



# Analyses statistiques de base

Atelier de formation A02

Plateforme de recherche clinique et évaluative

Anne-Sophie Julien, M.Sc., biostatistique



# Objectifs d'apprentissage

1. Connaître les étapes d'un projet reliées aux statistiques
2. Choisir les statistiques pour valider et décrire adéquatement un échantillon
3. Déterminer quelle analyse statistique doit être utilisée selon le contexte
4. Interpréter les résultats des principaux tests statistiques



# Plan de la présentation

1. Variables
2. Statistiques descriptives
3. Choix d'analyses
4. Interprétation de résultats
5. Liste d'analyses statistiques & Exemples
6. Exercices



# Étapes d'un projet

## Avant la collecte de données

- A01 • Établir les objectifs de recherche
- A01 • Choisir le devis de recherche optimal
- A03 • Déterminer comment les variables seront mesurées
- A03 • Calculer la taille d'échantillon
- Planifier les analyses statistiques
- **\*\*Consulter un statisticien\*\*** (*mieux vaut prévenir que guérir*)

## Après la collecte des données

- A02 • Valider les données
- A02 • Effectuer les analyses statistiques prévues a priori
- A02 • Vérifier que les résultats obtenus sont fiables (postulats)
- A02 • Interpréter les résultats de façon adéquate et transparente



# Section 1: Variables et Statistiques Descriptives



# Variables & Statistiques descriptives

## Variables

Objectif -> Mesure -> Variable -> Test statistique

## Statistiques descriptives

- **Validation**
  - Valeurs plausibles
  - Absence d'erreur de saisie
  - Minimum, Maximum
  - Croisements plausibles
  - Critères inclusion / exclusion respectés
- **Description**
  - Reproductibilité
  - Utilisation / Application des résultats
  - Généralisation



# Type de variables

- Variables **QUALITATIVES**

	Dichotomique (2 modalités)	Polytomique (>2 modalités)
Nominale / Catégorique	Sexe (Homme, Femme)	Statut Fumeur (Actuel, Ancien, Non)
Ordinale	Gravité Maladie (Modéré, Sévère)	Degré de satisfaction (Pas du tout, Satisfait, Très satisfait)

- Variables **QUANTITATIVES**

- Continue:

- Poids (163.1 lbs, 48.8 kg, ...)

- Discrète:

- Nombre de crises d'épilepsie (0, 1, 2, 3,...)



# Variable qualitative

- **Fréquence** : nombre d'occurrences pour une catégorie
  - **Proportion** (varie de 0 à 1)
  - **Pourcentage** (proportion x 100)

Satisfaction	%	% cumulé
Très insatisfait	22	22
Insatisfait	54	76
Satisfait	24	100

- **Fréquence cumulée**: Nombre d'observations qui se situent au-dessus (ou au-dessous) d'une valeur particulière (variable ordinale)
- **Mode** : la valeur la plus fréquente dans un ensemble de données





# Variable qualitative

## Représentation graphique

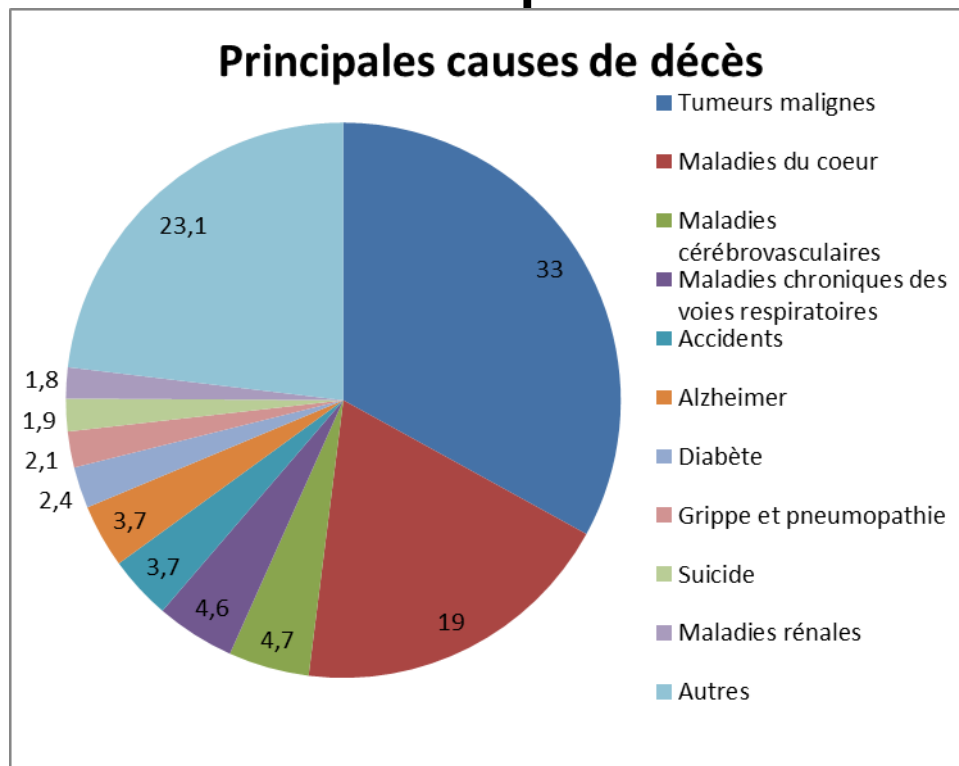


Diagramme circulaire  
(tarte / secteurs)

Source: ISQ, 2013

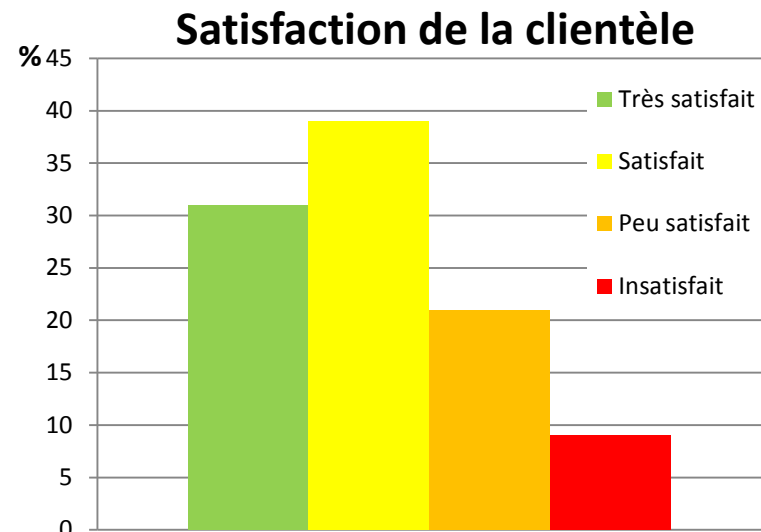


Diagramme à bandes



# Variable quantitative

## Mesures de localisation (de tendance centrale)

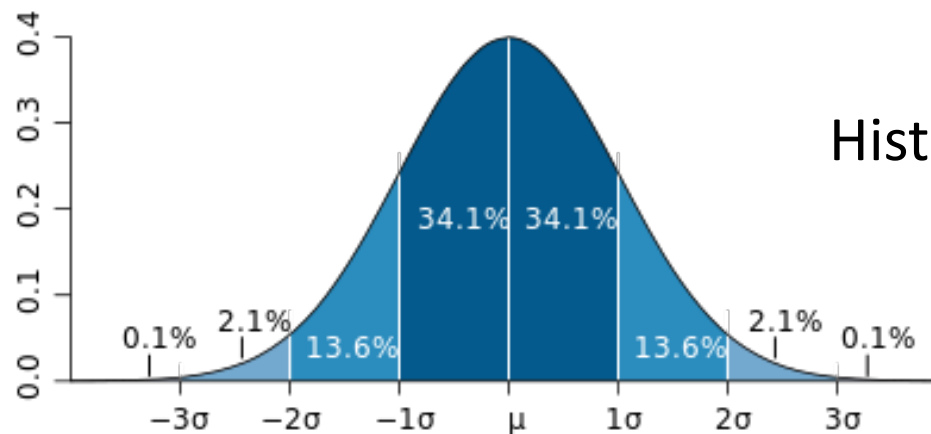
- **Moyenne** : somme des valeurs numériques divisée par le nombre de données (N)
- **Médiane** : une valeur qui permet de couper l'ensemble de données en deux parties égales



# Variable quantitative

## Mesures de dispersion (associées à la moyenne)

- **Variance** : une mesure de dispersion des données autour de la moyenne
- **Écart-type** : racine carré de la variance =  $\sigma$



Histogramme d'une  
loi normale



# Variable quantitative

## Mesures de dispersion (associées à la moyenne)

### Écart-type (Standard Deviation)

- Indicateur de la variabilité des données
- Présenté dans les statistiques descriptives

### Erreur-Type (Standard Error)

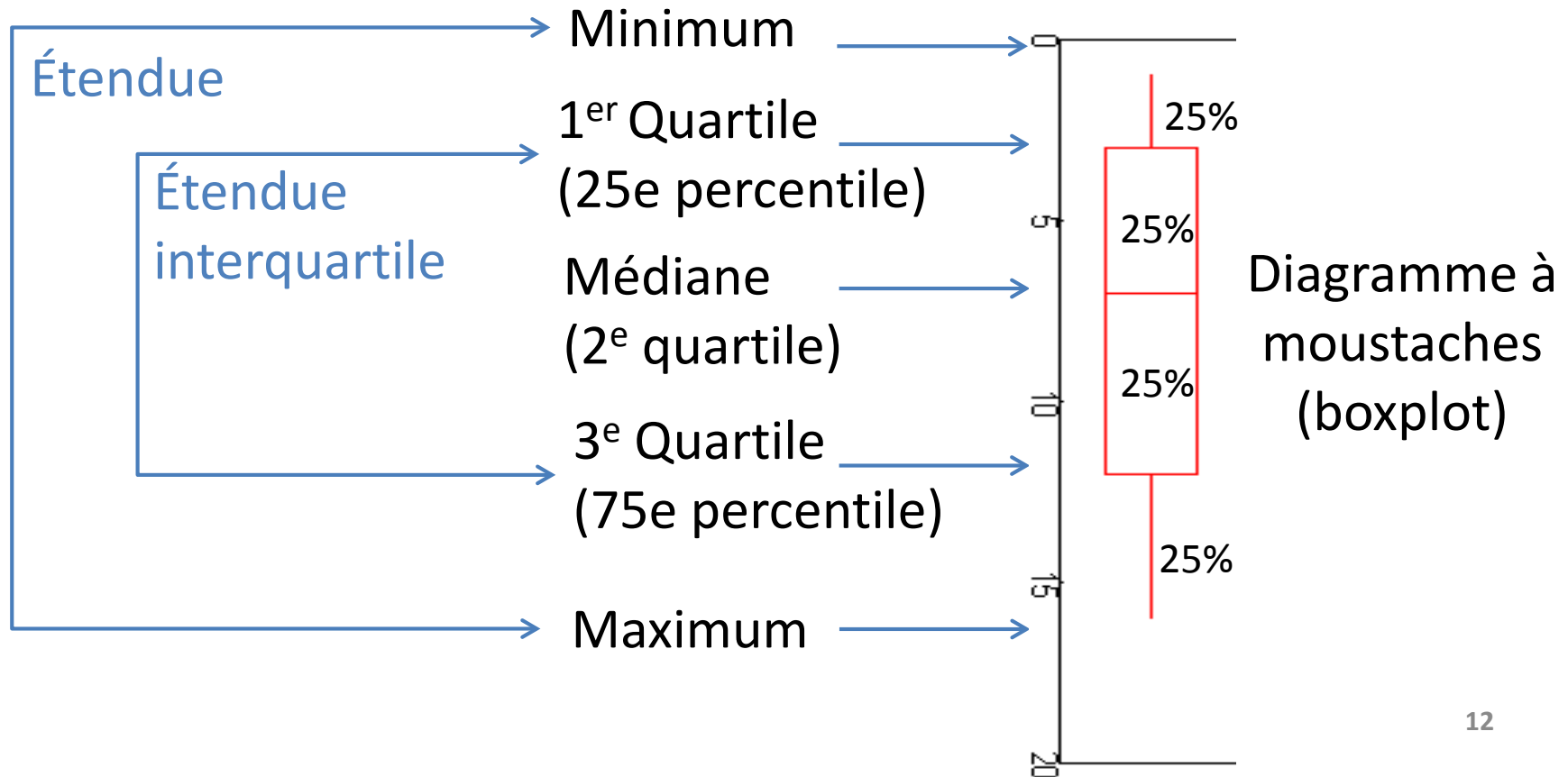
- Écart-type divisé par la racine carrée de N
- Mesure l'incertitude associée à l'estimation de la moyenne
- Présentée dans les résultats d'analyses

Référence: Nagele, P. (2001). Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *British Journal of Anaesthesia*, 90 (4): 514-16.



# Variable quantitative

## Mesures de dispersion (associées à la médiane)





**TABLE 1. PATIENT CHARACTERISTICS**

	Total	GOLD I	GOLD II	GOLD III	GOLD IV	P Value*
Patients, n (%)	170 (100)	34 (20)	57 (34)	43 (25)	36 (21)	
Age, yr, mean (SD)	64.0 (6.6)	66.3 (5.6)	63.3 (6.8)	63.2 (7.3)	63.7 (6.2)	0.15
Males, n (%)	128 (75)	25 (74)	41 (72)	35 (81)	27 (75)	0.74
Pack-years of smoking, mean (SD)	51.9 (24.8)	46.9 (26.6)	50.7 (27.0)	55.6 (23.4)	54.0 (20.5)	0.44
Current smokers, n (%)	72 (42)	15 (44)	27 (47)	23 (53)	7 (19)	<0.01
Regular drinkers, n (%)	79 (46)	17 (50)	29 (51)	22 (51)	11 (31)	0.20
Drinks per week of regular drinkers, median (IQR)	5 (2–12)	4 (2–7)	5 (2–10)	5.5 (2–14)	9.5 (3–14)	0.83
Retired or early retired, n (%)	133 (78)	27 (79)	40 (70)	33 (77)	33 (92)	0.11
Unemployed, n (%)	16 (9)	3 (9)	6 (11)	5 (12)	2 (6)	0.81
University entrance degree, n (%)	16 (9)	3 (9)	7 (12)	4 (9)	2 (6)	0.76

*Definition of abbreviations:* GOLD = Global Initiative for Chronic Obstructive Lung Disease; IQR = interquartile range.

\* P values were tested by analysis of variance if the variable is stated as mean (SD) or by  $\chi^2$  test if the variable is stated as n (%); P value for drinks per week was tested by Kruskal-Wallis test, as the variable was skewed.

*Exemple: Extrapulmonary Effects of Chronic Obstructive Pulmonary Disease on Physical Activity, A cross-sectional Study.*

147 individus avec MPOC.

GOLD = stade de la maladie.



## Section 2: Choisir & interpréter



# Choix d'analyse

Plusieurs facteurs à considérer:

- L'objectif
- La nature de la **variable dépendante** (ce qu'on tente d'expliquer)
- La nature de la **variable indépendante** (les facteurs prédictifs)
- Le **type de données**
  - Non corrélées (p.ex.: mesures sur des patients différents)
  - Corrélées (p.ex.: plusieurs mesures sur un même patient, mais dans certains cas, patients différents avec caractéristiques similaires, comme dans un CRT. Termes similaires: **pairées**, **appariées**)





# Choix d'analyse

Plusieurs facteurs à considérer (suite):

- Le **nombre de variables**
  - Univariée (1 variable: distribution, descriptif)
  - Bivariée (2 variables: association)
  - Multivariée (>2 variables: association): Pour sélectionner les meilleurs prédicteurs, ou tester une relation en ajustant pour des facteurs confondants
- Le **design expérimental** (étude observationnelle, randomisée, cas-témoins avec ou sans appariement, etc.)



# Validation de modèles paramétriques

- Postulats
  - Hypothèses du modèle doivent être respectés (i.e.: normalité, homoscédasticité, linéarité, indépendance,...)
- Sources de biais
  - Sélection non aléatoire des sujets
  - Absence de randomisation
  - Hypothèses  $H_0$  &  $H_1$  mal spécifiées
  - Valeurs influentes
  - Multicolinéarité
  - Données manquantes
- Pouvoir prédictif
  - Pouvoir explicatif des variables explicatives
  - Validation des prédictions
  - Validation croisée

**Sans validation, la conclusion d'une analyse pourrait être erronée!**



# Résultats d'un test

- **Mesure d'association**
  - Taille d'effet, rapport de risques ou cotes, coefficient de régression ou corrélation, etc
  - Indique le sens et la force de l'association
- **Intervalle de confiance**
  - Précision et incertitude des mesures d'association
  - Plusieurs IC générés, X% contiendront la valeur populationnelle
- **Valeur p**
  - Valeur continue plutôt que dichotomique
  - Indicateur de la compatibilité des données échantillonnelles avec la valeur d'intérêt dans la population sous l'hypothèse nulle



# Interprétation des résultats

- Valeur p
  - Éviter sig, ns,  $p < 0.05$ ,  $p > 0.05$ , \*\*\*, ...
- Une valeur  $p > 0,05$  n'implique pas qu'il n'y a pas d'effet significatif
  - ne rejette pas l'hypothèse nulle
  - examiner l'intervalle de confiance
  - vérifier la puissance (pas a posteriori)
- Tenir compte de la **significativité clinique**
- Présenter les **biais** et les discuter; généralisation, inférence



# Interprétation des résultats

- Association et relation ne sont pas synonymes de corrélation
- Association n'implique pas causalité
- Exemples:
  - Nombre de pompiers & dommages de l'incendie
  - Notes examen avec et sans tutorat
- Tests de comparaison au baseline (Table 1) entre groupes dans un RCT
  - Regarder les valeurs cliniques plutôt



## Section 3: Liste d'analyses statistiques



# Série d'analyses 1

Pour tester l'association entre des variables qualitatives

Tests	Variable dépendante	Variable(s) indépendante(s)	Type de données	Type d'analyse
Khi-Deux	Nominale	Nominale	Non corrélées	Bivariée
Fisher Exact	Dichotomique	Dichotomique	Non corrélées	Bivariée
Mc Nemar	Dichotomique	Dichotomique	Corrélées	Bivariée
Régression logistique	Dichotomique	Qualitative(s) + Quantitative(s)	Non corrélées	Bivariée ou Multivariée

Pour quantifier l'association entre deux variables dichotomiques

- Rapport de cotes
- Rapport de risques



# Khi-Deux de Pearson

Variables: IMC (3 catégories), hypertension (oui/non), sexe.

Questions:

- Existe-t-il une relation entre l'hypertension et l'IMC?
- La distribution de l'IMC est-elle la même entre les deux sexes?

	Insuffisance	Normal	Surpoids
Hypertension	35%	30%	45%
Normale	65%	70%	55%

	Homme	Femme
Insuffisance	4%	10%
Normal	36%	40%
Surpoids	60%	50%

Avec  $p < 0,05$ , on peut conclure que:

- Il existe une relation entre l'hypertension et l'IMC
- La distribution de l'IMC n'est pas la même selon le sexe





# Test exact de Fisher

Variables: Hypertension (oui/non), sexe.

Questions:

- Est-ce que la proportion d'hypertension est la même chez les deux sexes?
- Existe-t-il une association entre l'hypertension et le sexe?

	Homme	Femme
Hypertension	37%	31%
Normale	63%	69%

Avec  $P > 0,05$ , on ne peut pas rejeter l'hypothèse que la proportion d'hypertension soit la même chez les hommes que les femmes.



# Test de McNemar

Variables: Acouphènes (présence/absence), traitement (ou placebo). Mesurées dans les 2 oreilles randomisées d'un patient

Question: Est-ce que la proportion d'oreilles avec acouphènes est plus faible avec le traitement que le placebo?

		Acouphène	
		Présence	Absence
Oreille	Traitée	15%	85%
	Placebo	35%	65%

Avec  $p < 0.05$ , on peut conclure que la présence d'acouphènes diffère significativement, i.e. le traitement semble efficace ici



# Régression logistique

Variables: Malade (ou sain), Fumeur (ou non), Poids santé (ou surpoids), traitement (ou placebo)

## Questions:

- Existe-t-il une association entre le tabac et la maladie?
- Quel est l'impact du tabac et du poids sur la maladie?
- Est-ce que le traitement prévient la maladie, en ajustant pour le tabac et le poids?
- Quels sont les meilleurs prédicteurs de la maladie?
- Quels sont les facteurs de risques ou protecteur de la maladie?

## Interprétation:

- Coefficient de régression
  - Rarement utilisé
- Rapport de cotes
  - Mesure d'association
  - Voir diapositives suivantes



# Rapport de cotes

Deux variables dichotomiques

VD - Cas/Témoins, Malade/Sain, Décès/Vivant...

VI - Exposé/Non exposé, Groupe 1/Groupe 2, Trt A/Trt B...

	Cas	Témoin	Total
Exposé	A	B	A+B
Non exposé	C	D	C+D
Total	A+C	B+D	A+B+C+D

**Cote** = Chances d'être un cas vs témoin (odds)  
=  $A/B$  et  $C/D$

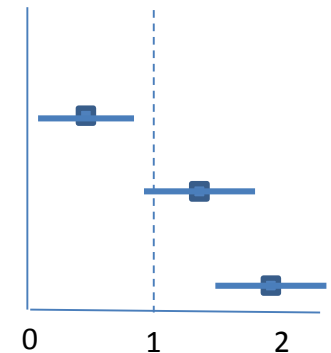
**RC** = Cote des exposés vs Cote des non exposés (odds ratio)  
=  $(A/B) / (C/D) = AD/BC$



# Rapport de cotes

## Interprétation

- De 0 à l'infini
- 1 = Absence d'effet
- Supérieur à 1 = La cote est plus grande dans le groupe exposé que non exposé
  - L'exposition est un facteur de risque\*
- Inférieur à 1 = La cote est plus petite dans le groupe exposé que non exposé
  - L'exposition est un facteur de protection\*
- Pas un risque, sauf si l'événement est rare



\*Avec codification: 1 = Défavorable (Événement), 0 = Favorable (Référence)



# Rapport de cotes

## Exemples

VD

VI

- $RC = 2,30$  (Malade vs Sain selon Fumeur vs Non Fumeur)
  - La cote de maladie chez les fumeurs est 2,30 fois celle des non fumeurs

VD

VI

- $RC = 0,44$  (Malade vs Sain selon Poids Santé vs Surplus de poids)
  - La cote de maladie chez les patients avec poids santé est 0,44 fois celle de ceux avec surplus de poids
  - La cote de maladie est 56% moins élevée chez les patients avec poids santé que ceux avec surplus de poids  $(1 - 0,44) * 100\%$



# Rapport de risques

- Comparaison du risque entre deux groupes
- Deux variables dichotomiques
- RR = Probabilité d'être malade chez les exposés divisée par la probabilité d'être malade chez les non exposés  
=  $(A/(A+B)) / (C/(C+D))$
- Contrairement au rapport de cotes, c'est une **probabilité**
- N'est pas équivalent au rapport de cotes, sauf si la probabilité de l'événement est faible



# Exemple: Étude Cas-Témoins

*Case-Control Study of Human Papillomavirus and Oropharyngeal Cancer.*

## Participants

Cas: Patients atteints de cancer oropharyngé (100)

Contrôles: -Patients en provenance du même hôpital, sans cancer  
-2 contrôles pairés par âge et sexe pour chacun des cas

## Mesures

- Histologie
- Questionnaires (socio-démographique, antécédants, habitudes de vie)

## Objectif

- Existe-t-il une relation entre le VPH et l'apparition du cancer oropharyngé?





**Table 1.** Explanatory Variables for Patients with Oropharyngeal Cancer and Control Patients.\*

Explanatory Variable	Patients with Oropharyngeal Cancer (N=100)	Control Patients (N=200)	Unadjusted Odds Ratio (95% CI)†
	<i>number (percent)</i>		
<b>Demographic characteristics</b>			
<b>Sex</b>			
Female	14 (14)	28 (14)	1.0
Male	86 (86)	172 (86)	1.0 (0.5–2.0)
<b>Age</b>			
<50 yr	34 (34)	68 (34)	1.0
50–64 yr	51 (51)	102 (51)	1.0 (0.6–1.7)
≥65 yr	15 (15)	30 (15)	1.0 (0.5–2.0)
<b>Highest educational level</b>			
Some high school	11 (11)	15 (8)	1.0
High-school graduate or some college	41 (41)	71 (36)	0.8 (0.3–1.9)
College graduate	48 (48)	114 (57)	0.6 (0.3–1.4)‡
<b>Race or ethnic group§</b>			
White, non-Hispanic	87 (87)	171 (86)	1.0
Black, non-Hispanic	9 (9)	17 (8)	1.0 (0.5–2.4)
Other	4 (4)	12 (6)	0.7 (0.2–2.1)
<b>Home state</b>			
Maryland	50 (50)	138 (69)	1.0
Other	50 (50)	62 (31)	2.2 (1.3–3.6)

## Analyses

- Régression logistique  
→ Est-ce que les caractéristiques des patients sont associées à la présence de cancer?



**Table 3.** Association of Oropharyngeal Cancer with Exposure to HPV and with Biomarkers of Cancer Associated with HPV-16.

Measure of HPV Exposure or Disease	Prevalence		Odds Ratio (95% CI)	
	Case Patients (N=100) <i>number (percent)</i>	Control Patients (N=200) <i>number (percent)</i>	Unadjusted	Adjusted*
<b>HPV-16 L1 serologic status</b>				
Seronegative	43 (43)	186 (93)	1.00	1.00
Seropositive	57 (57)	14 (7)	17.6 (8.8–34.5)	32.2 (14.6–71.3)
<b>Oral HPV-16 infection†</b>				
Negative	68 (68)	192 (96)	1.00	1.00
Positive	32 (32)	8 (4)	11.3 (5.0–25.7)	14.6 (6.3–36.6)
<b>Any oral HPV infection‡</b>				
Negative	63 (63)	189 (94)	1.00	1.00
Positive	37 (37)	11 (6)	10.0 (4.8–20.7)	12.3 (5.4–26.4)
<b>HPV-16 E6 or E7 serologic status</b>				
Seronegative for E6 and E7	36 (36)	192 (96)	1.00	1.00
Seropositive for E6 or E7	64 (64)	8 (4)	33.3 (16.2–68.6)	58.4 (24.2–138.3)
<b>HPV-16 DNA in tumor</b>				
Absent	28 (28)	—	—	—
Present	72 (72)	—	—	—

\* Odds ratios were adjusted for age, sex, tobacco use, alcohol use, dentition and toothbrushing, and presence or absence of a family history of head and neck cancer.

## Analyses

Régression  
logistique  
bivariée et  
multivariée

→ Est-ce que  
l'exposition au  
VPH affecte  
l'apparition de  
cancer, en  
contrôlant ou  
non pour  
certains facteurs



## Série d'analyses 2

Pour comparer des groupes (moyennes) avec une VD continue

Tests	Variable dépendante	Variable indépendante	Type de données	Type d'analyse
Test T	Continue	Dichotomique	Non corrélées	Bivariée
Test T Paired	Continue	Dichotomique	Corrélées	Bivariée
ANOVA (à 1 facteur)	Continue	Qualitative	Non corrélées	Bivariée
ANOVA à mesures répétées*	Continue	Qualitative	Corrélées	Bivariée

\*Analyse non couverte dans le présent atelier



## Série d'analyses 3

Pour comparer des groupes (localisation) avec une VD continue  
-> Normalité non respectée

Tests	Variable dépendante	Variable indépendante	Type de données	Type d'analyse
Wilcoxon	Continue	Dichotomique	Non corrélées	Bivariée
Wilcoxon Rangs signés	Continue	Dichotomique	Corrélées	Bivariée
Kruskall- Wallis	Continue	Qualitative	Non corrélées	Bivariée



# Test T de Student

Variables: Pouls, sexe

Question: Est-ce que le pouls moyen diffère entre les hommes et les femmes?

sexe	Moyenne	Moyenne de l'IC à 95%		Ecart-type	Méthode	Variances	DDL	Valeur du test t	Pr >  t
F	70.2000	65.7751	74.6249	3.5637	Pooled	Equal	8	0.82	0.4333
H	68.8000	67.1811	70.4189	1.3038	Satterthwaite	Unequal	5.052	0.82	0.4466

$P \geq 0,05$  & intervalles de confiance emboîtés: On ne peut rejeter l'hypothèse que le pouls moyen soit le même pour les deux sexes.

**NB**: Variances égales ou inégales



# Rangs Signés de Wilcoxon

Variables: Pression intraoculaire (IOP) provenant des deux yeux de plusieurs patients.

Oeil	Médiane	Q1-Q3
Gauche	15,50	14,20-16,50
Droit	16,80	14,40-17,80

Question: Est-ce que l'IOP de l'oeil gauche diffère de celle de l'oeil droit?

Différence Gauche - Droit					
Moyenne	Médiane	Minimum	Maximum	S	$P \geq  S $
-0,54	-0,50	-4,1	2,1	-1,5	0,81

$P \geq 0,05$ : Il n'y a pas suffisamment d'évidence pour rejeter que la **distribution** de l'IOP est la même dans les deux yeux



# ANOVA à 1 facteur

Variables: VO2Max , IMC (3 groupes)

Question: La moyenne de VO2Max diffère-t-elle entre les groupes d'IMC ?

Source	DDL	Valeur F	Pr > F
imc	2	11.54	0.0016

imc	vo2max LSMEAN
Insuffis	43.5000000
Normal	46.3000000
Surpoids	38.4400000

*Test global significatif*

➤ *comparaisons multiples par paires*

P = 0,0016: Le VO2Max diffère pour au moins un des trois groupes au seuil de 5%.  
Le groupe "Surpoids" a un VO2Max significativement inférieur aux deux autres.

Least Squares Means for effect imc Pr >  t  for H0: LSMean(i)=LSMean(j) Dependent Variable: vo2max			
i/j	1	2	3
1		0.1171	0.0101
2	0.1171		0.0005
3	0.0101	0.0005	





# Exemple: Essai clinique randomisé

*Comparison of the Atkins, Zone, Ornish, and LEARN Diets for Change in Weight and Related Risk Factors Among Overweight Premenopausal Women.*

## Participants

- 311 femmes préménopausées randomisées dans 4 groupes
- Groupes = diètes (livre + formation)

## Mesures aux temps 0, 2, 6, 12 mois

- Nourriture consommée (apport énergétique + % glucides)
- Exercice physique
- Données anthropométriques et métaboliques

## Objectif

- Voir l'impact de diètes avec différents apports en glucides sur la perte de poids et le métabolisme



**Table 1.** Baseline Participant Characteristics\*

Characteristics	Atkins (n = 77)	Zone (n = 79)	LEARN (n = 79)	Omish (n = 76)	All Diets (n = 311)
Demographics					
Age, y	42 (5)	40 (6)	40 (7)	42 (6)	41 (6)
Education, y	16 (2)	16 (2)	16 (2)	16 (2)	16 (2)
Race/ethnicity, No. (%)					
White	59 (76)	52 (66)	59 (75)	<u>Analyse</u> Randomisation → Pas de tests ici → Absence de différence clinique	
Black	2 (3)	7 (9)	6 (7)		
Hispanic	7 (9)	8 (10)	7 (9)		
Asian/Pacific Islander	7 (9)	9 (11)	6 (8)		
Other	2 (3)	3 (4)	1 (1)		
Smokers, No. (%)	2 (3)	4 (5)	4 (5)		
Physical activity, kcal/kg per d	34 (6)	34 (6)	34 (5)		
Anthropometrics					
Weight, kg	86 (13)	84 (12)	85 (14)	86 (10)	85 (12)
Body fat, %	41 (6)	40 (6)	38 (6)	40 (6)	40 (6)
Body mass index†	32 (4)	31 (3)	31 (4)	32 (3)	32 (4)
Waist-hip ratio	0.843 (0.067)	0.841 (0.068)	0.839 (0.066)	0.840 (0.060)	0.841 (0.065)

\*Data are expressed as mean (SD) unless otherwise indicated.



**Table 2.** Mean Dietary Intake and Energy Expenditure by Diet Group and Time Point\*

	Atkins	Zone	LEARN	Ornish	<i>P</i> Value†
Energy, kcal/d					
Baseline	1888 (512)	1975 (567)	1925 (553)	1850 (541)	.52
12 mo	1599 (494)	1594 (523)	1654 (492)	1505 (437)	.43
Carbohydrate (% energy)					
Baseline	45.6 (10.5)	47.1 (9.0)	48.3 (9.2)	47.9 (8.6)	.31
12 mo	34.5 (14.4) <sup>a</sup>	45.4 (11.1) <sup>b</sup>	47.2 (8.9) <sup>b,c</sup>	52.4 (12.3) <sup>c</sup>	<.001

\*Data presented are unadjusted raw data with no imputations for missing data. Standard deviations are presented in parentheses. Sample sizes for baseline and 2, 6, and 12 months, respectively, are: Atkins,  $n = 77, 73, 71,$  and  $68$ ; Zone,  $n = 79, 73, 67,$  and  $57$ ; LEARN,  $n = 79, 73, 66,$  and  $60$ ; and Ornish,  $n = 76, 72, 67,$  and  $56$ .

†By analysis of variance.

<sup>a,b,c,d</sup>When the analysis of variance (last column) was statistically significant ( $P < .05$ ), all pairwise comparisons among diet groups were tested for statistical significance using the Tukey studentized range test. Pairwise comparisons that were significantly different from one another are indicated by superscripts as follows: when the values for 2 diet groups within a row do not share a common superscript, they are significantly different, whereas if the values do share a common superscript, they are not significantly different.

\* Tableau modifié pour la présentation

## Analyses

ANOVA + Comparaisons multiples -> Est-ce que l'apport énergétique et le % de glucides diffèrent selon la diète (avant, puis après la diète)?



## Série d'analyses 4

Pour tester l'association entre des variables continues

Tests	Variable dépendante	Variable (s) indépendante(s)	Type de données	Type d'analyse
Corrélation	Continue	Continue	Non corrélées	Bivariée
Régression linéaire	Continue	Continue(s)	Non corrélées	Bivariée ou multivariée

Pour étudier le risque de survenue d'un événement

Tests	Variables dépendantes	Variable indépendante	Type de données	Type d'analyse
Kaplan-Meier	Continue (temps) + Dichotomique	Aucune ou Qualitative	Non corrélées	Univariée ou Bivariée



# Corrélation

Variables: Vitamine D sérique & ingérée, Temps passé à l'intérieur

Questions:

- Est-ce qu'il y a une corrélation entre la vitamine D ingérée et sérique?
- Est-ce que la vitamine D sérique diminue lorsque le temps passé à l'intérieur augmente?

Interprétation

- Coefficient de corrélation:  $\rho$  (rho)
  - Valeur de -1 à 1:
    - Près de 0, relation faible
    - Près de 1 ou -1, relation forte
  - Négatif: relation inverse
  - Positif: relation positive
- $P < 0,05$ : La corrélation n'est pas nulle



# Régression linéaire

Variables: Poids, taille, âge, VO2Max, cholestérol, tension

Questions:

- Peut-on prédire le poids d'un individu à partir de sa taille?
- (Idem), en ajustant pour son âge?
- Quelles variables permettent de prédire le cholestérol LDL?
- Quelle proportion de la variabilité du LDL est expliquée par ces variables?

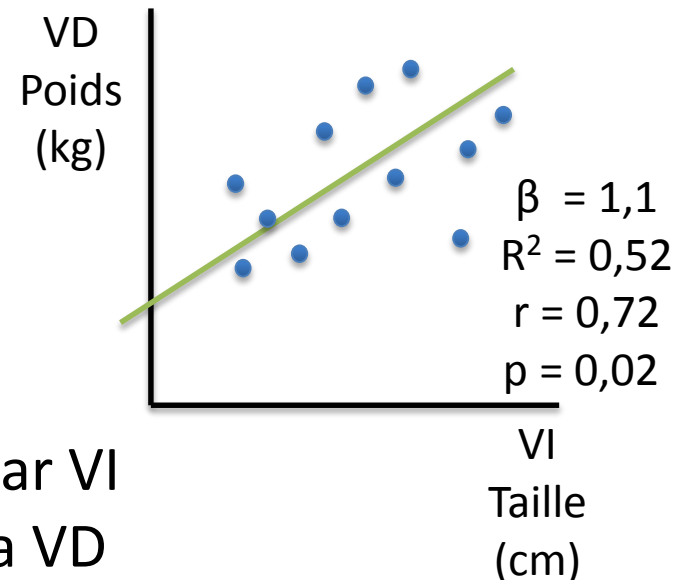
-> Sens à la relation (vs corrélation)



# Régression linéaire

## Interprétation

- Coefficient de régression:  $\beta$ 
  - Pente de la droite
  - Signe = sens de la relation
  - Lorsque VI augmente de 1 unité, VD augmente de  $\beta$  unités
- Coefficient de détermination:  $R^2$ 
  - % de variabilité de VD expliqué par VI
  - 0 = La VI n'explique pas du tout la VD
  - 1 = La VI explique parfaitement la VD
- Valeur P
  - $P < 0,05$ : Il y a une relation linéaire entre VI et VD



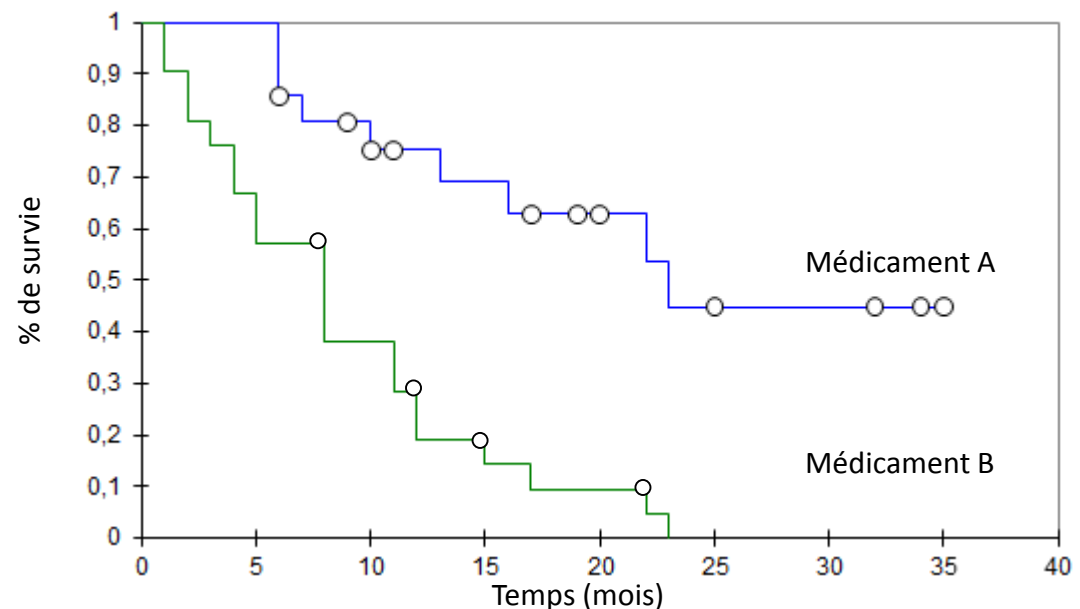


# Courbe de Kaplan-Meier

Variables: Décès, Temps de suivi (mois), Médicament (A ou B)

## Questions:

- Suite à une greffe, combien de patients décèdent et combien de temps s'écoule-t-il avant leur décès?
- Est-ce que la courbe de survie est la même selon la médication reçue après la greffe?





# Exemple: Étude de Cohorte

*Dog exposure in infancy decreases the subsequent risk of frequent wheeze but not of atopy*

## Participants

- 1246 enfants enrôlés après leur naissance et suivis jusqu'à 13 ans
- Classés selon le nombre d'animaux à domicile (exposition)

## Mesures

- Allergies (Immunoglobuline E, Tests cutanés), asthme
- Antécédants médicaux, habitudes de vie

## Hypothèse

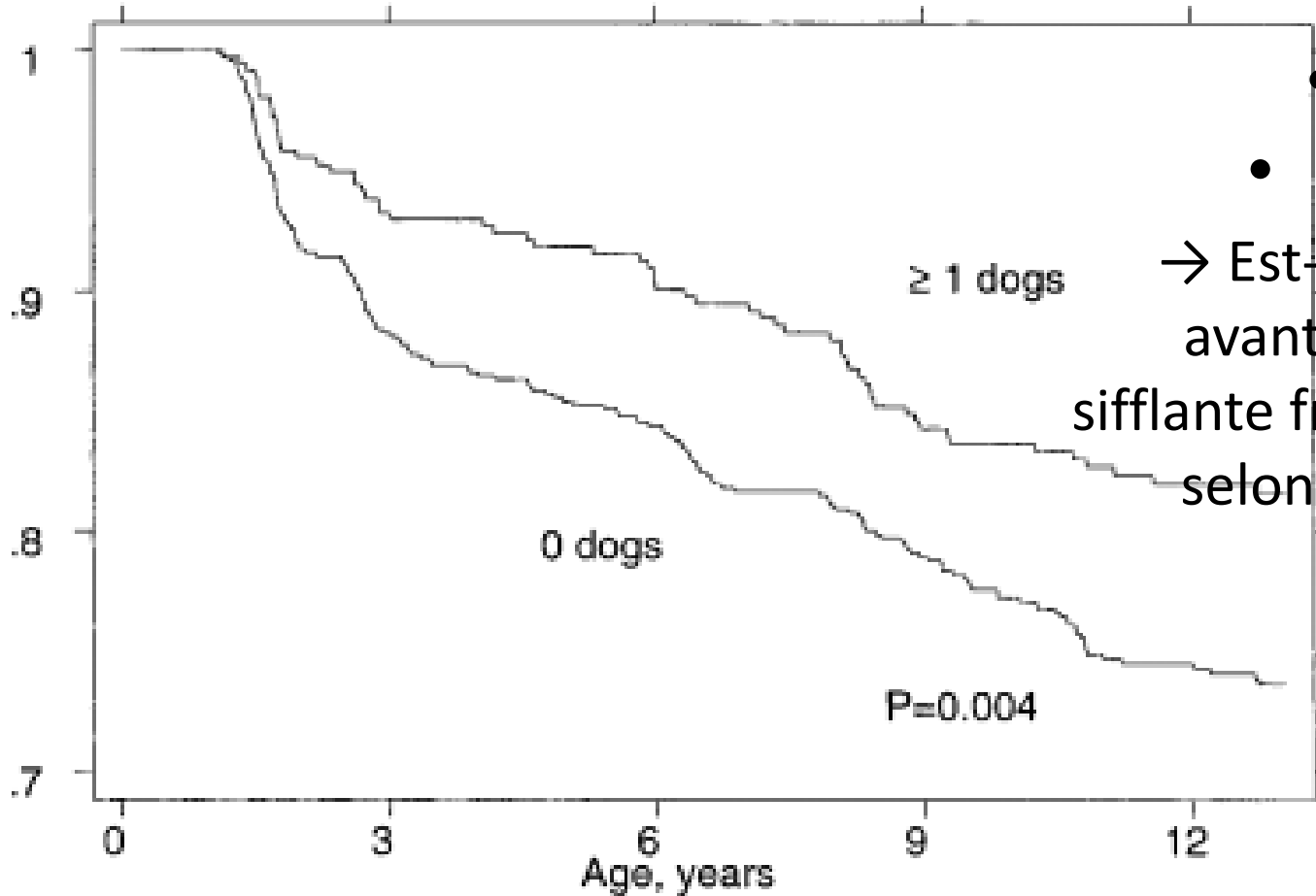
- L'exposition en bas-âge aux animaux prévient le développement d'allergies et de respiration sifflante fréquente





## Analyses

- Kaplan-Meier
  - Test du log-rank
- Est-ce que le temps avant une respiration sifflante fréquente diffère selon l'exposition aux chiens?



**FIG 1.** Kaplan-Meier survival estimates for frequent wheezing, by indoor dog exposure in infancy. P value is unadjusted (log-rank test).



# Références

- Le bilan démographique du Québec 2013, Institut de la statistique du Québec, Page 65.
- Nagele, P. (2001). Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *British Journal of Anaesthesia*, 90 (4): 514-16.
- Watz, H. et al. (2008). Extrapulmonary Effects of Chronic Obstructive Pulmonary Disease on Physical Activity, A cross-sectional Study. *American Journal of Respiratory and Critical Care Medicine*, 177: 743-751.
- Remes, S.T. (2001). Dog exposure in infancy decreases the subsequent risk of frequent wheeze but not of atopy. *Journal of Allergy and Clinical Immunology*, 108(4): 509-15.
- D'Souza, G., et al. (2007). Case-Control Study of Human Papillomavirus and Oropharyngeal Cancer. *New England Journal of Medicine*, 356: 1944-56.
- Gardner, C.D. et al. (2007) Comparison of the Atkins, Zone, Ornish, and LEARN Diets for Change in Weight and Related Risk Factors Among Overweight Premenopausal Women. *JAMA*, 297(9): 969-78.
- Altman, D.G. et al. (2001). The *Revised CONSORT Statement for Reporting Randomized Trials*: Explanation and Elaboration. *Annals of Internal Medicine*, 134(8): 663-694.
- Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. *Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration*. *Epidemiology*. 2007 Nov;18(6):805-35. PMID: 18049195
- Greenland, S., Senn, S.J., Rothman, K.J. et al. *Statistical tests, Pvalues, confidence intervals, and power: a guide to misinterpretations*. *European Journal of Epidemiology* (2016) 31: 337.
- Rosner, B. (2015). *Fundamentals of Biostatistics*, 8<sup>th</sup> edition. Brook Cole, 888 pages.
- Bernard, P.M. et C. Lapointe (1987), *Mesures statistiques en épidémiologie*, Sillery, Presses de l'Université du Québec.