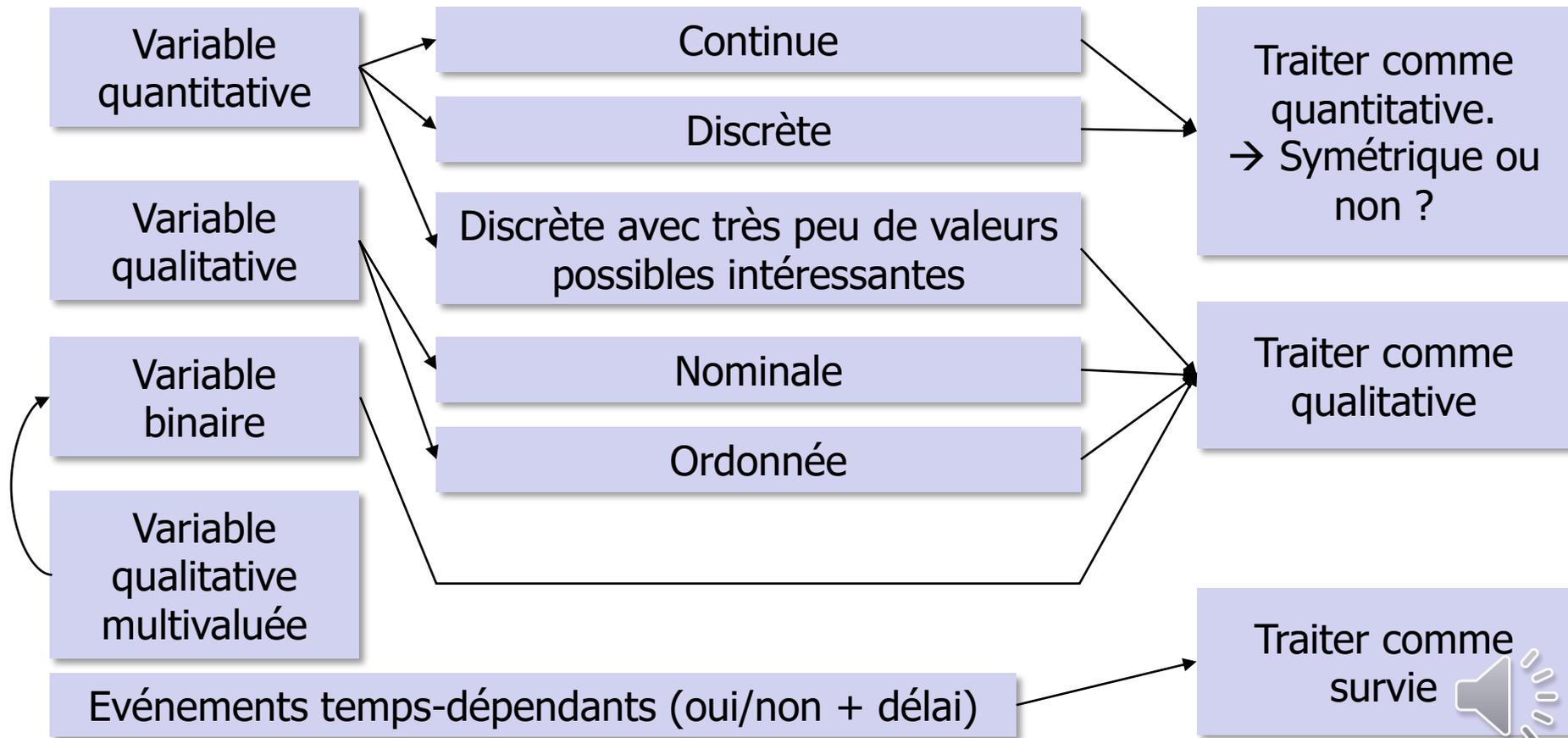
# Analyses univariées

(descriptives et parfois inférentielles)



# Analyses univariées

- Identifier d'abord le type de la variable :



# Variables quantitatives

Paragraphe « statistique » de la section « matériel et méthodes »

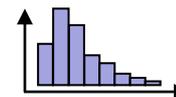
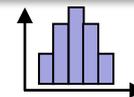
Les variables quantitatives sont exprimées en moyenne et écart type (DS, déviation standard) si l'histogramme révèle une distribution d'allure symétrique, et médiane premier et troisième quartile (Q1, Q3) dans le cas contraire.

- Histogramme :
  - A visualiser systématiquement => décision
  - A produire très rarement dans le document

Exemples de rédaction en « résultats »

Le poids moyen est de 73,5 kg (DS=12,4).

L'HbA1c médiane est de 5,5% (Q1-Q3 : [5 ; 8]).



# Variables quantitatives : intervalle de confiance d'une moyenne

Paragraphe « statistique » de la section « matériel et méthodes »

Les intervalles de confiance des moyennes à 95% (IC95) sont calculés à l'aide d'une loi [choisir] de Student / normale.

- ATTENTION :
  - pas recommandé de les produire systématiquement
  - que si l'objectif principal est de mesurer cette moyenne

Exemples de rédaction en « résultats »

Le poids moyen est de 73,5 kg (IC95 : [70,0 ; 77,0]).



# Variables qualitatives

Paragraphe « statistique » de la section « matériel et méthodes »

Les variables qualitatives, binaires, ou discrètes avec très peu de modalités sont exprimées en effectif et pourcentage.

Exemples de rédaction en « résultats »

Parmi les patients, 13 (26%) ont une atteinte de grade I, 15 (30%) une atteinte de grade II et 22 (44%) une atteinte de grade III.

Seulement 2 patients (4%) sont décédés.

L'échantillon comporte 36 (72%) femmes.

Parmi les antécédents, on retrouve 5 cas de diabète (10%), 3 cas d'infarctus (6%) (...), étant entendu qu'un patient peut présenter plusieurs antécédents.

*Garder un ordre logique (croissant/décroissant), ou par fréquence décroissante.*

*Si binaire, on n'écrit que la modalité « spéciale »*

*On met en avant le sexe le plus fréquent uniquement*

*Présentation similaire pour les qualitatives multivaluées*



# Variables qualitatives : intervalle de confiance d'une proportion

Paragraphe « statistique » de la section « matériel et méthodes »

Les intervalles de confiance des proportions à 95% (IC95) sont calculés à l'aide d'une loi [choisir] normale / binomiale.

- ATTENTION :
  - pas recommandé de les produire systématiquement
  - que si l'objectif principal est de mesurer cette proportion
  - pas adapté pour les événements temps-dépendants

Exemples de rédaction en « résultats »

La proportion de formes graves s'élève à 12,0% (IC95 : [9,5 ; 14,5]).



# Variables de survie

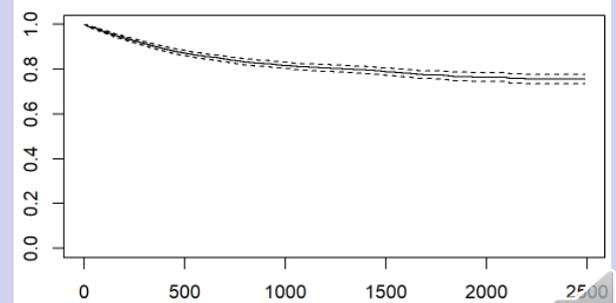
Paragraphe « statistique » de la section « matériel et méthodes »

Les survies sont étudiées avec l'estimateur de Kaplan-Meier. Les intervalles de confiance des survie à 95% (IC95) sont calculés à l'aide d'une loi normale.

- Il est bienvenu de dessiner la courbe de Kaplan-Meier. Les taux de survie et IC95 sont calculés en même temps.

Exemples de rédaction en « résultats »

Le taux de survie sans conversion était de 99,8% [99,7 ; 99,9] à 30 jours, de 93,7% [92,9 ; 94,5] à 1 an, et de 83,7% [82,0 ; 85,3] à 5 ans (Figure 1).



# Analyses bivariées : tester l'indépendance statistique entre deux variables

Recommandations pour une thèse de médecine et la plupart des journaux.

Pour certains journaux scientifiques, une seule p valeur est autorisée par article (quasiment), et ces recommandations ne s'appliquent donc pas.



# Préambule

- Comparer des proportions :
  - Comparer la proportion de malades entre les groupes A, B et C
  - Revient à
  - Tester l'indépendance entre la variable qualitative  $X$  (avec  $X=1$  pour les malades et  $X=0$  pour les autres), et la variable qualitative qui détermine le groupe  $\{A; B; C\}$
- Comparer des moyennes :
  - Comparer la moyenne de  $X$  entre les groupes A, B et C
  - Revient à
  - Tester l'indépendance entre la variable quantitative  $X$ , et la variable qualitative qui détermine le groupe  $\{A; B; C\}$
- Cela revient toujours à tester l'indépendance entre 2 variables



# Préambule

- Tester l'indépendance entre 2 variables X et Y :

		Nature de X	
		Qualitative	Quantitative
Nature de Y	Qualitative	Quali-quali	Quali-quant
	Quantitative	Quali-quant	Quanti-quant

- La notion d'indépendance est « commutative »  
=> 3 cadres seulement



# Bivarié qualitatif - qualitatif

Paragraphe « statistique » de la section « matériel et méthodes »

L'indépendance entre deux variables qualitatives est testée à l'aide d'un test [choisir] exact de Fisher / du  $\text{Khi}^2$ .

- Choix du test :
  - Le test exact de Fisher est toujours valide, pas toujours réalisable
  - $\text{Khi}^2$  : seulement si effectifs théoriques tous  $\geq 5$
- Graphique : on peut juxtaposer des camemberts... très facultatif

Exemples de rédaction en « résultats »

La proportion de formes sévères est significativement plus élevée chez les femmes que chez les hommes (respectivement 45,8% contre 23,5%,  $p=0,0003$ ).



# Bivarié qualitatif - quantitatif

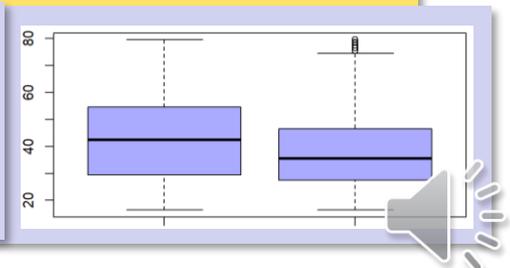
Paragraphe « statistique » de la section « matériel et méthodes »

L'indépendance entre une variable qualitative et une variable quantitative est testée à l'aide [choisir] d'un test de Student / d'une analyse de la variance (ANOVA) / d'un test de Wilcoxon-Mann-Whitney.

- Choix du test :
  - Student : seulement si 2 groupes ET [  $X \sim$  Normale OU  $n > 30$  dans chaque groupe ]
  - ANOVA : si 2 groupes ou plus, possible notamment si  $X \sim$  Normale
  - Wilcoxon-Mann-Whitney : seulement si 2 groupes, sur les rangs, toujours possible
- Boxplots bienvenues si c'est un sujet important

Exemples de rédaction en « résultats »

Les femmes sont significativement plus âgées que les hommes (en moyenne respectivement 42,5 ans contre 38,2 ans,  $p=0,0003$ , voir figure 1).



# Bivarié quantitatif - quantitatif

Paragraphe « statistique » de la section « matériel et méthodes »

L'indépendance entre deux variables quantitatives est testée à l'aide [choisir] du test de nullité du coefficient de corrélation de Pearson / du test de nullité du coefficient de corrélation de Spearman / du test de nullité de la pente d'une régression linéaire simple.

- Choix du test :
  - Si X et Y suivent une loi normale **OU** si une régression est valide : coefficient de corrélation de Pearson (le test de nullité de la pente est quasiment identique)
  - Sinon, coefficient de corrélation de Spearman, sur les rangs
- Nuage de points bienvenu si c'est un sujet important

Exemples de rédaction en « résultats »

Le score prédit et le score observé sont significativement corrélés  
( $r=0,662$ ,  $r^2=0,438$ ,  $p=6.5e-281$  )



# Analyses multivariées (ou multivariables) : expliquer, prédire, modéliser...



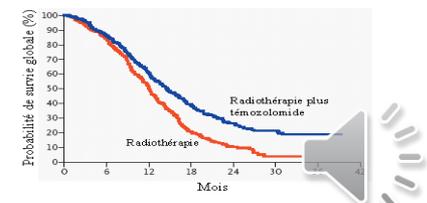
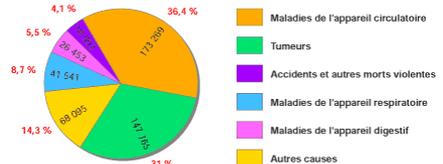
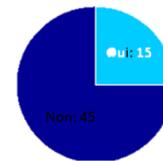
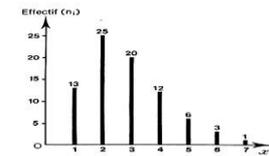
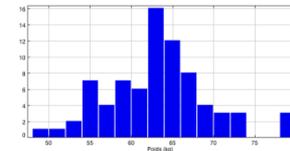
# Généralités

- Personne ne s'attend à ce que vous réalisiez de telles analyses seul :
  - Réalisation d'apparence simple, mais en réalité complexe
  - Validation complexe
  - Interprétation « dangereuse »
- Adressez-vous à un spécialiste
- Si vous faites un Master 2, vidéos et cours disponibles sur <http://emmanuel.chazard.org> et <https://www.youtube.com/user/emmanuelchazard>



# Les régressions multiples

- Permettent d'expliquer une variable Y à l'aide de plusieurs variables  $X_i$
- Y variable quantitative  
=> régression linéaire ex : poids
- Y variable de décompte  
=> régression de Poisson ex : nb hospitalisations
- Y binaire  
=> régression logistique ex : malade 1/0
- Y qualitative  
=> régression logistique multinomiale ex : cause décès
- Y événement dans le temps  
=> modèle de Cox ex : survie sous traitement



# Généralités sur la présentation

Paragraphe « statistique » de la section « matériel et méthodes »

[on énonce les tests et modèles utilisés. On précise comment les covariables sont sélectionnées]

Les résultats sont exprimés en termes de coefficients assortis d'un intervalle de confiance à 95%.

[voir le modèle de document Word fourni sur <http://www.chazard.org/emmanuel/theses.htm>]

- Présentation centrée sur les ratios et leurs IC95, en texte ou tableau, et volontiers en graphique.

Exemples de rédaction en « résultats »

Les facteurs associés à une augmentation du risque étaient (HR et IC95%) : une synovectomie (1,90 [1,34 ; 2,70]), et un geste osseux sur le col ou l'acétabulum (1,82 [1,36 ; 2,43]). Les facteurs protecteurs étaient...

