



UMF
UNIVERSITATEA DE
MEDICINĂ ȘI FARMACIE
"IULIU HAȚIEGANU"
CLUJ-NAPOCA

Etudes descriptives. Modélisation des phénomènes médicaux

1

Objectifs

- Recherche descriptive
- Analyse de Corrélation
- Analyse de Régression:
 - Régression linéaire simple
 - Régression linéaire multiple
 - Régression logistique simple
 - Régression logistique multiple

2

2

Recherche descriptive

- Les études descriptives sont des études observationnelles qui décrivent:
 - les caractéristiques d'un phénomène,
 - les caractéristiques d'un sujet (patient)
 - les caractéristiques d'un d'une population étudiée;
- est utilisé lorsqu'il y a peu d'informations sur phénomène
- une recherche préliminaire;

3

3

Type de recherche

Type de recherche		Question	Direction de recherche
Descriptive		Que se passe t-il ? Nombre de cas?	description des certains phénomènes de santé (ex. une nouvelle maladie infectieuse)
Analytique	Explicative	Quelles sont les facteurs qui expliquent le phénomène de santé/résultat clinique d'intérêt ... ?	Construire une théorie (modèle) en testant des hypothèses.
	Prédictive	Que se passera t-il si... ?	Tester et vérifier des hypothèses.

- les études descriptives essaient juste de comprendre le phénomène étudié sans chercher à expliquer pourquoi ce phénomène se produit
- Les études analytiques font des comparaisons, essayant d'expliquer le phénomène

4

4

Types d'études (recherches) descriptives

- Étude de cas
- Série de cas
- Études transversales (études de prévalence)
- Etudes écologiques et corrélationnelles

5

5

Étude de cas (Cas rapporté)

- décrit l'**expérience d'un seul patient** ou d'un groupe de patients avec un diagnostic similaire.
- décrit comment un clinicien identifie une **caractéristique** état **inhabituel** d'une **maladie** ou des **antécédents** d'un patient.
- peuvent être les **premiers indices** pour **identifier de nouvelles maladies** ou **effets néfastes d'une exposition**.

Étape suivante : série de cas

6

6

Exemple

Case Reports > Case Rep Med. 2016;2016:5750710. doi: 10.1155/2016/5750710.
Epub 2016 Dec 26.

Fetus in the Bladder: Rare Complication of Vesicouterine Fistula

Vladimir Lesovoy¹, Yuri Parashchuk², Dmytro Shchukin¹, Roman Safonov³, Karyna Didenkova⁴,
Maria Lisova², Inessa Safonova⁵

FULL TEXT LINKS



ACTIONS



Abstract

The work presents a rare case of spontaneous migration of an 11-week fetus from the uterine cavity into the urinary bladder cavity through the long-standing vesicouterine fistula.

Conclusions

The onset of pregnancy is possible in the presence of the long-term vesicouterine fistula.

The migration of the fetus from the uterine cavity into the bladder presents one of the complications of vesicouterine fistulas.

Lien vers l'article: <https://pubmed.ncbi.nlm.nih.gov/28105051/>

7

7

Série de cas

- groupe de patients semblables entre eux (**avec une caractéristique commune d'intérêt**) = les « cas »
- il **n'y a pas de groupe témoin** ou une autre groupe de patients impliqué dans l'étude.
- Revue **d'observations cliniques** de cas individuels survenant dans une période relativement courte de temps.

8

8

Quand il est d'intérêt d'utiliser les séries de cas?

- on observe un **phénomène inhabituel** (une nouvelle maladie)
- peu de connaissances sur un domaine médical (**nouvelle pratique**)
- on veut des données/preuves « scientifiques » (en regroupant des observations cliniques similaires) pour formaliser un phénomène médical « intuitif »

9

9

Série de cas : Avantages et désavantages

- Avantages :
 - ✓ rapide, peu coûteux,
- Désavantages:
 - ✓ Pas de groupe de comparaison
 - **Pas** de mise en évidence d'un lien de **causalité/lien** de dépendance
 - Impossible de savoir si les facteurs observés sont plus fréquents parmi les cas que dans la population (**ne peut pas** être utilisé pour tester **l'association**)
 - ✓ Proviens d'observations cliniques : ne représente pas la population à risque
 - ✓ Ne représente pas l'ampleur/la magnitude du phénomène

Étape suivante : étude de prévalence

10

10

Exemple

> Nature. 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7. Epub 2020 Feb 3.

A pneumonia outbreak associated with a new coronavirus of probable bat origin

Peng Zhou ^{#1}, Xing-Lou Yang ^{#1}, Xian-Guang Wang ^{#2}, Ben Hu ¹, Lei Zhang ¹, Wei Zhang ¹, Hao-Rui Si ^{1 3}, Yan Zhu ¹, Bei Li ¹, Chao-Lin Huang ², Hui-Dong Chen ², Jing Chen ^{1 3}, Yun Luo ^{1 3}, Hua Guo ^{1 3}, Ren-Di Jiang ^{1 3}, Mei-Qin Liu ^{1 3}, Ying Chen ^{1 3}, Xu-Rui Shen ^{1 3}, Xi Wang ^{1 3}, Xiao-Shuang Zheng ^{1 3}, Kai Zhao ^{1 3}, Quan-Jiao Chen ¹, Fei Deng ¹, Lin-Lin Liu ⁴, Bing Yan ¹, Fa-Xian Zhan ⁴, Yan-Yi Wang ¹, Geng-Fu Xiao ¹, Zheng-Li Shi ⁵

FULL TEXT LINKS

 nature publishing group

 PMC Full text

ACTIONS

 Cite

 Favorites

• COVID-19 - Premier rapport d'une série de patients

Abstract

Since the outbreak of severe acute respiratory syndrome (SARS) 18 years ago, a large number of SARS-related coronaviruses (SARSr-CoVs) have been discovered in their natural reservoir host, bats¹⁻⁴. Previous studies have shown that some bat SARSr-CoVs have the potential to infect humans⁵⁻⁷. Here we report the identification and characterization of a new coronavirus (2019-nCoV), which caused an epidemic of acute respiratory syndrome in humans in Wuhan, China. The epidemic, which started on 12 December 2019, had caused 2,794 laboratory-confirmed infections including 80 deaths by 26 January 2020. Full-length genome sequences were obtained from five patients at an early stage of the outbreak. The sequences are almost identical and share 79.6% sequence identity to SARS-CoV. Furthermore, we show that 2019-nCoV is 96% identical at the whole-genome level to a bat coronavirus. Pairwise protein sequence analysis of seven conserved non-structural proteins domains show that this virus belongs to the species of SARSr-CoV. In addition, 2019-nCoV virus isolated from the bronchoalveolar lavage fluid of a critically ill patient could be neutralized by sera from several patients. Notably, we confirmed that 2019-nCoV uses the same cell entry receptor-angiotensin converting enzyme II (ACE2)-as SARS-CoV.

11

11

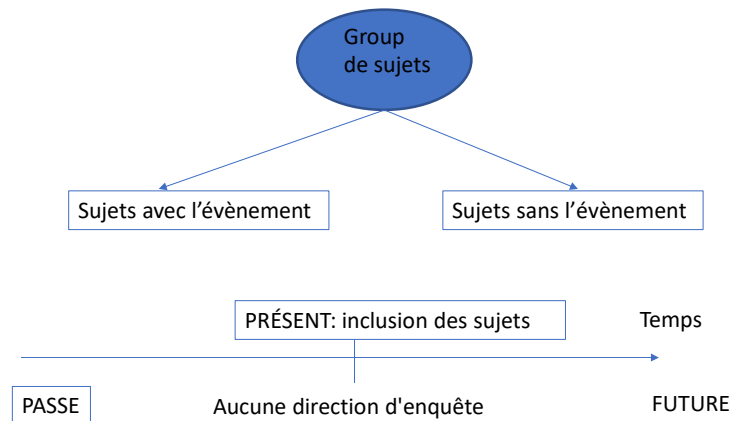
Étude transversale (étude de prévalence): définition

- engl. *Cross-sectional study*
- Mesure la **fréquence** d'un phénomène de sante:
 - maladie
 - facteur d'exposition
 - exposition à un traitement
- Enquête sur les sujets/patients présents au moment du recueil
- Une seule observation par sujet/patient = « photographie » de l'état de la population étudiée

12

12

Schéma d'une étude transversale: Que se passe-t-il en ce moment?



13

13

Exemple

> Acta Ophthalmol. 2020 Dec;98(8):e951-e959. doi: 10.1111/aos.14472. Epub 2020 May 18.

Ocular manifestations and clinical characteristics of 535 cases of COVID-19 in Wuhan, China: a cross-sectional study

Liwen Chen ¹, Chaohua Deng ¹, Xuhui Chen ¹, Xian Zhang ¹, Bo Chen ¹, Huimin Yu ¹, Yuanjun Qin ¹, Ke Xiao ¹, Hong Zhang ¹, Xufang Sun ¹

FULL TEXT LINKS



ACTIONS

“ Cite

- **Abstract**
- **Purpose:** To investigate the ocular manifestations and clinical characteristics of COVID-19 patients caused by SARS-CoV-2 in Wuhan, China.
- **Methods:** A total of 535 COVID-19 patients were recruited at Mobile Cabin Hospital and Tongji Hospital. Information on demographic characteristics, exposure history, ocular symptoms, eye drop medication, eye protections, chronic eye diseases, systemic concomitant symptoms, radiologic findings and SARS-CoV-2 detection in nasopharyngeal swabs by real-time PCR was collected from questionnaires and electronic medical records.
- **Results:** Of 535 patients, 27 patients (5.0%) presented with conjunctival congestion and 4 patients had conjunctival congestion as the initial symptom. The average duration of conjunctival congestion was 5.9 ± 4.5 days (mean [SD]). The other ocular symptoms, including increased conjunctival secretion, ocular pain, photophobia, dry eye and tearing, were also found in patients with conjunctival congestion. Notably, hand-eye contact was independently correlated with conjunctival congestion in COVID-19 patients. We also found that some COVID-19 patients had chronic eye diseases, including conjunctivitis (33, 6.2%), xerophthalmia (24, 4.5%) and keratitis (14, 2.6%). Similar to the published studies, the most common clinical symptoms were fever, cough and fatigue. A total of 343 patients (64.1%) had positive SARS-CoV-2 detection in nasopharyngeal swabs.
- **Conclusions:** Conjunctival congestion is one of the COVID-19-related ocular symptoms, which could occur as the initial symptoms. Frequent hand-eye contact may be the risk factor for conjunctival congestion in COVID-19 patients. Screening of patients with conjunctival congestion by ophthalmologists is advocated during the outbreak of COVID-19. It is essential to provide eye-care equipment and strengthen education on eye protection.

Lien vers l'article: <https://pubmed.ncbi.nlm.nih.gov/32421258/>

14

14

Etude transversale par questionnaire

- utile pour **comprendre comment les gens pensent**;
- on peut mesurer:
 - attitudes,
 - des avis,
 - comportements
- une série de **questions liées au même construction / aspect d' intérêt**;
- couvrant différents aspects;
- les réponses **peuvent être combinées** pour former un **score** ou une **échelle**;
- les **échelles** peuvent **mesurer la gravité de la maladie** ou elles peuvent être utilisées comme des **tests de diagnostic**

15

15

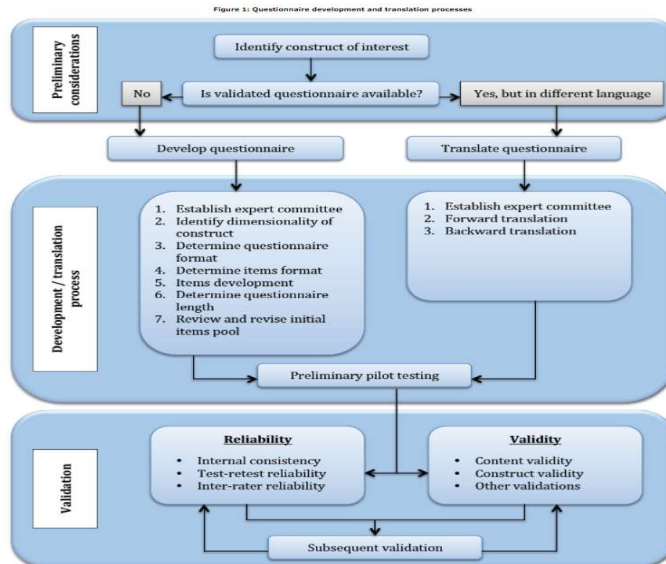
Etude transversale par questionnaire

- la qualité d'un questionnaire:
 - Analyse de fiabilité= précision de mesure et la reproductibilité des résultats
 - analyse de **cohérence interne**= pour déterminer dans quelle mesure les questions appartenant a la même échelle mesurent la même chose (les questions sont liés les uns aux autres)
 - analyse de **stabilité** (ex. **test-retest reliability**)= niveau de correspondance entre deux mesures distinctes effectuées par le même chercheur avec le même questionnaire/item
 - Analyse de validité = vérifier que le questionnaire mesure bien ce qu'il est censé évaluer
 - validité critérielle (*engl. Criterion validity*)
 - validité du construit (*engl. Construct validity*)
 - autres type de validité

16

16

Étapes de développement du questionnaire



Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi anesthesia. 2017;11(5): 80-89.

Lien vers l'article:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5463570/>

17

Exemple

International Journal of Mental Health and Addiction
https://doi.org/10.1007/s11469-020-00270-8

BRIEF REPORT

The Fear of COVID-19 Scale: Development and Initial Validation

Daniel Kwasi Ahorsu¹ · Chung-Ying Lin¹ · Vida Imani² · Mohsen Saffari³ · Mark D. Griffiths⁴ · Amir H. Pakpour^{5,6}

Published online: 27 March 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020



• Abstract

• **Background:** The emergence of the COVID-19 and its consequences has led to fears, worries, and anxiety among individuals worldwide. The present study developed the Fear of COVID-19 Scale (FCV-19S) to complement the clinical efforts in preventing the spread and treating of COVID-19 cases.

• **Methods:** The sample comprised 717 Iranian participants. The items of the FCV-19S were constructed based on extensive review of existing scales on fears, expert evaluations, and participant interviews. Several psychometric tests were conducted to ascertain its reliability and validity properties.

• **Results:** After panel review and corrected item-total correlation testing, seven items with acceptable corrected item-total correlation (0.47 to 0.56) were retained and further confirmed by significant and strong factor loadings (0.66 to 0.74). Also, other properties evaluated using both classical test theory and Rasch model were satisfactory on the seven-item scale. More specifically, reliability values such as internal consistency ($\alpha = .82$) and test-retest reliability (ICC = .72) were acceptable. Concurrent validity was supported by the Hospital Anxiety and Depression Scale (with depression, $r = 0.425$ and anxiety, $r = 0.511$) and the Perceived Vulnerability to Disease Scale (with perceived infectability, $r = 0.483$ and germ aversion, $r = 0.459$).

• **Conclusion:** The Fear of COVID-19 Scale, a seven-item scale, has robust psychometric properties. It is reliable and valid in assessing fear of COVID-19 among the general population and will also be useful in allaying COVID-19 fears among individuals.

Lien vers l'article: <https://pubmed.ncbi.nlm.nih.gov/32226353/>

18

18

Etudes corrélationnelles = écologiques

- **Les études écologiques corrélationnelles** = étude dans laquelle les facteurs étudiés proviennent de **population et non pas d'individu**, définie à partir de caractéristiques temporelles ou géographiques.
- utiliser des **données** déjà collectées et **agrégées au niveau de la (des) population (s)**.
- La mesure de l'association entre les variables observées peut être déterminé par le coefficient de corrélation de Pearson (r)

19

19

Etudes écologiques: Avantages et désavantages

- Avantages:
 - Rapide, peu onéreux.
 - De nouvelles hypothèses peuvent être générées
 - De nouveaux facteurs de risque peuvent être identifiés
- Désavantages:
 - Impossibilité de contrôler les facteurs de confusion.
 - Erreur écologique = erreur de raisonnement dans l'interprétation de résultats statistiques au niveau individuel à partir de données agrégées (conclusions fausses qui laissent penser que les relations entre des variables observées au niveau des groupes s'appliquent aussi à chaque individu)

20

20

Indicateurs calculés dans les études descriptives transversales

- **prévalence** = le rapport entre le nombre total de cas de la maladie à un moment donné (prévalence instantanée) ou sur une période de temps (prévalence périodique) et la taille de la population.
- **incidence** = le rapport entre le nombre de nouveaux cas de maladie sur une période donnée et la taille de la population.
- **taux de mortalité** = le rapport entre le nombre de décès enregistrés et la taille de la population sur une certaine période de temps.
- **taux de létalité** = le rapport entre le nombre de décès enregistrés et le nombre total de sujets malades sur une période donnée.

21

21

Exemple

> Nicotine Tob Res. 2020 Aug 24;22(9):1646-1649. doi: 10.1093/ntr/ntaa121.

Smoking Prevalence and COVID-19 in Europe

Panagiotis Tsigaris ¹, Jaime A Teixeira da Silva ²

FULL TEXT LINKS

OXFORD
ACADEMIC

PMC **FREE**
Full text

Abstract

- **Introduction:** This ecological study investigates the association between smoking prevalence and COVID-19 occurrence and mortality in 38 European nations as of May 30, 2020.
- **Methods:** Data were collected from Our World in Data. Regression analysis was conducted to adjust for potential confounding factors such as economic activity (gross domestic product), the rate of COVID-19 testing, and the stringency of COVID-19 control policies.
- **Results:** There was a statistically significant negative association between smoking prevalence and the prevalence of COVID-19 across the 38 European nations after controlling for confounding factors ($p = 0.001$). A strong association was found between the prevalence of COVID-19 per million people and economic activity ($p = 0.002$) and the rate of COVID-19 testing ($p = 0.0006$). Nations with stricter policy enactment showed fewer COVID-19 cases per million people, but the association was not significant ($p = 0.122$). Delaying policy enactment was associated with a greater prevalence of COVID-19 ($p = 0.0535$). Evidence of a direct association between smoking prevalence and COVID-19 mortality was not found ($p = 0.626$). There was a strong positive association between COVID-19 mortality rate and the prevalence of COVID-19 cases ($p < 0.0001$) as well as the proportion of the population over 65 years of age ($p = 0.0034$) and a negative association with the rate of COVID-19 testing ($p = 0.0023$).
- **Conclusions:** We found a negative association between smoking prevalence and COVID-19 occurrence at the population level in 38 European countries. This association may not imply a true or causal relationship, and smoking is not advocated as a prevention or treatment of COVID-19.

Lien vers l'article: <https://pubmed.ncbi.nlm.nih.gov/32609839/>

22

22

Modélisation des phénomènes de santé à l'aide de la corrélation et de la régression

23

23

Analyse de Corrélation versus Analyse de Régression

- **Analyse de Corrélation: mesurer la liaison entre deux variables quantitatives X et Y**
 - **corrélation est symétrique:** les deux variables peuvent être interchangées sans affecter leur corrélation
- **Analyse de Régression = modéliser la liaison entre 1 variable dépendante et ≥ 1 variables indépendantes**
 - !! la **régression est asymétrique** (si Y dépend de X \neq X dépend de Y)
 - !! X=variable indépendante (explicative); Y=var. dépendante (expliquée)
 - !! dépendance \neq causalité
- Dans une corrélation: il n'y a pas des variables dépendantes ou variables indépendantes

24

Analyse de corrélation

25

LES TYPES DE CORRELATIONS ENTRE DEUX VARIABLES QUANTITATIVES

- Une corrélation est linéaire:
 - **positive**: les deux variables varient dans le même sens (proportionnalité directe)
 - **négative**: les deux variables varient en sens inverse (proportionnalité inverse)
- Une corrélation est non-linéaire et monotone: les valeurs des variables ont tendance à se déplacer dans la même direction, mais pas forcément à une vitesse constante
 - **positive**: Si X augmente alors Y aussi
 - **négative**: Si X augmente alors Y diminue

26

26

Evaluation du lien/liaison entre 2 variables quantitatives:

- Le poids et la tension artérielle systolique?

Poids = X: X_1, X_2, \dots, X_n

TAS = Y: Y_1, Y_2, \dots, Y_n

o. L'évaluation graphique de la relation :

- La diagramme de dispersion/nuage des points/scatter plot

1. L'existence d'une relation entre les variables X et Y

- A l'aide d'un test statistique sur le coefficient de corrélation

2. Le sens et l'intensité / la force/l'importance/le degré de la corrélation

- Le coefficient de corrélation Pearson ou Spearman

3. Prédiction : prédire les valeurs d'une variable sachant les valeurs de l'autre

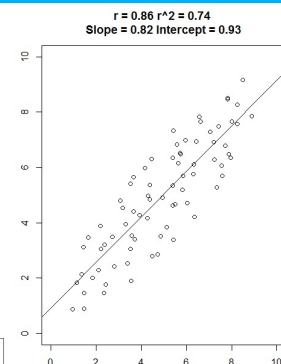
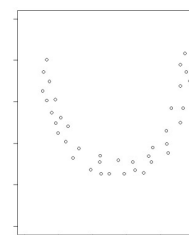
- La régression
 - déterminer une fonction (fonction de régression) telle que $Y=f(X)$

27

Evaluation graphique : diagramme de dispersion (nuage des points/scatter)

➤ 1^{er} objectif - Evaluer la linéarité

- ✓ Si le nuage des points semble suggérer qu'il y a une tendance que les points sont plutôt situés autour d'une droite imaginaire – la relation est peut être linéaire
- ✓ Si le nuage des points semble suggérer des tendances qui ne sont pas linéaires, la relation est peut être non linéaire (exponentielle, quadratique, ...)

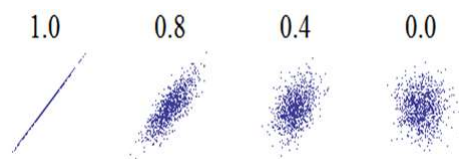


28

Evaluation graphique : diagramme de dispersion (nuage des points/scatter)

➤ 2eme objectif - Evaluer l'importance/la force/puissance de la corrélation:

- ✓ Si la relation est plus probable linéaire, on peut évaluer d'une manière subjective la force/puissance/importance de la corrélation linéaire.
- ✓ Plus les points se rapprochent a un droite de tendance-> plus la corrélation est forte (plus les points sont distants, plus la corrélation est faible)

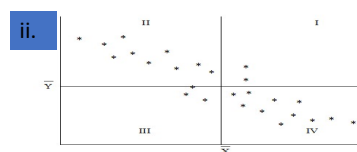
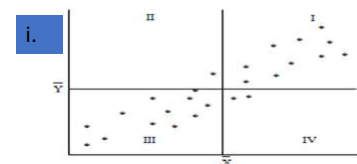


29

Evaluation graphique : diagramme de dispersion (nuage des points/scatter)

➤ 3eme objectif : évaluation visuelle de la relation entre deux variables quantitatives en utilisant des cadrans (créés par les moyennes du x et du y) pour identifier la tendance/sens/direction (relation de proportionnalité directe ou inverse):

- La plupart des points sont dans les cadrans I et III OU une sensation visuelle que les points ont une tendance croissante (plus les valeur du x augmente le y augmente aussi) \Rightarrow tendance croissante/ pente ascendante/ pente positive/ lien (direct) proportionnel
- La plupart des points sont dans les cadrans II et IV OU une sensation visuelle que les points ont une tendance décroissante (plus les valeur du x augmente le y diminuent) \Rightarrow tendance décroissante / pente descendante/ pente négative/ lien inversement proportionnelle



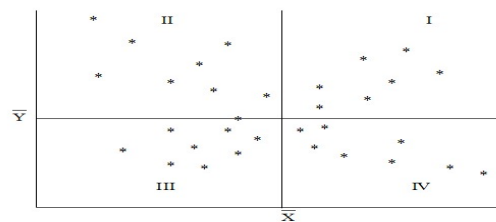
30

Evaluation graphique : diagramme de dispersion (nuage des points/scatter)

- **3eme objectif : évaluation visuelle** de la relation entre deux variables quantitatives en utilisant des cadrans (créés par les moyennes du x et du y) pour identifier **la tendance/sens/direction (relation de proportionnalité directe ou inverse)**:

- iii. les points sont distribués uniformément dans tous les cadrans ou les points semblent distribués au hasard partout \Rightarrow aucune tendance

iii.



31

Le coefficient de Corrélation linéaire de Pearson

But: évaluer la **relation linéaire** entre des deux variables **quantitatives** du point de vue de la **direction/sens/intensité** de la relation.

Le **coefficient de corrélation de Pearson**:

- est une mesure synthétique de **l'intensité et du sens** de la relation entre deux **variables quantitatives**
- montre le degré de rapprochement des points à une droite qui passe entre les points.

Condition d' application:

- ✓ les paires des observations indépendantes dans l'échantillon
- ✓ variables quantitatives
- ✓ les deux variables soient normalement distribuées
- ✓ la relation entre les deux variables est linéaire simple

32

Coefficient de corrélation linéaire de Pearson

Covariance d'échantillonnage COV (X,Y):

$$COV(X,Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

Coefficient de corrélation linéaire de Pearson

$$r = \frac{COV(X,Y)}{S_X \cdot S_Y}$$

$$r \in [-1,1]$$

X_i, Y_i – sont les valeurs des deux séries des données. \bar{X} et \bar{Y} sont les moyennes des deux séries. n – nombre des observations. S_x et S_y sont les déviations standard d'échantillonnage, r – coefficient de corrélation Pearson, d – coefficient de détermination

33

Corrélation linéaire - interprétations

r (coefficient de corrélation Pearson):

montre la **direction /sens** et l' **intensité** de la corrélation linéaire;

a. Interprétation du direction/sens/tendance:

$r > 0$ corrélation linéaire positive / relation linéaire positive/proportionnalité directe/tendance croissante

$r < 0$ corrélation linéaire négative / relation linéaire négative / proportionnalité inverse/tendance décroissante

$r \approx 0$ aucune corrélation linéaire

b. Interprétation de l'intensité de la relation/corrélation

Si $|r|$ approche **1** l'intensité de l'a corrélation augmente

Si $|r|$ se rapproche de **0** l'intensité de l' corrélation diminue

34

Corrélation linéaire - interprétations

Interprétation de l'intensité/force/degré/puissance de la corrélation linéaire avec les règles empiriques de Colton [Colton T. Statistics in Medicine. Little Brown and Company, New York, NY 1974]

r compris entre $-0,25$ et $0,25 \Rightarrow$ une corrélation linéaire **négligeable** ou **nulle**

$[0.25 \text{ et } 0.50)$ ou $[-0.25 \text{ et } -0.50) \Rightarrow$ un degré de corrélation linéaire **faible/acceptable**

$[0.50 \text{ et } 0.75)$ ou $[-0.50 \text{ et } -0.75) \Rightarrow$ un degré de corrélation linéaire **modérée à bonne**

$[0.75 \text{ et } 1]$ ou $[-0.75 \text{ et } -1] \Rightarrow$ une **très bonne à excellente** corrélation linéaire

!!! les parenthèses [...] indique que la valeur est incluse, et les parenthèses (...) indiquent que la valeur n'est pas incluse dans l'intervalle)

Il y a autre divisions possibles aussi.

Ces règles doit être utilisée avec soins.

Elle sont pour donner une idée, mais pour chaque problème, l'intensité de la relation est relative au domaine.

Pour certain situations les valeurs en dessous de 0,8 peut être faibles.

35

Test statistique de signifiante pour le coefficient de corrélation linéaire de Pearson (r)

- **But:** tester si deux variables ont une **corrélation linéaire** statistiquement significative
- On utilise un test statistique sur le coefficient de corrélation pour voir s'il est différent de 0
- **H_0 = absence de la corrélation linéaire au niveau de la population cible**
- **Condition d' application:**
 - les paires des observations indépendantes dans l'échantillon
 - variables quantitatives
 - les deux variables sont normalement distribuées
 - sans valeurs aberrantes (très éloignées du nuage des points)

36

Test statistique de signifiante pour le coefficient de corrélation linéaire de Pearson (r)

- L' hypothèse nulle (H₀):
 - il n'y a pas de différence statistiquement significative entre le coefficient de corrélation r (des variables X et Y) et 0
 - il n'y a pas de corrélation linéaire statistiquement significative entre les variables X et Y
- L' hypothèse alternative (H₁):
 - il y a une différence statistiquement significative entre le coefficient de corrélation r (des variables X et Y) et 0
 - il y a une corrélation linéaire statistiquement significative entre les variables X et Y
- La décision avec p-value.
 - Si p-value < α (=0,05) => on rejette H₀ et nous sommes en faveur du H₁
 - Si p-value $\geq \alpha$ (=0,05) => on ne peut pas rejeter H₀

37

Test statistique de signifiante pour le coefficient de corrélation linéaire du Pearson

Exemple

Le coefficient de corrélation linéaire Pearson pour la relation entre les triglycérides et le poids pour 50 sujets est 0,72, et la valeur du p associée est 0,001.

Les paires des observations sont indépendantes, les données sont normalement distribuées, la relation est linéaire

Interprétation de la valeur du p associée à le coefficient de corrélation:

=> p<0,05 il y a une corrélation linéaire statistiquement significative entre les triglycérides et le poids

La corrélation linéaire entre les triglycérides et le poids est statistiquement significative

Interprétation de la direction et de l'intensité de la corrélation: la relation est (directe) proportionnelle (r=0,72 >0), et l' intensité de la corrélation est modérée à bonne (r=0,72 – est dans [0,50 et 0,75))

38

Si les **conditions d' application** ne sont pas satisfaites

- Solutions possibles:
- transformations: dans une échelle logarithmique
- exclure des valeurs aberrants de l'analyse ou
- Le calcul du coefficient de Spearman

39

39

Coefficient de corrélation de Spearman

But: évaluer la relation monotone entre deux variables du point de vue de la **direction/sens et son intensité.**

Condition d' application:

- Les paires des observations indépendantes dans l'échantillon
- Les deux variables sont ordinales/
- Ou Une variable ordinale et une quantitative (normalement ou non normalement distribuées)
- Ou Deux variables quantitatives (au moins une non normalement distribuées)

Utilité: évaluer la relation entre

- deux variables quantitatives qui ne sont pas normalement distribuées
- deux variables quantitatives: une variable normale distribuée, et un autre non normalement distribuée
- deux variables ordinales
- une ordinale et une quantitative (n'importe s'il est normale distribuée, ou non)

40

Coefficient de corrélation de Spearman

Etapes de calcul:

- Remplacer la série bivariable $(x_1, \dots, x_n; y_1, \dots, y_n)$
 - avec la série des rangs $(R_{x1}, \dots, R_{xn}; R_{y1}, \dots, R_{yn})$, des valeurs x_i et y_i après leur rangement dans ordre croissant
 - pour les valeurs égaux on prend la moyenne arithmétique des rangs.

□ Calculer le coefficient ρ de Spearman:
$$\rho = \frac{\sum_{i=1}^n (R_{x,i} - \bar{R}_X)(R_{y,i} - \bar{R}_Y)}{\sqrt{\sum_{i=1}^n (R_{x,i} - \bar{R}_X)^2} \sqrt{\sum_{i=1}^n (R_{y,i} - \bar{R}_Y)^2}}$$

Interpretations:

a. Interprétation du direction/sens/tendance:

- $\rho > 0$ corrélation monotone positive / relation monotone positive/tendance croissante
- $\rho < 0$ corrélation monotone négative / relation monotone négative / tendance décroissante
- $\rho \approx 0$ aucune corrélation linéaire

b. Interprétation de l'intensité de la relation/corrélation

Si $|\rho|$ approche 1 l'intensité de la corrélation augmente

Si $|\rho|$ se rapproche de 0 l'intensité de la corrélation diminue

41

Coefficient de corrélation Spearman

Il y a aussi un test statistique pour vérifier si le coefficient de corrélation Spearman (ρ) est statistiquement différent de 0.

• L'hypothèse nulle:

- il n'y a pas d'une **différence statistiquement significative** entre le **coefficient de corrélation ρ** et 0
- il n'y a pas de **corrélation monotone statistiquement significative** entre les variables **X** et **Y**

• L'hypothèse alternative:

- il y a une **différence statistiquement significative** entre le **coefficient de corrélation ρ** et 0
- il y a une **corrélation monotone statistiquement significative** entre les variables **x** et **y**

• La décision avec p-value.

- Si $p\text{-value} < \alpha (=0,05) \Rightarrow$ on rejette H_0 et nous sommes en faveur de H_1
- Si $p\text{-value} \geq \alpha (=0,05) \Rightarrow$ on ne peut pas rejeter H_0

42

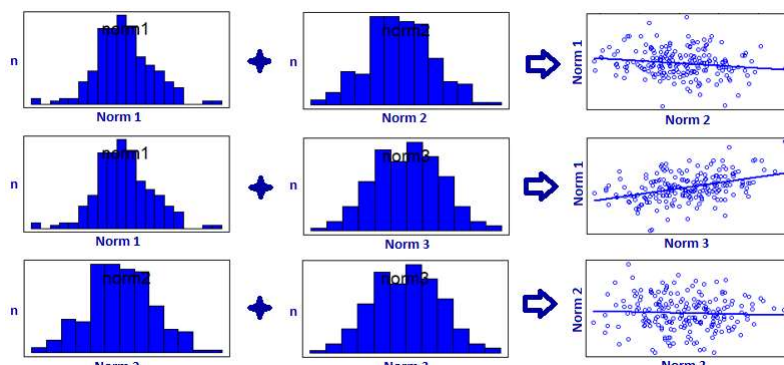
Comment identifier la distribution des données en regardant un graphique nouage des points, et d'identifier le bon coefficient de corrélation à utiliser

- Si on regarde une graphique de type nouage des points, on peut deviner approximative si les deux variables ont une distribution normale (gaussienne) ou pas. On sait que la distribution normale est symétrique autour de la moyenne, et on peut observer ça dans le graphique.
- Ex. Si les deux variables sont normale distribuées le graphique a approximative la forme d'une ellipse.

43

Comment identifier la distribution des données en regardant un graphique nouage des points, et d'identifier le bon coefficient de corrélation à utiliser

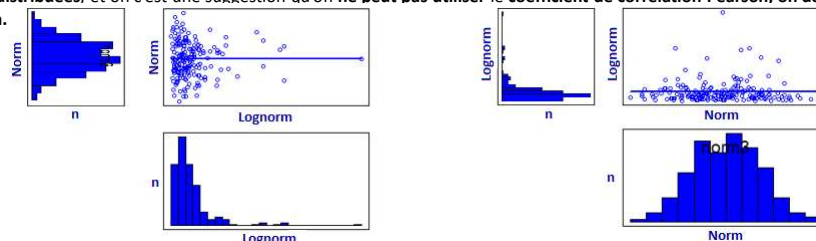
- Ex. Ici il y a **3 variables**: Norm 1, Norm 2, et Norm 3 qui sont **normale distribuées** (come on peut observer sur les histogrammes).
- Pour chaque combinaison entre deux variables Norm 1 + 2, Norm 1 + 3, et Norm 2 + 3 il y a une nouage des points.
- On peut observer que si **les deux variables sont normale distribuées** le **graphique a la forme d'une ellipse** (n'importe si la tendance est décroissante, croissante, ou sans).
- Donc si vous observez des nouages des points en forme d'ellipse, et la relation est approximative lineaire vous pouvez deviner que c'est une **suggestion** que les **deux variables** sont **normale distribuées**, et on peut utiliser le **coefficient de corrélation Pearson**.



44

Comment identifier la distribution des données en regardant un graphique nuage des points, et d'identifier le bon coefficient de corrélation à utiliser

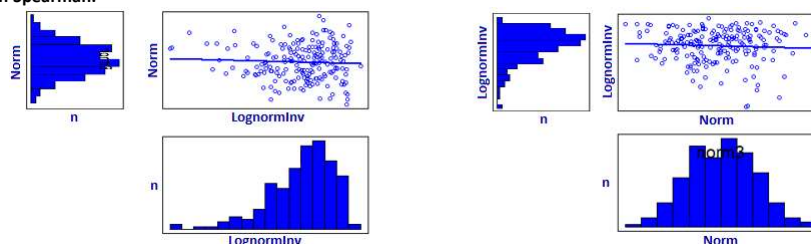
- Ex. Ici il y a **2 variables**: la variable **Norm** qui est **normale distribuée** (come on peut observer sur l'histogramme – symétrique, forme gaussienne – de chapeau), et la variable **Lognorm** – qui **n'est pas normale distribuée**, ayant une forte **asymétrie à droite** (come on peut observer sur l'histogramme – n'est pas symétrique, il y a une **queue à la droite**).
- On peut observer deux nuages des points: sur axe X – Lognorm, sur axe Y – Norm
- sur axe X – Norm, sur axe Y – Lognorm
- On peut observer que si **une des deux variables ne sont pas normale distribuées le graphique n'a pas la forme d'une ellipse**. On observe a dans le nuage des points a gauche que pour la variable Norm les données sont approximative symétrique distribuées autour de centre du nuage sur l'axe verticale, mais la variable Lognorm a la plupart des points a gauche, et de mois en moins des points, de plus en plus distancées vers la droit du nuage des points. Le graphique nuage des points a la droite est identique, mais avec une rotation, parce que les mêmes variables sont inversées sur les axes.
- Donc si vous observez des nuages des points qui **n'est pas en forme d'ellipse** vous pouvez deviner que c'est une **suggestion** que au moins une des deux **variables ne sont pas normale distribuées**, et on c'est une suggestion qu'on **ne peut pas utiliser le coefficient de corrélation Pearson, on doit utiliser plutôt le coefficient de corrélation Spearman**.



45

Comment identifier la distribution des données en regardant un graphique nuage des points, et d'identifier le bon coefficient de corrélation à utiliser

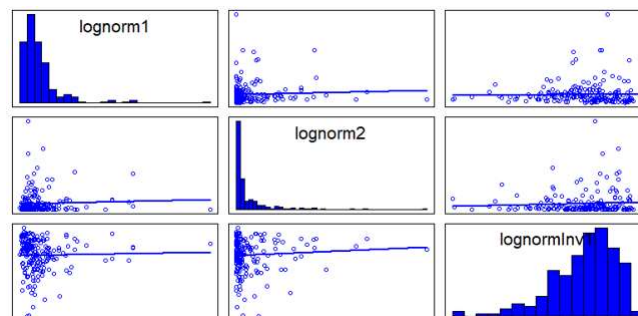
- Ex. Ici il y a **2 variables** **Norm** qui est **normale distribuée** (come on peut observer sur l'histogramme – symétrique, forme gaussienne – de chapeau), et la variable **LognormInv** – qui **n'est pas normale distribuée**, ayant une forte **asymétrie à gauche** (come on peut observer sur l'histogramme – n'est pas symétrique, il y a une **queue à gauche**).
- On peut observer deux nuages des points: sur axe X – LognormInv, sur axe Y – Norm
- sur axe X – Norm, sur axe Y – LognormInv
- On peut observer que si **une des deux variables ne sont pas normale distribuées le graphique n'a pas la forme d'une ellipse**. On observe a dans le nuage des points a gauche que pour la variable Norm les données sont approximative symétrique distribuées autour de centre du nuage sur l'axe verticale, mais la variable LognormInv a la plupart des points a droite, et de mois en moins des points, de plus en plus distancées vers la gauche du nuage des points. Le graphique nuage des points a la droite est identique, mais avec une rotation, parce que les mêmes variables sont inversées sur les axes.
- Donc si vous observez des nuages des points qui **n'est pas en forme d'ellipse** vous pouvez deviner que c'est une **suggestion** que au moins une des deux **variables ne sont pas normale distribuées**, et on c'est une suggestion qu'on **ne peut pas utiliser le coefficient de corrélation Pearson, on doit utiliser plutôt le coefficient de corrélation Spearman**.



46

Comment identifier la distribution des données en regardant un graphique nouage des points, et d'identifier le bon coefficient de corrélation à utiliser

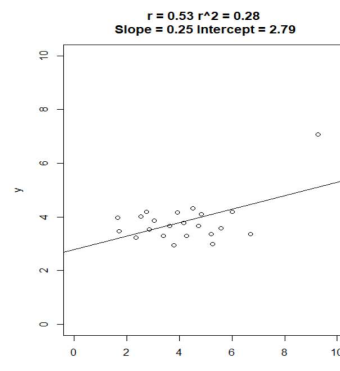
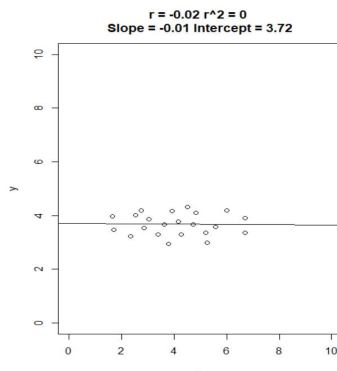
- Ex. Ici il y a **3 variables**: les variables **Lognorm 1, et 2** – qui **ne sont pas normale distribuées**, ayant une forte **asymétrie a droite** (come on peut observer sur l'histogramme – n'est pas symétrique, il y a une **queue a la droite**), et la variable **LognormInv** – qui **n'est pas normale distribuée**, ayant une forte **asymétrie a gauche** (come on peut observer sur l'histogramme – n'est pas symétrique, il y a une **queue a gauche**).
- On peut observer une **matrice des nouages des points**, et sur la **diagonale** sont les **histogrammes des variables**.
- On peut observer que si **les deux variables ne sont pas normale distribuées** le graphique **n'a pas la forme d'une ellipse**. Donc si vous observez des **nouages des points qui n'est pas en forme d'ellipse** vous pouvez deviner que c'est une **suggestion** que au moins une des deux **variables ne sont pas normale distribuées**, et on c'est une suggestion qu'on **ne peut pas utiliser** le **coefficient de corrélation Pearson**, on doit utiliser plutôt le **coefficient de corrélation Spearman**.



47

L'influence des valeurs aberrantes sur le coefficient de corrélation Pearson, et sur la droite de régression

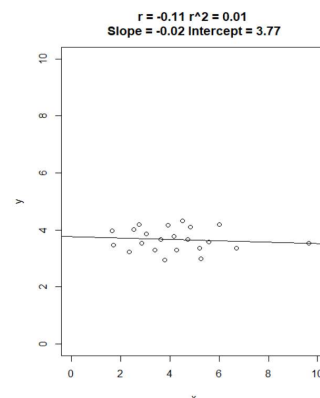
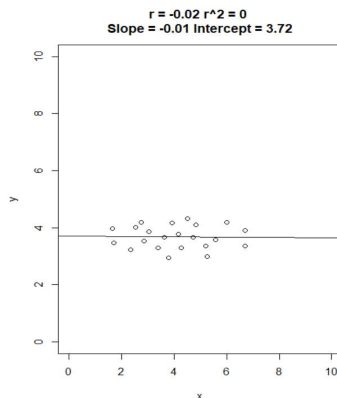
- Ex. Dans le **graphique a gauche** on peut observer une nouage des points en forme d'ellipse, avec une relation linéaire. Pour ce graphique le **coefficient de corrélation Pearson** est **-0,02**, et la **pente** de la droite de régression est **-0,01**. Donc on peut observer qu'il **n'y a pas de relation** apparente entre les variables.
- Dans le **graphique a droite** on peut observer une le même nouage des points, mais avec une valeur aberrante a la droite et en haut. Pour ce graphique le **coefficient de corrélation Pearson** est **0,53**, et la **pente** de la droite de régression est **0,25**. Si on regarde seulement le coefficient et la pente on peut penser que il y a une corrélation, une relation entre les deux variables, mais cette relation est induite d'une manière fausse par la valeur aberrante (qui est aussi un point de levier). S'il vous plait de **ne pas faire confiance dans les statistiques dans ce gens de situations**. Les auteurs de l'article doivent évaluer la cause de ces valeurs, et aussi l'effet de la ou les valeurs aberrantes. Au lieu du coefficient de corrélation Pearson, ici sera mieux d'utiliser le coefficient de corrélation Spearman.



48

L'influence des valeurs aberrantes sur le coefficient de corrélation Pearson, et sur la droite de régression

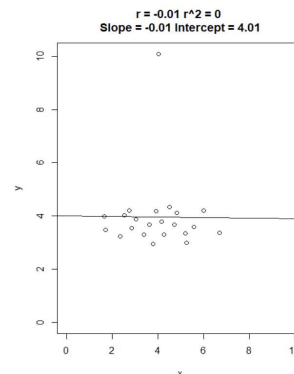
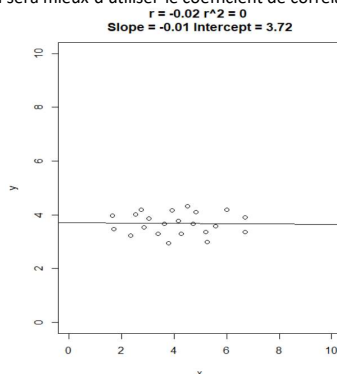
- Ex. Dans le **graphique a gauche** on peut observer une nouage des points en forme d'ellipse, avec une relation linéaire. Pour ce graphique le **coefficient de corrélation Pearson** est **-0,02**, et la **pente** de la droite de régression est **-0,01**. Donc on peut observer qu'il **n'y a pas de relation** apparente entre les variables.
- Dans le **graphique a droite** on peut observer une le même nouage des points, mais avec une valeur aberrante a la droite. Pour ce graphique le **coefficient de corrélation Pearson** est **-0,11**, et la **pente** de la droite de régression est **-0,02**. Ici la valeur aberrante n'a pas influencée beaucoup les statistiques (elle n'a pas une effet de levier). S'il vous plait de **ne pas faire confiance dans les statistiques dans ce gens de situations**. Les auteurs de l'article doivent évaluer la cause de ces valeurs, et aussi l'effet de la ou les valeurs aberrantes. Au lieu du coefficient de corrélation Pearson, ici sera mieux d'utiliser le coefficient de corrélation Spearman.



49

L'influence des valeurs aberrantes sur le coefficient de corrélation Pearson, et sur la droite de régression

- Ex. Dans le **graphique a gauche** on peut observer une nouage des points en forme d'ellipse, avec une relation linéaire. Pour ce graphique le **coefficient de corrélation Pearson** est **-0,02**, et la **pente** de la droite de régression est **-0,01**, et le **coefficient libre** est **3,72**. Donc on peut observer qu'il **n'y a pas de relation** apparente entre les variables.
- Dans le **graphique a droite** on peut observer une le même nouage des points, mais avec une valeur aberrante au milieu, et en haut. Pour ce graphique le **coefficient de corrélation Pearson** est **-0,01**, et la **pente** de la droite de régression est **-0,01**, et le **coefficient libre** est **4,01**. Ici la valeur aberrante n'a pas influencée beaucoup les statistiques (elle n'a pas une effet de levier), sauf le coefficient libre. S'il vous plait de **ne pas faire confiance dans les statistiques dans ce gens de situations**. Les auteurs de l'article doivent évaluer la cause de ces valeurs, et aussi l'effet de la ou les valeurs aberrantes. Au lieu du coefficient de corrélation Pearson, ici sera mieux d'utiliser le coefficient de corrélation Spearman.



50

Analyse de Regression

51

51

Régression

- **But:** - méthode pour étudier liaison entre 2 ou plusieurs variables

- **Forme générale:**

$Y = f(X) + \varepsilon$ ou f = fonction de régression

X = variable indépendante, explicative, prédictive

Y = variable dépendante, expliquée, prédite

ε = écart (erreur) de cette approximation

52

Type de Régression

- le type du variable dépendante
 - variable quantitative – régression linéaire
 - variable dichotomique – régression logistique
 - variable de survie – régression Cox
- la linéarité de la fonction
 - régression linéaire
 - régression non linéaire
- le nombre de variables dépendantes:
 - régression univariée (une variable dépendante)
 - régression multivariée (≥ 2 variables dépendantes)
- le nombre de variables indépendantes:
 - régression **simple** (une variable indépendante)
 - régression **multiple** (≥ 2 variables indépendantes)

53

Régression linéaire simple

But étudier la liaison linéaire entre deux variables, depuis lesquels la variable dépendante est quantitative

Objectifs:

- **prédire les valeurs** d' une variable quantitative en fonction des valeurs de l'autre variable
- évaluer
 - l'**existence** d'une liaison/lien/relation entre les deux variables
 - la **direction** de la liaison/lien/relation entre deux variables
 - l' **importance** de la liaison/lien/relation entre deux variables

Conditions d'application:

- variable dépendante quantitative
- observations indépendants
- existence d'une relation linéaire entre les deux variables (X et Y)
- erreurs (résidus=écart entre chaque valeur observée et chaque valeur prédite) normalement distribués & homoscédasticité (variance constante) des résidus.

54

Régression linéaire simple – le model

$Y = b_0 + b_1 X + \varepsilon$ ou b_0, b_1 = coefficients de la régression

- Méthode de calcul des coefficients de régression partielle (b_0, b_1)
= méthode des moindres carrés
- critère de la méthode: minimisé la somme de tous les carrés de la distance du chaque point par rapport à la droite:

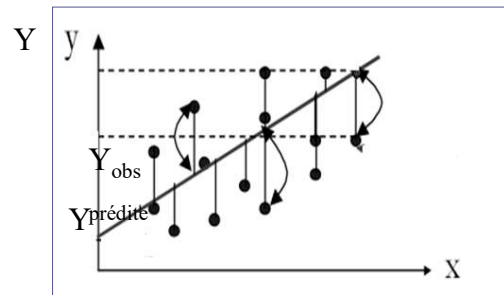
$$\sum_{i=1}^n (Y_i^{\text{predite}} - Y_i)^2 \rightarrow Y_i^{\text{predite}} = b_0 + b_1 X_i$$

-on obtient: $b_1 = \frac{\text{COV}(X,Y)}{S_X^2}$

Annotations:
 - Valeur observée (pointing to Y_i)
 - covariance (pointing to $\text{COV}(X,Y)$)

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Moyenne du Y, moyenne du X



55

Régression linéaire simple

- **Interprétation**

- La droite de régression $Y(X)$: $Y = b_0 + b_1 X$

b_0 = est l'ordonnée à l'origine – la valeur du Y quand X est égal a 0 (d'habitude cette information n'est pas utile pour les médecins, elle présente une situation qui en réalité est impossible)

b_1 = la pente de la droite de régression.

Interprétation de b_1 = coefficient de la variable X

quand la variable indépendante (X) augmente d'une unité, nous pouvons nous attendre à ce que la valeur de la variable dépendante Y augmente (si $b_1 > 0$) ou diminue (si $b_1 < 0$) en moyenne.

- plus b_1 est grand (en valeur absolue) plus la relation est forte
- plus b_1 est proche du 0, plus la relation est faible

56

Test statistique de signifiante pour le coefficient de de la pente (b₁) de la régression linéaire

- L' hypothèse nulle:
 - il n'y a pas de différence statistiquement significative entre le coefficient (b₁) de de la variable X et 0
 - le coefficient (b₁) de la variable X n'est pas statistiquement significative
- L' hypothèse alternative:
 - il y a une différence statistiquement significative entre le coefficient (b₁) de de la variable X et 0
 - le coefficient (b₁) de la variable X est statistiquement significative
- La décision avec p-value.
 - Si p-value < α (=0,05) => on rejete H₀ et nous sommes en faveur du H₁
 - Si p-value ≥ α (=0,05) => on ne peut pas rejeter H₀

57

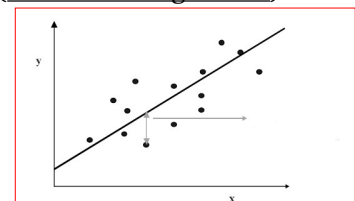
Le coefficient de détermination

Variance totale de la variable dépendante = Variance donné par la droite (variance de régression) + variance NON donnée par la droite (variance résiduelle)

$$r^2 = \frac{\text{Variance}_{\text{regression}}}{\text{Variance}_{\text{totale}}}$$

$$S_R^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - Y_i^{\text{predite}})^2$$

Variance residuelle



- Nommée : **coefficient de détermination**
- Notation: **d = r²** (le carrée du coefficient de corrélation Pearson)
- nombre de 0 à 1. On le transforme dans pourcentage pour l'interpréter (0 → 100%)

Interprétation: pourcentage de la variance (variation) d' une variable dépendante explique par la relation linéaire avec la variable indépendante

- plus d est grand plus la relation/association/lien est forte
- plus d est proche du 0, plus la relation est faible

58

Corrélation, régression, et causalité

- Le fait qu'on trouve dans une étude une corrélation statistiquement significative, ou un coefficient de détermination importante, ou si le coefficient de la régression linéaire est statistiquement significative, ne signifie pas fortement que la relation est causale !!!
- Ces informations, nous indique qu'il y a des relations, entre les variables, mais on ne sait pas qui est la cause et qui est l'effet.
- Pour montrer la causalité, la méthodologie de la réalisation de l'étude est très importante.

59

Régression linéaire multiple

But étudier la liaison/relation linéaire entre ≥ 2 variables indépendantes et 1 variable dépendante quantitative

Objectifs:

- **prédire les valeurs** d'une variable quantitative en fonction des valeurs des **plusieurs** autres variables
- évaluer
 - l'**existence** d'une association/liens/relation entre les variables
 - la **direction** de l'association/liens/relation entre variables
 - l'**importance** de l'association/liens/relation entre variables
- **ajuster/corriger/tenir compte** - l'effet d'autres variables

60

Régression linéaire multiple - Quantification de l'importance de la relation pour plusieurs variables

- Les relations entre différentes variables médicales sont souvent complexes, et plusieurs facteurs/ variables/ caractéristiques sont impliquées.
- Avec le coefficient de corrélation ou avec la régression linéaire simple on peut évaluer la relation entre seulement deux variables – donc une analyse univariée (entre une variable dépendante – et une variable explicative/ indépendante)
- Mais on a besoin de tenir compte de l'effet d'autres variables (nommée de confusion, ou autres variables explicatives) qui sont connues qu'ils influencent la relation qu'on étudie. On doit faire une étude bibliographique pour identifier ces variables de confusion.
- Pour tenir compte des plusieurs variables on doit utiliser des techniques multivariées (eg. Régression linéaire multiple, régression logistique multiple,)

61

61

Régression linéaire multiple - Quantification de l'importance de la relation pour plusieurs variables

- S'il y a plusieurs variables qui sont liées à la variable dépendante d'intérêt, on peut évaluer leur importance avec une **régression linéaire multiple** qui offre un coefficient (nommé ajusté – « adjusted » en anglais) pour chaque variable indépendante:
- La variable dépendante (prédite, expliquée) de la régression:
 - Les triglycérides (mg/dL)
- Les variables indépendantes (explicatives, prédictives)
 - Qualitatives (facteurs) (ex. le sexe)
 - Quantitatives (ex. le poids (kg))
- Le coefficient ajusté peut nous rapprocher plus à la vérité que le coefficient brut (« crude »/ « unadjusted » en anglais - sans ajustement) d'une régression linéaire simple, ou un coefficient de corrélation simple, parce que on peut tenir compte d'autres variables qui agissent dans le même temps

62

62

Régression linéaire multiple - Quantification de l'importance de la relation pour plusieurs variables

- L' équation du **régression linéaire multiple**:
- Variable dépendante= coefficient_1 * variable_1 + coefficient_2 * variable_2 + ... + coefficient_n * variable_n + coefficient_0
- Ex: triglycérides (mg/dL) = 23,10 * obésité (oui/non) + 1,14 * cholestérol (mg/dL)
- **L' interprétation du coefficient ajusté (adjusted – en anglais)** pour des variables **Qualitatives dichotomiques** (ex. obésité):
 - l'augmentation de la variable dépendante – les triglycérides en moyenne (ici il est de 23,1 mg/dL) pour ceux qui ont le facteur présent (être obèse - la variable indépendante) comparée à ceux qui n'ont pas le facteur (ne sont pas obèses), si on tiennent les autres variables constantes / si on ajuste les autres variables/ si on contrôle les autres variables) (ici – le cholestérol)
 - ceux qui ont le facteur présent (être obèse - la variable indépendante) ont la variable dépendante – les triglycérides en moyenne plus grand avec 23,1 mg/dL comparée à ceux qui n'ont pas le facteur (ne sont pas obèses), si on tiennent les autres variables constantes / si on ajuste les autres variables/ si on contrôle les autres variables) (ici – le cholestérol)

63

63

Régression linéaire multiple - Quantification de l'importance de la relation pour plusieurs variables

- L' équation du **régression linéaire multiple**:
- Variable dépendante= coefficient_1 * variable_1 + coefficient_2 * variable_2 + ... + coefficient_n * variable_n + coefficient_0
- Ex: triglycérides (mg/dL) = 23,10 * obésité (oui/non) + 1,14 * cholestérol (mg/dL)
- **L' interprétation du coefficient ajusté (adjusted – en anglais)** pour des variables **Quantitatives** (ex. le cholestérol):
 - l'augmentation de la variable dépendante – les triglycérides en moyenne (ici il est de 1,14 mg/dL) pour chaque changement d'un unité de mesure de la variable indépendante (chaque 1 mg/dL du cholestérol) si on tiennent les autres variables constantes / si on ajuste les autres variables/ si on contrôle les autres variables) (ici – l' obésité)
 - pour chaque unité de mesure de la variable indépendante (chaque 1 mg/dL du cholestérol) en plus, la variable dépendante – les triglycérides en moyenne augmente avec 1,14 mg/dL si on tiennent les autres variables constantes / si on ajuste les autres variables/ si on contrôle les autres variables) (ici – l' obésité)

64

64

Test statistique de signifiante pour les coefficients (b_1, b_2, \dots, b_n) de chaque variable indépendante de la régression linéaire multiple

- L' hypothèse nulle:
 - il n'y a pas de différence statistiquement significative entre le coefficient de regression (b_i) de de la variable X_i et 0
 - le coefficient (b_i) de la variable X_i n'est pas statistiquement significative
- L' hypothèse alternative:
 - il y a une différence statistiquement significative entre le coefficient (b_i) de de la variable X_i et 0
 - le coefficient (b_i) de la variable X_i est statistiquement significative
- La décision avec p-value.
 - Si $p\text{-value} < \alpha (=0,05) \Rightarrow$ on rejete H_0 et nous sommes en faveur de H_1
 - Si $p\text{-value} \geq \alpha (=0,05) \Rightarrow$ on ne peut pas rejeter H_0

65

Exemple

> Front Endocrinol (Lausanne). 2021 May 13;12:657856. doi: 10.3389/fendo.2021.657856. eCollection 2021.

Serum Uric Acid Is a Mediator of the Association Between Obesity and Incident Nonalcoholic Fatty Liver Disease: A Prospective Cohort Study

Qian Zhang ^{1,2}, Xiaoqian Ma ¹, Jie Xing ¹, Haiyun Shi ¹, Runkuan Yang ³, Yue Jiao ¹, Shuohua Chen ⁴, Shouling Wu ³, Shutian Zhang ¹, Xiujing Sun ¹

Abstract

Objective: Obesity has been demonstrated to show a consistent link with the increased possibility of nonalcoholic fatty liver disease (NAFLD). Since both serum uric acid (SUA) and obesity are essential components of metabolic syndrome (MetS), it is uncertain whether the incidence of NAFLD results from serum uric acid, obesity, or other potential factors based on previous studies.

Patients and methods: This study enrolled 16,839 participants with no history of alcohol consumption and no fatty liver disease in 2010. All participants completed a survey which included health and lifestyle questionnaires, and underwent physical examination, ultrasonography, and laboratory examinations of blood samples.

After the four-year follow up, 5,104 (30.31%) participants were diagnosed with NAFLD. The associations between SUA, BMI or obesity, and incident NAFLD were assessed by multivariate linear regression, logistic regression analysis, and mediation analysis, respectively.

66

66

Exemple: Modelés de régression linéaire multiple déterminer sur des sous-groupes de patients (femmes vs. hommes)

Résultats partiels extraits de l'article; lien vers l'article: <https://pubmed.ncbi.nlm.nih.gov/34054728/>

TABLE 3 | Linear regression of the influencing factors for SUA divided by gender.

Total population	β (95%CI)	P-value	Female	β (95%CI)	P-value	Male	β (95%CI)	P-value
Obesity	20.26 (15.74, 24.77)	<0.01		13.31 (6.63, 19.99)	<0.01		22.21 (16.41, 28.02)	<0.01
Age	-0.23 (-0.35, -0.12)	<0.01		0.30 (0.12, 0.47)	<0.01		-0.40 (-0.54, -0.25)	<0.01
Gender	57.61 (54.57, 60.64)	<0.01		—	—		—	—
Smoking	6.59 (3.33, 9.84)	<0.01		5.59 (2.43, 8.75)	<0.01		4.97 (1.73, 8.21)	<0.01
Marital status	34.43 (14.19, 54.67)	<0.01		23.43 (14.19, 32.67)	<0.01		38.18 (29.77, 46.58)	<0.01
Working type	-11.02 (-14.70, -7.34)	<0.01		-9.71 (-11.13, -8.29)	<0.01		-16.29 (-21.78, -10.79)	<0.01
Education level	0.32 (-2.85, 3.48)	0.84		0.01 (-2.85, 2.96)	0.01		-3.52 (-7.79, 0.741)	0.11
Physical activity	-7.71 (-10.34, -5.09)	<0.01		-6.51 (-9.14, -3.88)	<0.01		-7.85 (-11.25, -4.45)	<0.01
TG	6.51 (5.53, 7.48)	<0.01		8.26 (7.12, 9.39)	<0.01		9.25 (7.79, 10.72)	<0.01
TC	2.87 (1.92, 3.82)	<0.01		2.77 (1.07, 3.26)	<0.01		2.95 (1.51, 4.40)	<0.01
FBG	-2.18 (-2.95, -1.42)	<0.01		-0.47 (-1.42, 0.48)	0.34		-3.62 (-4.70, -2.54)	<0.01
SBP	0.32 (0.24, 0.39)	<0.01		0.18 (0.07, 0.28)	<0.01		0.36 (0.27, 0.45)	<0.01
CRP	1.75 (1.50, 2.00)	<0.01		1.30 (0.90, 1.71)	<0.01		1.94 (1.64, 2.25)	<0.01
Cr	-0.31 (-0.36, -0.26)	<0.01		-0.13 (-0.22, -0.04)	<0.01		-0.37 (-0.44, -0.31)	<0.01

CI, Confidence Interval; Cr, Creatinine; CRP, C-reactive Protein; FBG, Fasting Blood Glucose; Obesity: body mass index $\geq 28 \text{ kg/m}^2$; SBP, Systolic Blood Pressure; SUA, Serum Uric Acid; TC, Total Cholesterol; TG, Triglyceride.

les variables indépendantes (explicatives)

Source: Zhang Q, Ma X, Xing J, Shi H, Yang R, Jiao Y, Chen S, Wu S, Zhang S, Sun X. Serum Uric Acid Is a Mediator of the Association Between Obesity and Incident Nonalcoholic Fatty Liver Disease: A Prospective Cohort Study. Front Endocrinol (Lausanne). 2021 May 13;12:657856. doi: 10.3389/fendo.2021.657856.

67

Regression logistique

- **But:** évaluer la liaison entre **1 variable qualitative** (dichotomique/nominale/ordinaire) **dépendante** et ≥ 1 **variables** quantitatives ou qualitatives **indépendantes**
 - Deux **objectifs** possibles:
 - on cherche d'expliquer la survenue (mesurer par l' odds /probabilité) d'un événement binaire (oui/non) dépendante: tester d' hypothèses
 - on cherche de prédire la survenue (mesurer par l' odds /probabilité) d'un événement binaire (oui/non) dépendante: model prédictif
- en présence de covariables = toute variable indépendante associée à la variable dépendante dans un modèle de régression.

68

Regression logistique

- Très populaire en médecine: la variable dépendante peut être l'apparition d'une maladie, et les variables dépendantes sont des facteurs de risque.
- La variable dépendante doit être une **variable qualitative dichotomique**
 - Présence de maladie ou de décès,
 - Diabète (oui / non)
 - Hypertension (oui / non)
- **Variables indépendantes** : ≥ 1 variables de différents types
 - **Qualitatives**: Consommation d'alcool (oui/non), tabagisme (oui/non), obésité (oui/non)
 - **Qualitatives**: Consommation d'alcool (oui/non), tabagisme (oui/non), obésité (oui/non)
- *Conditions d'applicabilité*
- les observations sont indépendantes les unes des autres
- La relation entre chaque prédicteur continue et la variable dépendante est linéaire sur l'échelle logit

69

69

Model de regression logistique simple

X_1 –variable independante et Y =variable dépendante

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X_1$$

ou

$p = \Pr(Y=1/X_1)$ la probabilité de survenu d'un evenement en fonction de la valeur prise par la variable indépendante (X_1)

b_0 = le terme libre (constante du modèle)

b_1 = coefficient de régression partielle -> le rapport des chances (OR)

70

70

Model de regression logistique simple

- Odds ($Y=1/X_1 = 0$ =absente) = e^{b_0}
- Odds ($Y=1/X_1 = 1$ =present) = $e^{b_0+b_1}$
- l'odds ratio pour $Y=1 \rightarrow OR = \frac{\text{Odds}(Y=1/X_1 = 1)}{\text{Odds}(Y=1, X_1 = 0)} = \frac{e^{b_0+b_1}}{e^{b_0}} = e^{b_1}$

71

71

Model de regression logistique multiple

X_1, X_2, \dots, X_k –variables independantes et Y =variable dependante

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

ou

$p = \Pr(Y=1/X_1, X_2, \dots, X_k)$ = la probabilité de survenu d'un eveniment en fonction des valeurs prise par les variables indépendantes (X_1, X_2, \dots, X_k)

b_0 = le terme libre (constante du modèle)

b_1, \dots, b_k = coefficient de régression logistique \rightarrow le rapport des chances (OR)

72

Model de regression logistique multiple

Interprétation de l'odds ratio (OR) du modèle logistique quand la variable indépendante (X_i) est:

- **dichotomique:**

$$OR = \frac{\text{l'odds de l'event pour group exposé}}{\text{l'odds de l'event pour group nonexposé (de reference)}}$$

si $b_i > 0$ alors $OR > 1$ (facteur de risk),
si $b_i < 0$ alors $OR < 1$ (facteur protecteur).

- **nominal avec p categories:**

- On obtient « p-1 » des OR (la variable nominale doit être transformée en variable dichotomique p-1 avant la régression)

- **quantitative continues:**

- représente le changement de l'odds de l'événement d'intérêt lorsque X_i augmente d'une unité après ajustement pour les autres prédicteurs

73

La performance du modele logistique

- Le test de Hosmer – Lemeshow ne devrait pas être significatif
 - Hypothese nulle H_0 : Le modèle testé est adéquat pour les données.
- précision de la prédiction (en %)
- Calcul des coefficients de McFadden R^2 , Cox-Snell R^2 , Nagelkerke / Cragg & Uhler's R^2

74

Quantification de l'importance de la relation pour plusieurs facteurs

Régression logistique simple, multiple, odds ratio crue/ajusté...

- Une **régression logistique** avec une **seule variable indépendante**
 - permet une **analyse univariée** – on parle d'un **régression logistique simple**
 - nous offre un **odds ratio brut/crue** (« crude »/ » **unadjusted**» en anglais - sans ajustement, pour la variable indépendante
- Une **régression logistique** avec une **plusieurs variables indépendantes**
 - permet une **analyse multivariée** – on parle d'un **régression logistique multiple**
 - nous offre un **odds ratio ajusté** (« adjusted » en anglais) pour la variable indépendante – avec ajustement pour les autres variables indépendants dans la régression
- L'**odds ratio ajusté** peut nous rapprocher plus a la vérité que le **odds ratio brut** (« crude »/ » **unadjusted**» – obtenu a l'aide d'une régression logistique simple, ou calcul $a*d/b*c$), parce que on peut tenir compte (on peut contrôler) d'autres facteurs de risque qui agit dans le même temps

75

75

Quantification de l'importance de la relation pour plusieurs facteurs

Interprétation statistique du odds ratio crue/brut/non ajuste d'une variable qualitative dichotomique pour un régression logistique simple (uni variée)

- L' équation du **régression logistique simple (uni variée)**, avec un **variable indépendante qualitative dichotomique**:
 - $\ln(\text{odds}(\text{Maladie} = \text{oui}=1)) = b_0 + b_1 * \text{facteur } X_1 (\text{oui}=1/\text{non}=0) \rightarrow$
 - $L'odds(\text{Maladie} = \text{oui}=1) = e^{b_0} * e^{b_1 * \text{facteur } X_1 (\text{oui}=1/\text{non}=0)} \rightarrow OR_1 = \frac{\text{Odds}(Y=1/X_1 = 1)}{\text{Odds}(Y=1, X_1 = 0)}$
- L'**odds ratio** (Maladie =oui) = OR_1 associé au facteur X_1 (oui/non)
- Ex: L'**odds ratio** (cheilopalatoschisis =oui) = $OR_1 = 1,90$ associé au facteur Alcool (oui/non)
- L' **interprétation du Odds ratio crue/brut, sans ajustement**
- Si la **variable est qualitatives dichotomiques** (ex. consommation d'alcool):
 - il y a **1,90 fois plus de la chance d'avoir la maladie - cheilopalatoschisis pour ceux avec exposition** – avec consommation d'alcool comparée a ceux qui n'ont pas été exposée a l'alcool.

76

76

Quantification de l'importance de la relation pour plusieurs facteurs

Interprétation statistique du odds ratio crue/brut/non ajusté d'un variable quantitative pour un régression logistique simple (uni variée)

- L' équation du **régression logistique simple (uni variée)**, avec un **variable indépendante quantitative**:
- **Ln (odds (Maladie =oui))** = $b_0 + b_2 * \text{facteur } X_2$ (quantitative)
- **L'odds (Maladie =oui)** = $e^{b_0} * e^{b_2 * \text{facteur } X_2} \rightarrow OR_2 = \frac{\text{Odds (Y=1/X}_2 = a+1)}{\text{Odds (Y=1, X}_2 = a)}$ ou a=valeur fixé de la variable X_2
- **L'odds ratio (Maladie =oui)** = OR_2 associé au facteur X_2
- Ex: **L'odds ratio (cheilopalatoschizis =oui)** = $OR_2 = 1,10$ associé au facteur Age de la mère (années)
- L' **interprétation de l' Odds ratio crue/brut, sans ajustement**
- **Si la variable est quantitative** (ex. âge de la mère):
 - **pour chaque** unité de mesure de la variable (pour chaque **année de la mère**) **en plus la chance d'avoir cheilopalatoschizis augmente 1,10 fois**

77

77

Régression logistique simple– exemple interprétation

Caractéristique	OR crude (95% CI)	P-value
Alcool pendant la grossesse (oui/non)	2,25 (1,64 – 7,83)	<0,001
Age de la mère (années)	1,22 (1,03 – 1,85)	0,03

Régression logistique uni variée pour la prédiction de l'odds du cheilopalatoschizis.

L'OR – odds ratio - **crude – brut - unadjusted** (voir tableau) – (brut signifie qu'il s'agit - entre deux variables seulement – une variable dépendant (le cheilopalatoschizis) et **une variable indépendante** (la consommation d'alcool pendant la grossesse) - d'un analyse **uni variée**. Ici dans le tableau il y a deux régressions uni variées, entre cheilopalatoschizis et la consommation d'alcool pendant la grossesse, et une entre cheilopalatoschizis et l'âge de la mère

Pour la variable dichotomique: - OR crude – brut - unadjusted pour la consommation d'alcool pendant la grossesse : il y a 2,25 fois plus de chance de cheilopalatoschizis pour ceux qui ont été exposés à l'alcool par rapport aux ceux qui n'ont pas été exposés.

Pour la variable quantitative: - OR crude – brut - unadjusted pour l'âge de la mère: il y a 1,22 fois plus de chance de cheilopalatoschizis pour chaque unité de mesure en plus de l'âge de la mère (pour chaque année)

78

Régression logistique simple & multiple – exemple interprétation

Caractéristique	OR crude (95% CI)	P-value	OR adjusted (95% CI)	P-value
Alcool pendant la grossesse (oui/non)	2,25 (1,64 – 7,83)	<0,001	2,10 (1,52 – 9,95)	0,001
Age de la mère (années)	1,22 (1,03 – 1,85)	0,03	1,15 (1,02 – 1,78)	0,046

Régression logistique **multiple (multi variée)**, pour la prédiction de l'odds du cheilopalatoschizis.

OR – odds ratio - ajusté - adjusted (voir tableau) – (ajusté signifie qu'il s'agit d'une analyse entre une variable dépendante (le cheilopalatoschizis) et plusieurs variables indépendantes (consommation d'alcool pendant la grossesse et l'âge de la mère) – analyse **multivariée**

Variable dichotomique: - OR ajusté - adjusted pour la consommation d'alcool pendant la grossesse : il y a 2,10 fois plus de chance de cheilopalatoschizis pour ceux qui ont été exposés à l'alcool par rapport à ceux qui n'ont pas été exposés, on tenant constantes (en tenant fixe/en ajustant/ en contrôlant) les autres variables (âge de la mère).

Variable quantitative: - OR ajusté - adjusted pour l'âge de la mère : il y a 1,15 fois plus de chance de cheilopalatoschizis pour chaque unité de mesure en plus de l'âge de la mère (pour chaque année), on tenant constantes (en tenant fixe/en ajustant/ en contrôlant) les autres variables (consommation d'alcool pendant la grossesse).

79

Analyse multivariée: la Quantification de l'importance de la relation pour plusieurs facteurs

Interprétation statistique du odds ratio ajusté d'un variable qualitative dichotomique, et d'un variable quantitative pour un régression logistique multiple (multivariée)

- L'équation du **régression logistique multiple (multivariée)**:
- $\ln(\text{odds}(\text{Maladie}=\text{oui})) = b_0 + b_1 * X_1 + b_2 * X_2$
- Nous modélisons l'odds de la maladie à l'aide de deux odds ratios ajustés (**OR₁ ajusté** et **OR₂ ajusté**)
- L'interprétation du **Odds ratio ajusté** pour des variables
 - **Qualitatives dichotomiques** (ex. consommation d'alcool): **OR₁ = ajusté** $\frac{\text{Odds}(Y=1/X_1=1, X_2)}{\text{Odds}(Y=1, X_1=0, X_2)}$ > il y a 2,10 fois plus de la chance d'avoir la maladie - cheilopalatoschizis pour ceux avec exposition – avec consommation d'alcool comparée à ceux qui n'ont pas été exposés à l'alcool si on tient l'autre variable constante (on ajuste pour la 2ème variable ici l'âge)
 - **Quantitatives** (ex. âge de la mère): **OR₂ = ajusté** $\frac{\text{Odds}(Y=1/X_2=a+1, X_1)}{\text{Odds}(Y=1, X_2=a, X_1)}$ pour chaque unité de mesure de la variable (pour chaque année de la mère) en plus la chance d'avoir cheilopalatoschizis augmente 1,15 fois si on tient l'autre variable constante/fixé (on ajuste pour l'autre variable - ici l'alcool)

80

80

Exemple d'article scientifique

> Can J Infect Dis Med Microbiol. 2021 May 7;2021:5515941. doi: 10.1155/2021/5515941. eCollection 2021.

Risk Factors for COVID-19 in Patients with Hypertension

Xinxin Wang ¹, Haihua Zhang ¹, Huan Du ¹, Ruina Ma ^{1 2}, Yandong Nan ^{1 2}, Tao Zhang ^{1 2}

Abstract

- **Background:** Hypertension, as the most common comorbidity for patients with coronavirus disease 19 (COVID-19), has resulted in cases with more severe symptoms and higher mortality. The risk factors associated with COVID-19 in patients with hypertension are unknown.
- **Methods:** All the available and confirmed patients with COVID-19 from February 3 to March 10, 2020, were enrolled from Huoshenshan Hospital, Wuhan, China. The demographic characteristics, clinical manifestations, laboratory data, radiological assessments, and treatments on admission were extracted and compared. Univariate and multivariate logistic regression methods were used to explore risk factors associated with COVID-19 in patients with hypertension and the severity of the cohort.

Lien vers l'article: <https://pubmed.ncbi.nlm.nih.gov/34055114/>

81

Exemple: Modelés de régression logistique simple et multiple

• Résultats partiels extraits de l'article;

TABLE 4: Univariate logistic regression analysis of patients with COVID-19 and hypertension.

	Univariate OR (95% CI)	p value
Diabetes	2.148 (1.074–4.297)	0.031
Coronary heart disease	3.009 (1.344–6.733)	0.007
Cerebrovascular disease	4.279 (1.181–15.510)	0.027
Percentage of PLV, %	1.025 (1.004–1.046)	0.021
White blood cell count, ×10 ⁹ /L	1.133 (1.035–1.241)	0.007
Neutrophil count, ×10 ⁹ /L	1.128 (1.032–1.232)	0.008
ALP, IU/L	1.007 (0.998–1.016)	0.112
Urea nitrogen, mmol/L	1.142 (1.037–1.257)	0.007
UA (uric acid), umol/L	1.003 (1–1.005)	0.028
Creatinine, umol/L	1.016 (1.004–1.027)	0.008
K ⁺	0.582 (0.383–0.885)	0.011

ALT: alanine aminotransferase; OR: odds ratio; CI: confidence interval.

Analyse univarié de la régression logistique (la régression logistique simple)

L'estimation ponctuelle de l'odds ratio non-ajusté (brut/crue) & 95% intervalle de confiance (IC)

TABLE 5: Multivariate logistic regression analysis of patients with COVID-19 and hypertension.

	Multivariate OR (95% CI)	p value
Diabetes	2.438 (0.824–7.210)	0.107
Coronary heart disease	5.281 (1.462–19.076)	0.011
Cerebrovascular disease	5.661 (0.952–33.662)	0.057
Percentage of PLV, %	0.994 (0.963–1.026)	0.71
Neutrophil count, ×10 ⁹ /L	1.471 (1.183–1.828)	0.001
Urea nitrogen, mmol/L	0.994 (0.795–1.244)	0.961
UA (uric acid), umol/L	1.003 (0.998–1.008)	0.247
Creatinine, umol/L	1.018 (0.994–1.044)	0.149
K ⁺	0.273 (0.134–0.555)	<0.001

OR, odds ratio; CI, confidence interval.

Analyse multivarié de la régression logistique (le model multivariable de la régression logistique multiple)

L'estimation ponctuelle de l'odds ratio ajusté & 95% intervalle de confiance (IC)

Source: Wang X, Zhang H, Du H, Ma R, Nan Y, Zhang T. Risk Factors for COVID-19 in Patients with Hypertension. Can J Infect Dis Med Microbiol. 2021 May 7;2021:5515941. doi: 10.1155/2021/5515941.

82

82

Regression Cox (pour des données de survie)

Un étude a évalué la survie en fonction de la thérapie de remplacement hormonale (HRT) pour des patientes avec cancer du poumon. L'article montre les résultats d'un régression multivariée

Cox:

Variable	Hazard Ratio	95% CI	P
HRT indicates hormone replacement therapy; CI, confidence interval.			
History of HRT use (yes vs no)	1.09	0.82-1.44	.55
Age at cancer diagnosis (years)	1.03	1.02-1.05	<.01
TNM stage (baseline=stage I)			<.01
Stage II	1.56	0.88-2.75	.13
Stage III	3.47	2.51-4.79	<.01
Stage IV	5.91	4.31-8.11	<.01
Smoking history (yes vs no)	1.41	1.03-1.91	.03

Comme il est une régression multivariée – ça signifie qu'il y a plusieurs variables dans la régression, et donc l'analyse est plus complexe parce que il tenant compte de l'effet de tous ces variables ensemble sur la survie (on dit que la régression est ajustée / contrôlée pour plusieurs variables - toutes les variables dans le tableau ici sont analysée au même temps). L'interprétation des hazard ratios est similaire au régressions uni variées (avec une seule variable indépendante), mais on ajoutés le texte à la fin « en tenant/ en gardant toutes les autres variables constantes » ou on ajustant, on contrôlant les autres variables dans la régression. Les variables qu'on ajuste/qu'on contrôle avec la régression sont des facteurs de confusion (en anglais – confounding factors)

Huang B, Carloss H, Wyatt SW, Riley E. Hormone replacement therapy and survival in lung cancer in postmenopausal women in a rural population. Cancer. 2009 Sep 15;115(18):4167-75.

83

83

Regression Cox (pour des données de survie)

Variable	Hazard Ratio	95% CI	P
HRT indicates hormone replacement therapy; CI, confidence interval.			
History of HRT use (yes vs no)	1.09	0.82-1.44	.55
Age at cancer diagnosis (years)	1.03	1.02-1.05	<.01
TNM stage (baseline=stage I)			<.01
Stage II	1.56	0.88-2.75	.13
Stage III	3.47	2.51-4.79	<.01
Stage IV	5.91	4.31-8.11	<.01
Smoking history (yes vs no)	1.41	1.03-1.91	.03

Comme il est une régression multivariée pour l'analyse de survie – la **variable dépendante** (prédite/ expliquée – qu'on essaye de le prédire) est la survie (la probabilité de survie en fonction de temps), et il y a plusieurs **variables indépendantes** (explicatives, avec lesquels on essaye de prédire la variable dépendante).

L'interprétation hazard ratio (HR) pour une variable quantitative - ex. l'âge au moment du diagnostic de cancer (age at cancer diagnosis): le hazard est 1,03 fois plus grand pour chaque unité de mesure (année - year) en plus de l'âge, en gardant toutes les autres variables constantes (HRT, stade TNM, l'histoire du tabagisme). Comme le hazard ratio est supérieure à 1 indique que l'âge avancée au moment du diagnostic est associée à une pire survie.

84

84

Regression Cox (pour des données de survie)

Variable	Hazard Ratio	95% CI	P
HRT indicates hormone replacement therapy; CI, confidence interval.			
History of HRT use (yes vs no)	1.09	0.82-1.44	.55
Age at cancer diagnosis (years)	1.03	1.02-1.05	<.01
TNM stage (baseline=stage I)			<.01
Stage II	1.56	0.88-2.75	.13
Stage III	3.47	2.51-4.79	<.01
Stage IV	5.91	4.31-8.11	<.01
Smoking history (yes vs no)	1.41	1.03-1.91	.03

L'interprétation hazard ratio (HR) pour une variable dichotomique - ex. si la personne qui fume (antécédents de tabagisme – smoking history) le hasard est 1,41 fois plus élevé pour ceux qui fumaient (oui) par rapport à ceux qui n'avaient jamais fumé (non), en gardant toutes les autres variables constantes (âge, HRT, le stade TNM) . Comme le hazard ratio est supérieur à 1 indique que les fumeurs ont une plus mauvaise survie.

85

85

Regression Cox (pour des données de survie)

L'interprétation hazard ratio (HR) pour une variable qualitatives ordinales (ou nominale - interprétation est la même) - ex. stade de la tumeur (TNM stage – baseline - ligne de base = étape I - est une mise en scène d'extension de tumeur, le stade IV étant sujets avec métastases et la ligne de base est le groupe auquel on se rapporte pour toutes les comparaisons (est le groupe de référence)).

- Les interprétations se font par comparaison de chaque catégorie II, III, IV, vers le stade I comme l'interprétation des variables dichotomiques):
- le hasard est 1,56 fois supérieure à ceux de la phase II à la phase I, en gardant toutes les autres variables constantes (âge, HRT, le tabagisme) - mais il est pas statistiquement significative;
- le hasard est 3,47 fois supérieure à celles de stade III par rapport a le stade I, en maintenant toutes les autres variables constantes (âge, HRT, tabagisme) - statistiquement significatif – parce que le hazard ratio est supérieur à 1 indique que le stade de III ont une pire survie que ceux de stade I.

86

86

MERCI POUR VOTRE ATENTION

87