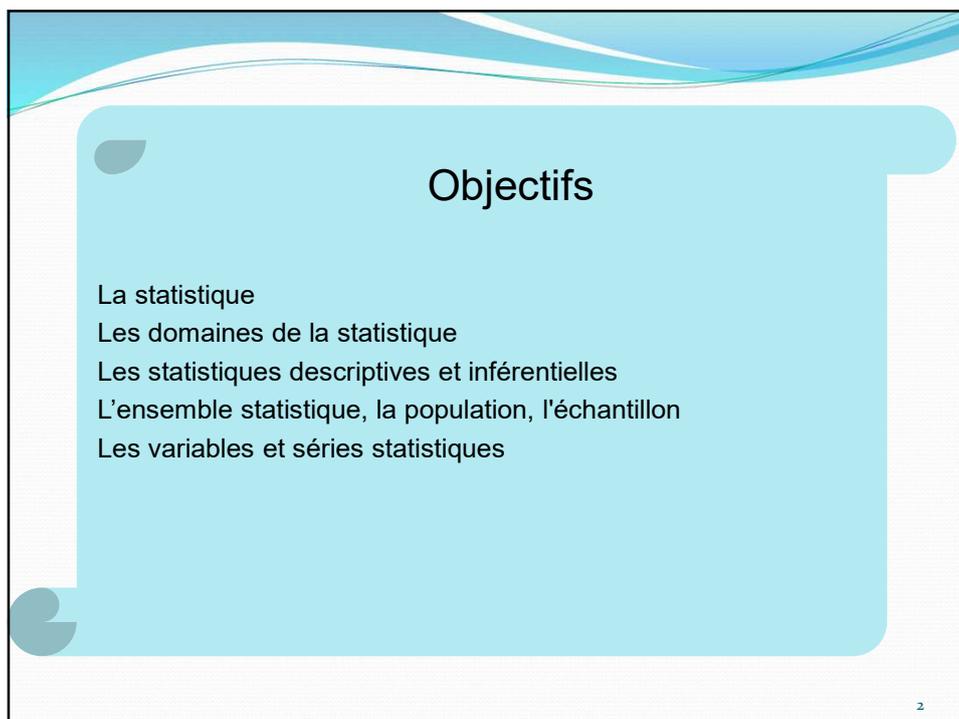




1



2

Eléments de Bio statistique

Statistique

- domaine scientifique mathématique
- **permet l'étude des phénomènes qui ont comme propriété fondamentale la variabilité**
- but d'étudier **les collectivités/populations** (personnes, etc.)
 - l'étude d'ensembles d'observations se fait sur un nombre des entités similaires (personnes, objets, etc.).
 - par rapport a, au moins une caractéristique ou attribut commun
 - mais **variable**

3

3

Domaines de statistique

- **Statistique Descriptive**
- **Théorie de Probabilités**
- **Statistique Analytique (Inférentielle)**

4

4

Domaines de statistique

Statistique Descriptive s'occupe avec:

- La **collecte** de **données** statistiques,
- La **présentation** des données statistiques (**tableaux, graphiques**)
- La détermination de certains **indicateurs synthétique** de données statistiques

Base des données collectées dans une étude

N patient	Âge (années)	Sexe	Poids (kg)	Glycémie (mg/dL)
1	65	h	82	156
2	73	f	64	137
3	57	h	78	129
4	82	f	73	117
...
237	79	f	81	198



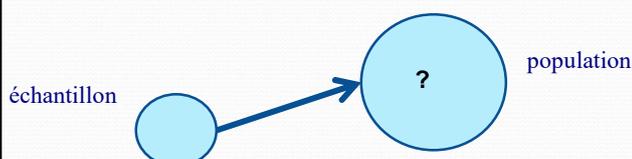
Caractéristique	
Âge (années)	Moyenne = 68,3
Femmes	N= 342, % = 53,2%
Poids (kg)	Moyenne = 74,4
Glycémie (mg/dL)	Moyenne = 146

5

Domaines de statistique

Statistique Analytique = Statistique Inductive ou Inferentielle

- méthodes pour obtenir des **conclusions sur une population** entière
- **en observant** seulement **une** partie de cette population (**échantillon**)
- méthodes :
 - l'**estimation** des **paramètres** statistiques
 - (Quelle est la moyenne dans la population?)
 - la vérification des **hypothèses statistiques**
 - (Est-ce que deux traitements ont des résultats différents)



6

Domaines de statistique

Statistique descriptive
Statistique analytique

Table 1
Baseline characteristics of patients with SARS-CoV-2 infection.

Characteristic	COVID-19 patients			P value
	Total (N = 80)	FPV (N = 35)	LPV/RTV (N = 45)	
Age, median (IQR)	47 (35.75–61)	43 (35.5–59)	49 (36–61)	0.61
Age subgroup				
15–44	36 (45.0%)	18 (51.4%)	18 (40.0%)	–
45–64	33 (41.3%)	13 (37.1%)	20 (44.4%)	–
≥ 65	11 (13.7%)	4 (11.4%)	7 (15.6%)	0.47
Male	35 (43.8%)	14 (40.0%)	21 (46.7%)	0.55
BMI, median (IQR)	22.9 (16.2–31.6)	22.7 (16.2–31.6)	23.1 (16.4–28.4)	0.51
Epidemiology				
History of visiting Wuhan City	46 (57.5%)	20 (57.1%)	26 (57.8%)	–
Not been to Wuhan City	34 (42.5%)	15 (20.0%)	19 (17.8%)	0.95
Onset symptoms				
Fever	59 (73.8%)	22 (62.9%)	37 (82.2%)	0.11
Cough	22 (27.5%)	12 (34.3%)	10 (22.2%)	0.23
Headache/myalgia	8 (10.0%)	3 (8.6%)	5 (11.1%)	1.00
Chill	1 (1.3%)	0 (0%)	1 (2.2%)	1.00
Diarrhea	1 (1.3%)	1 (2.9%)	0 (0%)	0.44
Stuffy nose/sore throat	8 (10.0%)	6 (17.1%)	2 (4.4%)	0.13

Statistique analytique independent variables. The result showed that there were two statistically significant factors in the model: antiviral therapy (odds ratio (OR) = 3.190, 95% confidence interval (CI) = 1.041–9.780) and fever (OR = 3.622, 95%CI = 1.054–12.442). This means that

Cai Q, Yang M, Liu D, Chen J, Shu D, Xia J, Liao X, Gu Y, Cai Q, Yang Y, Shen C, Li X, Peng L, Huang D, Zhang J, Zhang S, Wang F, Liu J, Chen L, Chen S, Wang Z, Zhang Z, Cao R, Zhong W, Liu Y, Liu L. Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study. Engineering (Beijing). 2020 Oct;6(10):1192-1198. doi: 10.1016/j.eng.2020.03.007. Epub 2020 Mar 18. PMID: 32346491; PMCID: PMC7185795.

7

Notions fondamentales de statistique

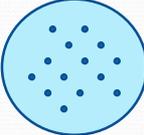
Ensembles statistiques (population, échantillon)

Population

- un ensemble des entités (personnes, objets, etc.)
- qui ont au moins un attribut (caractéristique) commun

La **dimension/taille** de la population - le nombre des éléments (finie, infinie)

Unité statistique = un élément de la population



Pour définir une population on utilise des critères

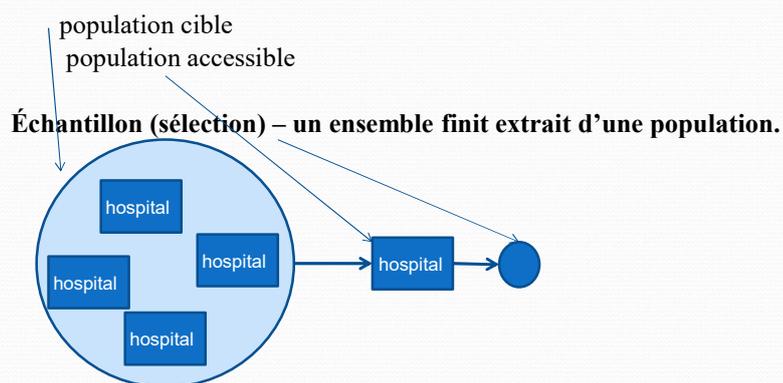
- d'inclusion – pour le critère commun
- d'exclusion – pour les éléments qui ont déjà les critères d'inclusion

8

Notions fondamentales de statistique

Ensembles statistiques

Population



9

Ensembles statistiques

Raisons pour utiliser les échantillons

- plus **vite** que les populations
- **moins chère**
- parfois, le processus “d’observation” peut **détruire** les éléments observés
- les résultats sur échantillons sont souvent **plus précis** que les résultats sur une population
- **éthique**

10

Échantillonnage

Un échantillon **représentative** = qui copie bien les caractéristiques de la population

	Population	Echantillon
% des femmes	51%	~ 51% (ex. 52%)
Moyenne d'âge	45	~ 45 (ex. 43,7)
Tabagisme	25%	~ 25% (ex. 29%)

Pour construire un échantillon **représentative** il faut **sélectionner** ses éléments d'une manière **aléatoire** avoir un **nombre suffisant d'éléments**.

11

Échantillonnage

Types d'échantillonnage:

- Échantillonnage **probabiliste**
 - avec un sélectionne aléatoire (au hasard) depuis la population . On peut calculer la probabilité d'être inclus . On peut faire des bonnes estimations sur la population. Il es plus complexe, plus chère, que l'échantillonnage non probabiliste. Plus utilise dans des études de sante publique, économie médicale, ...
 - aléatoire simple
 - stratifié
 - systématique
 - en grappes (cluster – en anglais)
- Échantillonnage **non probabiliste**
 - C'est le plus utilise dans les études cliniques
 - pratique
 - consécutive

12

Échantillonnage

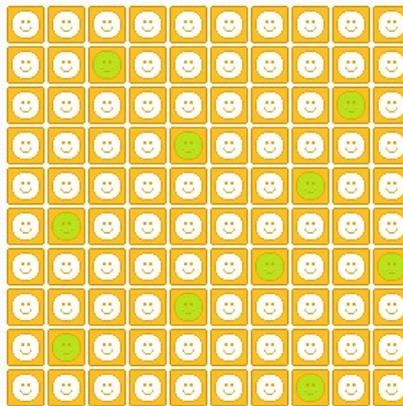
Types d'échantillonnage:

- Échantillonnage **probabiliste**
 - Les meilleurs échantillonnages probabiliste dans ordre descendante,
 - stratifié
 - aléatoire simple
 - systématique

13

ÉCHANTILLONNAGE

Échantillonnage probabiliste aléatoire simple – par l'extraction aléatoire (randomise) d'éléments d'une population.



Le smiley est modifiée après:
http://commons.wikimedia.org/wiki/File:Map_marker_icon_%E2%80%93_Nicolas_Mollet_%E2%80%93_Smiley_%E2%80%93_People_%E2%80%93_Default.png – le caree, est production personnelle

14

ÉCHANTILLONNAGE

Échantillonnage probabiliste aléatoire simple

› Zhonghua Yu Fang Yi Xue Za Zhi. 2019 Nov 653(11):1147-1151.
doi: 10.3760/cma.j.issn.0253-9624.2019.11.014.

[The association of pre-pregnancy body mass and weight gain during pregnancy with macrosomia: a cohort study]

[Article in Chinese]
P Feng¹, X Y Wang¹, Z W Long¹, S F Shan¹, D T Li¹, Y Liang¹, M X Chen¹, Y H Gong², R Zhou², D G Yang³, R N Duan¹, T Qiao¹, Y Chen¹, J Li¹, G Cheng⁴

Affiliations + expand
PMID: 31683403 DOI: 10.3760/cma.j.issn.0253-9624.2019.11.014

Abstract in English, Chinese

Objective: To examine the association of pre-pregnancy body mass and weight gain during pregnancy with macrosomia. **Methods:** From January 2015 to December 2015, a total of 20 477 pregnant women were recruited by probabilistic proportional scale sampling with **simple randomization** in Sichuan, Yunnan and Guizhou Provinces. Basic information of pregnant women, weight gain during pregnancy and weight of newborn were collected. A multiple logistic regression model was used to assess the association between the pre-pregnancy body mass and gestational weight gain indicators with macrosomia. **Results:** 20 321 mother-infant were included in the final analysis. 20 321 pregnant women were (30.09±4.10) years old and delivered at (39.20±1.29) weeks, among which 12 341 (60.73%) cases were cesarean delivery. The birth weight of 20 321 infants were (3 292.26±431.67) grams, and 970 (4.77%) were macrosomia. The multiple logistic regression model showed that after adjusting for the age of women, compared to the normal weight group in the pre-pregnancy, the overweight and obesity group elevated the risk of macrosomia, with OR (95%CI) about 1.99 (95%CI: 1.69-2.35) and 4.05 (95%CI: 3.05-5.39), respectively. After adjusting for the age, the pre-pregnancy BMI, delivery weeks, delivery mode and infant's gender, compared to the weight-gain appropriate group, higher weight gain rate in the mid-pregnancy and excessive total gestational weight gain elevated the risk of macrosomia, with OR (95%CI) about 1.99 (95%CI: 1.66-2.39) and 1.80 (95%CI: 1.55-2.08), respectively. **Conclusion:** The overweight before pregnancy, obesity before pregnancy, the rate of weight gain in the second trimester and the high total weight gain during pregnancy could increase the risk of macrosomia.

15

ÉCHANTILLONNAGE

Échantillonnage probabiliste

Aléatoire stratifié

– par l'extraction aléatoire (randomisée) d'éléments d'une population, mais sur des strates avec certain proportion.

Eg. Si on sait que le pourcentage des sujets avec maladie moins grave est 70%, et 30% pour ceux avec maladie grave,

on va extraire dans un manière aléatoire,

des sujets jusqu'a on va trouver 70% des sujets avec maladie moins grave (7).

Si on trouve moins des sujets avec maladie grave (seulement 2 initialement) on va ignorer un parti de ceux avec maladie grave jusqu'à on trouve des sujets avec maladie grave pour réaliser le pourcentage de ces sujets qu'on est intéressé

16

ÉCHANTILLONNAGE

Échantillonnage probabiliste

Aléatoire stratifié

> Vaccine. 2021 Jan 8;39(2):247-254. doi: 10.1016/j.vaccine.2020.12.009. Epub 2020 Dec 5.

Individual preferences for COVID-19 vaccination in China

Anli Leng ¹, Elizabeth Maitland ², Siyuan Wang ³, Stephen Nichc

Affiliations + expand
 PMID: 33328140 PMCID: PMC7719001 DOI: 10.1016/j.vaccine.2
 Free PMC article

Using a stratified random sampling method, six Chinese provinces were chosen based on high, medium and low GDP per capita, geographically covering east, central and west regions of China. In each province, three cities were chosen, also divided into high, medium and low GDP per capita, with 314 individuals randomly surveyed in each city.

Abstract

Background: Vaccinations are an effective choice to stop disease outbreaks, including COVID-19. There is little research on individuals' COVID-19 vaccination decision-making.

Objective: We aimed to determine individual preferences for COVID-19 vaccinations in China, and to assess the factors influencing vaccination decision-making to facilitate vaccination coverage.

Methods: A D-efficient discrete choice experiment was conducted across six Chinese provinces selected by the stratified random sampling method. Vaccine choice sets were constructed using seven attributes: vaccine effectiveness, side-effects, accessibility, number of doses, vaccination sites, duration of vaccine protection, and proportion of acquaintances vaccinated. Conditional logit and latent class models were used to identify preferences.

Results: Although all seven attributes were proved to significantly influence respondents' vaccination decision, vaccine effectiveness, side-effects and proportion of acquaintances vaccinated were the most important. We also found a higher probability of vaccinating when the vaccine was more effective; risks of serious side effects were small; vaccinations were free and voluntary; the fewer the number of doses; the longer the protection duration; and the higher the proportion of acquaintances vaccinated. Higher local vaccine coverage created altruistic herd incentives to vaccinate rather than

17

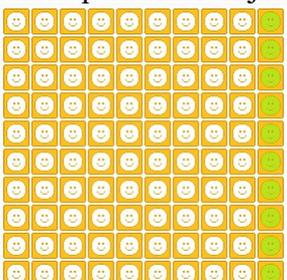
ÉCHANTILLONNAGE

systématique

on sélection chaque k-eme élément

e.g. on utilise la liste des patients inscrits dans le cabinet d'un médecin.

Si on veut un échantillon de 200 sujets, et la liste a 2000 sujets, on va extraire chaque 10eme sujet



18

ÉCHANTILLONNAGE

systematique

> Pan Afr Med J. 2021 May 13;39:38. doi: 10.11604/pamj.2021.39.38.27648. eCollection 2021.

Availability of PPEs and training status of health professionals on COVID-19 in Silte Zone, Southern Ethiopia

Mohammed Muze ¹, Bahredin Abdella ¹, Abdilmejid Mustefa ¹, Abas Ali ¹, Abdulfeta Abdo ², Abas Lukman ², Abdulfeta Shafti ³, Shukure Uomer ³, Yesufe Badege ⁴, Abdulmejid Mutteba ⁴, Bayesa Tolasa ⁵, Sister Hossae ⁵, Sultan Shukur ⁶, Ebrahim Ahmed ⁷, Abdu Kemal ⁸, Tadeia Erena ⁹

Affiliations + expand
 PMID: 34422161 PMCID: PMC8356931 DOI: 10.11604/pamj.2021.39.38.27648
 Free PMC article

Abstract

Introduction: recent infectious disease outbreaks like COVID-19 highlights the importance of personal protective equipments and competent professionals on public health preparedness and response in health care systems. Hence, understanding availability of personal protective equipments and training status of health professionals is very important to fill the gap of COVID-19 preparedness and response. Therefore, this study was conducted to assess availability and adequacy of personal protective equipments and health professional's training status on COVID-19 in Silte Zone, southern Ethiopia.

Methods: cross sectional study was conducted from August to October 2020 in Silte Zone. First four weredas from 13 were selected randomly. **Systematic sampling** technique was used to select 351 health professionals from 13 health facilities of selected weredas.

Results: overall, only 36.1% of the health professionals have received adequate training on COVID-19. About 30% of the health professionals had taken training on emergency plan of COVID-19, about 33% had been taught on COVID-19 treatment procedures. Majority 80.9% of the respondents indicated that personal protective equipments are inadequately available. Face masks, hand sanitizers and eye goggles were most scarce PPEs.

19

ÉCHANTILLONNAGE

En grappes (clusters – en anglais)

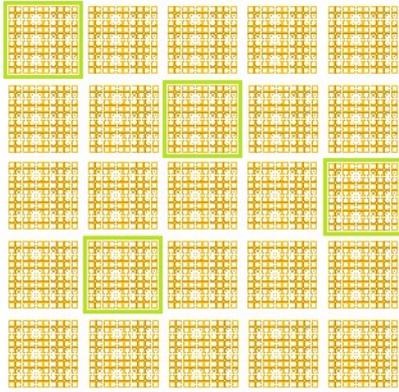
– par l'extraction **aléatoire** des clusters d'une population.

Un cluster peut être:

- un hôpitaux,
- un cabinet d'un médecin,
- un école,
- un ville,
- un famille, ...

L'échantillonnage en grappes peut être

- Non probabiliste si l'extraction n'est
- Pas aléatoire



20

Échantillonnage

- Un échantillonnage en grappes (cluster – en anglais) – pour plusieurs hôpitaux c'est plus bon que l'échantillonnage dans un seul grappe/hôpitaux/cabinet...
- en grappes (cluster – en anglais) – pour plusieurs hôpitaux, et à l'intérieur de chaque hôpital on peut utiliser systématique, aléatoire simple, stratifié

21

ÉCHANTILLONNAGE

En grappes (clusters – en anglais)

► Psychiatry Res. 2020 May;287:112934. doi: 10.1016/j.psychres.2020.112934. Epub 2020 Mar 20.

The psychological impact of the COVID-19 epidemic on college students in China

Wenjun Cao ¹, Ziwei Fang ², Guoqiang Hou ³, Mei Han ², Xinrong Xu ², Jiaxin Dong ², Jianzhong Zheng ⁴

Affiliations + expand

PMID: 32229390 PMCID: PMC7102633 DOI: 10.1016/j.psychres.2020.112934

[Free PMC article](#)

Abstract

A COVID-19 epidemic has been spreading in China and other parts of the world since December 2019. The epidemic has brought not only the risk of death from infection but also unbearable psychological pressure. We sampled college students from Changzhi medical college by using **cluster sampling**. They responded to a questionnaire packet that included the 7-item Generalized Anxiety Disorder Scale (GAD-7) and those inquiring the participants' basic information. We received 7,143 responses. Results indicated that 0.9% of the respondents were experiencing severe anxiety, 2.7% moderate anxiety, and 21.3% mild anxiety. Moreover, living in urban areas (OR = 0.810, 95% CI = 0.709 - 0.925), family income stability (OR = 0.726, 95% CI = 0.645 - 0.817) and living with parents (OR = 0.752, 95% CI = 0.596 - 0.950) were protective factors against anxiety. Moreover, having relatives or acquaintances infected with COVID-19 was a risk factor for increasing the anxiety of college students (OR = 3.007, 95% CI = 2.377 - 3.804). Results of correlation analysis indicated that economic effects, and effects on daily life, as well as delays in academic activities, were positively associated with anxiety symptoms ($P < .001$). However, social support was negatively correlated with the level of anxiety ($P < .001$). It is suggested that the mental health of college students should be monitored during epidemics.

22

ÉCHANTILLONNAGE

Échantillonnage non probabiliste

Le plus utilise dans les études clinique

- pratique/convenable

– en utilisant les plus convenables sources des sujets, en fonction de leur relative facilité d'accès: volontaires parmi les médecins/étudiants; ou parmi des amis, famille, collègues d'un médecin, ou des patients.

– ca pose des problèmes – les volontaires, et les proches sont parfois différent du reste de la population,

- une étude dans un hôpital sans préciser que c'est une collecte consécutive des sujets, est plus probable un échantillonnage pratique

- consécutive

– le meilleur échantillonnage non probabiliste

– on prend toutes les patients qui viennent avec les critères d'intérêt pour l'étude

23

ÉCHANTILLONNAGE

Échantillonnage non probabiliste

> BMC Public Health, 2020 Sep 4;20(1):1351. doi:10.1186/s12889-020-09429-3.

Seizure in patients with COVID-19

Amir Emami ¹, Nima Fadakar ², Ali Akbari ³, Mehrzad Lotfi ⁴, Mohsen Farazdaghi ², Fatemeh Javanmardi ¹, Tahereh Rezaei ⁵, Ali A Asadi-Pooya ^{6, 7}

Affiliations + expand
PMID: 32949289 PMCID: PMC7501768 DOI: 10.1007/s10072-020-04731-9

Free PMC article

Abstract

Objective: The purpose of the current study was to collect the data on the occurrence of seizures in patients with COVID-19 and to clarify the circumstances of the occurrence of seizures in these patients.

Methods: All consecutive patients who referred to healthcare facilities anywhere in Fars province (located in South Iran with a population of 4.851 million people) from February 19 until June 2, 2020, and had confirmed COVID-19 by positive result on polymerase chain reaction testing and seizure were included.

Results: During the study period, 6,147 people had confirmed COVID-19 in Fars province, Iran; 110 people died from the illness (case fatality rate 1.79%). During this time period, five people had seizures (seizure rate 0.08%), in four patients, seizure was one of the presenting manifestations, and in one person, it happened during the course of hospital admission. Two patients had status epilepticus. All patients experienced hypoxemia and four of them needed respirator. Two patients had related metabolic derangements and one had cerebrospinal fluid (CSF) lymphocytic pleocytosis. Brain imaging was abnormal in three patients. Four patients died.

Conclusion: New-onset seizures in critically ill patients with COVID-19 should be considered as acute symptomatic seizures and the treating physician should try to determine the etiology of the seizure and manage the cause immediately and appropriately. Detailed clinical, neurological, imaging, and electrophysiological investigations and attempts to isolate SARS-CoV-2 from CSF may clarify the role played by this virus in causing seizures in these patients.

Abstract

Background: COVID-19 restrictions such as the closure of schools and parks, and the cancellation of youth sports and activity classes around the United States may prevent children from achieving recommended levels of physical activity (PA) and sedentary behavior (SB). This study examined the effects of the COVID-19 pandemic on PA and sedentary behavior (SB) in U.S. children.

Method: Parents and legal guardians of U.S. children (ages 5-13) were recruited through convenience sampling and completed an online survey between April 25-May 16, 2020. Measures included an assessment of their child's previous day PA and SB by indicating time spent in 11 common types of PA and 12 common types of SB for children. Parents also reported perceived changes in levels of PA and SB between the pre-COVID-19 (February 2020) and early-COVID-19 (April-May 2020) periods. Additionally, parents reported locations (e.g., home/garage, parks/trails, gyms/fitness centers) where their children had performed PA and their children's use of remote/streaming services for PA.

Results: From parent reports, children (N = 211) (53% female, 13% Hispanic, $M_{age} = 8.73$ [SD = 2.58] years) represented 35 states and the District of Columbia. The most common physical activities during the early-COVID-19 period were free play/unstructured activity (e.g., running around, tag) (90% of children) and going for a walk (55% of children). Children engaged in about 90 min of school-related sitting and over 8 h of leisure-related sitting a day. Parents of older children (ages 9-13) vs. younger children (ages 5-8) perceived greater decreases in PA and greater increases in SB from the pre- to early-COVID-19 periods. Children were more likely to perform PA at home indoors or on neighborhood streets during the early- vs. pre-COVID-19 periods. About a third of children used remote/streaming services for activity classes and lessons during the early-COVID-19 period.

Conclusion: Short-term changes in PA and SB in reaction to COVID-19 may become permanently entrenched, leading to increased risk of obesity, diabetes, and cardiovascular disease in children. Programmatic and policy strategies should be geared towards promoting PA and reducing SB over the next 12 months.

24

ÉCHANTILLONNAGE

Échantillonnage non probabiliste

Quand vous ne pouvez pas identifier quel type d'échantillonnage a été réalisé dans une étude (section Matériel et méthodes) clinique (dans des hôpitaux), il y a plusieurs possibilités:

- Soit l'étude a utilisé un échantillonnage pratique (convenable), et les auteurs ont caché cette information (ou parfois ils ont oublié, ou ils ne savent pas que cette information doit être écrite – surtout dans des articles plus anciennes) depuis l'article, parce que ils savent que méthodologiquement n'est pas la meilleure technique (variante plus probable)
- Soit l'étude a utilisé un échantillonnage consécutif (idéal dans des études cliniques), mais les auteurs ont oublié de le pré Materials and Methods

Exemple: ici on peut pas identifier la technique d'échantillonnage => on pense que il y a une échantillonnage plutôt pratique

Subjects

The study was conducted on 150 T2DM patients diagnosed as per American Diabetes Association criteria (age 44.12 ± 8.5 year, M:F ratio 57.33 vs. 42.67). The patients with secondary diabetes mellitus (DM), other endocrinopathy, on drugs such as steroids, Pioglitazone, Metformin and with conditions that could influence insulin resistance, IHD, infections were excluded from the study. Also, T2DM patients who were directly exposed to chemicals or having cancer, renal failure or HIV were excluded. 150 healthy normal glucose tolerant controls (age 44.1 ± 10.4 years, M:F ratio 55.33 vs. 44.67) were studied as controls. The institutional ethics committee of

Jain J, Gupta N, Mathur R, Nimesh S, Mathur SK. A Study on Impact of BPA in the Adipose Tissue Dysfunction (Adiposopathy) in Asian Indian Type 2 Diabetes Mellitus Subjects. Indian J Clin Biochem. 2020 Oct;35(4):451-457. doi: 10.1007/s12291-019-00843-y. Epub 2019 Aug 3. PMID: 33013015; PMCID: PMC7502640.

25

Paramètres d'un population et statistiques

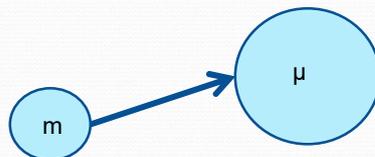
Un paramètre d'une population

– indice synthétique caractérisant la **population entière** (moyenne/ fréquence) Notation – lettres grecques

Une statistique associées a un échantillon

– indices synthétiques caractérisant un **échantillon** (moyenne/ fréquence) Notation – lettres latines

Les statistiques estiment les paramètres d'une population.



26

Variables statistiques

La Caractéristique ou la Variable - un attribut d'une population P

Types de Variables:

– qualitative

• dichotomique



• nominale



• ordinale



27

Variables statistiques

La Caractéristique ou la Variable - un attribut d'une population P

Types de Variables:

– quantitative

• discrète: 1, 2, 3 ...

• continue: 63,5 kg; 82,7 kg

– de survie



28

Variables statistiques

Table 1

Baseline characteristics of patients with SARS-CoV-2 infection.

Characteristic	COVID-19 patients			P value
	Total (N = 80)	FPV (N = 35)	LPV/RTV (N = 45)	
Age, median (IQR)	47 (35.75–61)	43 (35.5–59)	49 (36–61)	0.61
Age subgroup				
15–44	36 (45.0%)	18 (51.4%)	18 (40.0%)	–
45–64	33 (41.3%)	13 (37.1%)	20 (44.4%)	–
≥ 65	11 (13.7%)	4 (11.4%)	7 (15.6%)	0.47
Male	35 (43.8%)	14 (40.0%)	21 (46.7%)	0.55
BMI, median (IQR)	22.9 (16.2–31.6)	22.7 (16.2–31.6)	23.1 (16.4–28.4)	0.51
Epidemiology				
History of visiting Wuhan City	46 (57.5%)	20 (57.1%)	26 (57.8%)	–
Not been to Wuhan City	34 (42.5%)	15 (20.0%)	19 (17.8%)	0.95
Onset symptoms				
Fever	59 (73.8%)	22 (62.9%)	37 (82.2%)	0.11
Cough	22 (27.5%)	12 (34.3%)	10 (22.2%)	0.23
Headache/myalgia	8 (10.0%)	3 (8.6%)	5 (11.1%)	1.00
Chill	1 (1.3%)	0 (0%)	1 (2.2%)	1.00
Diarrhea	1 (1.3%)	1 (2.9%)	0 (0%)	0.44
Stuffy nose/sore throat	8 (10.0%)	6 (17.1%)	2 (4.4%)	0.13
Laboratory test, median (IQR)				
WBC ($\times 10^9 L^{-1}$)	6.0 (3.5–5.2)	8.1 (3.8–6.6)	4.3 (3.4–4.9)	0.21
Neutrophils ($\times 10^9 L^{-1}$)	2.8 (2.1–3.4)	3.0 (2.1–3.7)	2.6 (2.1–3.1)	0.43
Lymphocyte ($\times 10^9 L^{-1}$)	1.3 (0.9–1.6)	1.5 (1.0–1.8)	1.2 (0.9–1.4)	0.06
ALT (U L ⁻¹)	22.2 (15.0–26.3)	21.6 (15.0–24.0)	22.6 (15.5–27.0)	0.54

Cai Q, Yang M, Liu D, Chen J, Shu D, Xia J, Liao X, Gu Y, Cai Q, Yang Y, Shen C, Li X, Peng L, Huang D, Zhang J, Zhang S, Wang F, Liu J, Chen L, Chen S, Wang Z, Zhang Z, Cao R, Zhong W, Liu Y, Liu L. Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study. Engineering (Beijing). 2020 Oct;6(10):1192-1198. doi: 10.1016/j.eng.2020.03.007. Epub 2020 Mar 18. PMID: 32346491; PMCID: PMC7185795.

29

Variables statistiques

On peut transformer une variable **quantitative** dans une variable **qualitative**

- ca va déterminer une **perte d'information**

- Ex. Variable age quantitative peut être transformée dans une variable dichotomique: Age ≥ 50 : oui; Age < 50 : non

Ou dans plusieurs catégories, voir a droit

On **ne peut pas transformer** une variable **qualitative** dans une variable **quantitative**

Certaines variables qualitatives sont appelée avec des nombres, mais ca ne le fait pas des variables quantitatives – la classe NYHA de l'insuffisance cardiaque I, II, III, IV, ou le stade d'un cancer I, II, III, IV. On ne peut pas faire des calculs arithmétiques, ou des statistiques quantitatives (moyenne) sur ces variables.

Characteristic

Age, median (IQR)

Age subgroup

15–44

45–64

≥ 65

30

Variables statistiques

- **Variable** (Caractéristique) = **attribut d'une population**
- **Donnée** = la valeur d'une variable pour un certain sujet

Variable : tabagisme;

Donnée: réponse de type Oui / Non
(ou le nombre de cigarettes fumées/jour)

Variable : la cause de la mort des bébés

Donnée: la malformation congénitale, l' asphyxie, ...

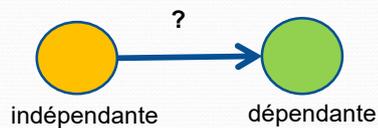
31

Types d'analyse statistique

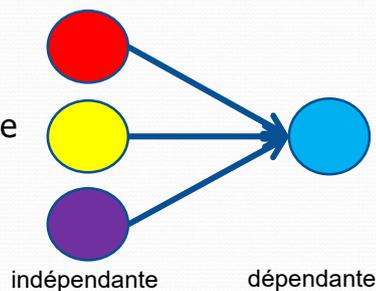
- Analyse univariée



- Analyse bivariée



- Analyse multivariée



32

Types d'analyse statistique

Analyse univariée
(Statistique descriptive)
On décrit des variables
Soit l'âge, soit la fièvre ...

Ici il y a plusieurs analyses bivariées, parce qu'on évalue la relation entre 2 variable: traitement (soit FPV ou LTV/RTV) et Age, puis, traitement et Sexe, puis traitement et

Table 1
Baseline characteristics of patients with SARS-CoV-2 infection.

Characteristic	COVID-19 patients			P value
	Total (N = 80)	FPV (N = 35)	LPV/RTV (N = 45)	
Age, median (IQR)	47 (35.75-61)	43 (35.5-59)	49 (36-61)	0.61
Age subgroup				
15-44	36 (45.0%)	18 (51.4%)	18 (40.0%)	—
45-64	33 (41.3%)	13 (37.1%)	20 (44.4%)	—
≥ 65	11 (13.7%)	4 (11.4%)	7 (15.6%)	0.47
Male	35 (43.8%)	14 (40.0%)	21 (46.7%)	0.55
BMI, median (IQR)	22.9 (16.2-31.6)	22.7 (16.2-31.6)	23.1 (16.4-28.4)	0.51
Epidemiology				
History of visiting Wuhan City	46 (57.5%)	20 (57.1%)	26 (57.8%)	—
Not been to Wuhan City	34 (42.5%)	15 (20.0%)	19 (17.8%)	0.95
Onset symptoms				
Fever	59 (73.8%)	22 (62.9%)	37 (82.2%)	0.11
Cough	22 (27.5%)	12 (34.3%)	10 (22.2%)	0.23
Headache/myalgia	8 (10.0%)	3 (8.6%)	5 (11.1%)	1.00
Chill	1 (1.3%)	0 (0%)	1 (2.2%)	1.00
Diarrhea	1 (1.3%)	1 (2.9%)	0 (0%)	0.44
Stuffy nose/sore throat	8 (10.0%)	6 (17.1%)	2 (4.4%)	0.13

Cai Q, Yang M, Liu D, Chen J, Shu D, Xia J, Liao X, Gu Y, Cai Q, Yang Y, Shen C, Li X, Peng L, Huang D, Zhang J, Zhang S, Wang F, Liu J, Chen L, Chen S, Wang Z, Zhang Z, Cao R, Zhong W, Liu Y, Liu L. Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study. Engineering (Beijing). 2020 Oct;6(10):1192-1198. doi: 10.1016/j.eng.2020.03.007. Epub 2020 Mar 18. PMID: 32346491; PMCID: PMC7185795.

33

Types d'analyse statistique

Ici il y a une **analyse multivariee (on observe multivariate regression model)**, qui évalue la relation entre 9 variable: le temps jusqu'à la survie en fonction de l'âge (Age), maladie cardiaque (CHD), maladie vasculaire cérébrale (CVD), dyspnea, procalcitonin (PCT), aspartate aminotransferase (AST), la bilirubine totale (TBIL), et la Creatinine)

Variables	Level	HR	95% CI	P value
Age, y	65-74 vs <65	3.43	1.24-9.5	.018
Age, y	≥75 vs <65	7.86	2.44-25.35	< .001
CHD	Yes vs No	4.28	1.14-16.13	.032
CVD	Yes vs No	3.1	1.07-8.94	.037
Dyspnea	Yes vs No	3.96	1.42-11	.008
PCT, ng/mL	>0.5 vs ≤0.5	8.72	3.42-22.28	< .001
AST, U/L	>40 vs ≤40	2.2	1.1-6.73	.03
TBIL, μmol/L	≥17.1 vs <17.1	2.25	0.94-5.35	.068
Creatinine, μmol/L	≥133 vs <133	2.87	0.84-9.83	.093

Figure 2 – Risk factor of the fatal outcome in the multivariate Cox proportional hazards regression model. The figure presents the HRs and the 95% CIs associated with the end point. AST = aspartate aminotransferase; CHD = coronary heart disease; CVD = cerebrovascular disease; HR = hazard ratio; PCT = procalcitonin; TBIL = total bilirubin.

Chen R, Liang W, Jiang M, Guan W, Zhan C, Wang T, Tang C, Sang L, Liu J, Ni Z, Hu Y, Liu L, Shan H, Lei C, Peng Y, Wei L, Liu Y, Hu Y, Peng P, Wang J, Liu J, Chen Z, Li G, Zheng Z, Qiu S, Luo J, Ye C, Zhu S, Liu X, Cheng L, Ye F, Zheng J, Zhang N, Li Y, He J, Li S, Zhong N; Medical Treatment Expert Group for COVID-19. Risk Factors of Fatal Outcome in Hospitalized Subjects With Coronavirus Disease 2019 From a Nationwide Analysis in China. Chest. 2020 Jul;158(1):97-105. doi: 10.1016/j.chest.2020.04.010. Epub 2020 Apr 15. PMID: 32304772; PMCID: PMC7158802.

34

Exemples des questions pour l'examen

1) Lesquelles des réponses suivantes sont correctes :

- a) la variable longueur d'os cubitus avec les suivantes catégories: court, moyen, long est une variable qualitative nominale
 -> parce que les valeurs de la variable (court, moyen, long) ne sont pas des nombres, mais du texte, elle est qualitative; parce que il y a plus de 2 catégories (3 en effet: court, moyen, long), elle n'est pas dichotomique, donc elle peut être soit nominale soit ordinale; entre les catégories on peut voir une ordre (une est plus grande que l'autre), donc elle est ordinale; au final est une variable qualitative ordinale. => faux
- b) le temps passée entre le diagnostic d'une maladie et l'apparition d'une métastase est une variable de survie
 -> le temps passée entre une moment donnée (ici le diagnostique), et l' apparition d'un évènement (soit positive, soit négative – ici négative: métastase), et une variable de survie. => vraie
- c) les os du carpe (scaphoïde, trapèze, pisiforme) est une variable qualitative ordinale
 -> parce que les valeurs de la variable (scaphoïde, trapèze, pisiforme) ne sont pas des nombres, mais du texte,, elle est qualitative; parce que il y a plus de 2 catégories (3 en effet: scaphoïde, trapèze, pisiforme), elle n'est pas dichotomique, donc elle peut être soit nominale soit ordinale; entre les catégories on ne peut pas avoir voir une ordre (on ne peut pas hiérarchiser les catégories), donc elle est nominale; au final est une variable qualitative nominale. => faux
- d) le diamètre du ulna en millimètres est une variable quantitative continue
 -> parce que les valeurs de la variable (diamètre en millimètre) sont des nombres (et aussi, c'est possible de calculer une moyenne sur des valeurs de ce type), elle est quantitative; parce que on peut mesurer les valeurs avec une précision qui nous donne des valeurs avec décimales, (même si on n'écrit pas les décimales, seulement s'ils sont possibles, en théorie), la variable est continue; au final est une variable quantitative continue. => vraie
- e) épaisseur de la paroi ventriculaire avec les suivantes catégories : mince, moyen, épais est une variable qualitative continue
 -> parce que les valeurs de la variable (mince, moyen, épais) ne sont pas des nombres, mais du texte, elle est qualitative; parce que il y a plus de 2 catégories (3 en effet: court, moyen, long), elle n'est pas dichotomique, donc elle peut être soit nominale soit ordinale; entre les catégories on peut voir une ordre (une est plus grande que l'autre), donc elle est ordinale; au final est une variable qualitative ordinale; en plus il n'existe pas le type de variable qualitative continue => faux

Réponse: b, d

35

Exemples des questions pour l'examen

2) Lesquelles des réponses suivantes sont correctes :

- a) la variable longueur d'os cubitus en centimètres est une variable quantitative discrète
 -> parce que les valeurs de la variable (longueur d'os cubitus en centimètres) sont des nombres (et aussi, c'est possible de calculer une moyenne sur des valeurs de ce type), elle est quantitative; parce que on peut mesurer les valeurs avec une précision qui nous donne des valeurs avec décimales, (même si on n'écrit pas les décimales, seulement s'ils sont possibles, en théorie), la variable est continue; au final est une variable quantitative continue; quantitative discrète signifie – seulement des valeurs entières (ex. nombre des enfants dans une famille) => faux
- b) le temps passée entre le diagnostic d'une maladie et l'apparition d'un symptôme (la malaise) est une variable quantitative
 -> le temps passée entre une moment donnée (ici le diagnostique), et l' apparition d'un évènement (soit positive, soit négative – ici négative: la malaise), et une variable de survie. => faux
- c) les os du carpe (scaphoïde, trapèze, pisiforme) est une variable qualitative nominale
 -> parce que les valeurs de la variable (scaphoïde, trapèze, pisiforme) ne sont pas des nombres, mais du texte, elle est qualitative; parce que il y a plus de 2 catégories (3 en effet: scaphoïde, trapèze, pisiforme), elle n'est pas dichotomique, donc elle peut être soit nominale soit ordinale; entre les catégories on ne peut pas avoir voir une ordre (on ne peut pas hiérarchiser les catégories), donc elle est nominale; au final est une variable qualitative nominale. => vraie
- d) le poids du cœur après l'autopsie en grammes est une variable qualitative ordinale
 -> parce que les valeurs de la variable (le poids du cœur en grammes) sont des nombres (et aussi, c'est possible de calculer une moyenne sur des valeurs de ce type), elle est quantitative; parce que on peut mesurer les valeurs avec une précision qui nous donne des valeurs avec décimales, (même si on n'écrit pas les décimales, seulement s'ils sont possibles, en théorie), la variable est continue; au final est une variable quantitative continue. => faux
- e) le diamètre d'aorte abdominale avec les suivantes catégories : mince, moyen, épais est une variable qualitative ordinale
 -> parce que les valeurs de la variable (mince, moyen, épais) ne sont pas des nombres, mais du texte, elle est qualitative; parce que il y a plus de 2 catégories (3 en effet: court, moyen, long), elle n'est pas dichotomique, donc elle peut être soit nominale soit ordinale; entre les catégories on peut voir une ordre (une est plus grande que l'autre), donc elle est ordinale; au final est une variable qualitative ordinale. => vraie

Réponse: c, e

36

Exemples des questions pour l'examen

3*) Une échantillonnage a été réalisé parmi des étudiants en réalisant un ensemble des étudiants volontaires. Laquelle des réponses suivantes représente la méthode d'échantillonnage utilisée:

- a) aléatoire simple
- b) stratifié
- c) systématique
- d) pratique
- e) consécutive

-> parce que les sujets de l'étude sont des volontaires, ça indique une échantillonnage pratique, un sou type d'échantillonnage non probabiliste

Réponse: d

4) Une échantillonnage a été réalisé parmi des patients avec hypertension artérielle. Tous les patients avec ce diagnostic dans un hôpital pendant une période d'un mois ont été introduit dans l'étude. Lesquelles des réponses suivantes représente la méthode d'échantillonnage utilisée:

- a) aléatoire simple
- b) probabiliste
- c) non probabiliste
- d) pratique
- e) consécutive

-> parce que toutes les sujets avec le diagnostic ont été introduits dans l'étude, ça suggère que ils ont pris toutes les patients venu d'une manière consécutive dans l'hôpital, ça indique une échantillonnage consécutive, un sou type d'échantillonnage non probabiliste

Réponse: c, e

37

Exemples des questions pour l'examen

5) Une échantillonnage a été réalisé parmi des patients avec hypertension artérielle. Tous les patients avec ce diagnostic dans 5 hôpitaux (choisis d'une manière aléatoire parmi toutes les hôpitaux dans la pays) pendant une période d'un mois ont été introduits dans l'étude. Lesquelles des réponses suivantes représente la méthode d'échantillonnage pour la chois des hôpitaux:

- a) aléatoire simple
- b) probabiliste
- c) systématique
- d) stratifié
- e) en grappes (cluster)

Réponse: b, e

-> parce que la chois d'échantillonnage initiale, n'est pas au niveau des sujets, mais au niveau des hôpitaux, ça indique une échantillonnage en grappes (en clusters). Parce que les hôpitaux ont été choisi par hasard depuis une ensemble des hôpitaux, ça indique une échantillonnage en grappes, un sou type d'échantillonnage probabiliste;

L'échantillonnage en grappes peut être non probabiliste si la chois des hôpitaux est par convenance, pratique.

Au niveau d'un grappe dans cette question l'échantillonnage est non probabiliste, parce que, dans chaque hôpital ils ont pris toutes les sujets (donc toutes les sujets ont venu d'une manière consécutive)

38

Exemples des questions pour l'examen

6) Lesquelles des réponses suivantes sont correctes :

- a) Si on transforme la variable âge (années) dans des tranches d'âge ça va déterminer une perte d'information
 -> Imaginez la variable âge avec deux tranches d'âge (0 – 50 ans, et 50 – 100 ans), si un sujet est dans la première tranche d'âge il peut avoir toute valeur entre 0 et 50. C'est trop imprécis! Même si on a plusieurs tranches d'âge 3, 4, on va avoir une perte d'information. => vrai
- b) On peut transformer la variable degré de douleur (avec les valeurs: sans, faible, modérée, forte) dans une variable quantitative
 -> Imaginez qu'on va associer des valeurs numériques pour les catégories de la variable 1 pour sans, 2 pour faible, 3 pour modérée, 4 pour forte. La problématique est que ça ne préserve pas l'ordre, mais pas la magnitude. Clairement faible n'est pas deux fois sans douleur. Aussi, forte n'est pas deux fois la douleur faible. Donc c'est impossible de faire cette transformation sans fausser la réalité. => faux
- c) La variable stade d'un cancer de foie (avec les valeurs: I, II, III, IV) est une variable qualitative
 -> une stadialisation simplifiée possible pour le cancer de foie de stade I - pour une petite tumeur à l'intérieur du foie, II - une tumeur plus grande, III - plusieurs tumeurs dans le foie, IV - n'importe la dimension de la tumeur, mais la tumeur peut se trouver dans des ganglions, ou métastase dans le poumon ou dans un autre organe. Clairement ils sont des catégories non numériques (le stade III n'est pas trois quarts du stade IV, ils contiennent des concepts différents, soit de dimension, nombre des tumeurs, place dans des différents organes), même si on associe des nombres à ces catégories. => vrai
- d) On peut calculer la moyenne arithmétique de la variable stade du cancer de foie (avec les valeurs: I, II, III, IV)
 -> une stadialisation simplifiée possible pour le cancer de foie de stade I - pour une petite tumeur à l'intérieur du foie, II - une tumeur plus grande, III - plusieurs tumeurs dans le foie, IV - n'importe la dimension de la tumeur, mais la tumeur peut se trouver dans des ganglions, ou métastase dans le poumon ou dans un autre organe. Clairement le stade IV n'est pas deux fois le stade II. Donc on ne peut pas le transformer dans une variable quantitative. Il n'y a pas de relations mathématiques proportionnelles entre les stades, donc on ne peut pas faire ce calcul => faux
- e) Si on transforme la variable stade de l'insuffisance cardiaque (avec les valeurs: I, II, III, IV) dans une variable quantitative, on va déterminer une augmentation d'information -> on ne peut pas le transformer parce que il est qualitative, en plus voir les points antérieurs

Réponse: a, c

39

Exemples des questions pour l'examen

7) Regardez le tableau suivant depuis un article scientifique médical. Il compare deux traitements pour des sujets avec diabète. Répondez aux questions sur le diapositive suivante

Table 1. Demographic characteristics of participants.

Variable	Metformin (N = 43)	Insulin (N = 40)	p-value
Age-yr (mean ±SD)	33.51 ±4.67	33.10±4.56	0.686
Parity (no. (%))			
Nulliparous	7 (16.3)	8 (20.0)	0.897
P1–P4	33 (76.7)	29 (72.5)	
≥P5	3 (7.0)	3 (7.5)	
Marital Status (no. (%))			
Married	43 (100.0)	39 (97.5)	0.482
Single/Divorced	0 (0.0)	1 (2.5)	
BMI-kg/m ²	33.47±6.95	32.61±6.21	0.56
Weight at enrolment (no. (%))			
<90kg	24 (57.1)	28 (70.0)	0.227
≥90kg	18 (42.9)	12 (30.0)	
Gestational Age (weeks) [median, inter-quartile range (25–75)]	28, (26–29)	26, (23–28)	0.017
Classification of diabetes (no. (%))			
GDM	32(74.4)	23(57.5)	0.103
T2DM	11(25.6)	17(42.5)	
Co-morbid condition in pregnancy (no. (%))			
Essential Hypertension	6(14.0)	5(12.5)	1.000
Sickle cell disease	1 (2.3)	0 (0.0)	1.000
Others*	3 (7.0)	3 (7.5)	1.000
Obstetric history (no. (%))			
Miscarriage	12 (27.9)	15 (37.5)	0.351
Stillbirths	3 (7.0)	2 (5.0)	1.000
Early neonatal deaths	1 (2.3)	2 (5.0)	0.607
Big Baby (>4.0kg)	5 (11.6)	2 (5.0)	0.095
Caesarian section	14 (32.6)	14 (35.0)	0.820
Others**	3 (7.0)	3 (7.5)	1.000

*One patient each in the metformin group had malaria, multiple uterine fibroid, anemia as a co-morbid condition while one each for insulin had hepatitis B, G6PD and goiter

**One patient each in the metformin group had intrauterine fetal demise (IUFD), gestational diabetes, and gestational hypertension in their obstetric history while insulin group had one patient each with gestational diabetes, post-partum hemorrhage and IUFD.

doi:10.1371/journal.pone.0125712.t001

Lien vers l'article: <https://doi.org/10.1371/journal.pone.0125712>

Beyuo T, Obied SA, Adjepong-Yamoah KK, Bugyei KA, Oppong SA, Marfoh K. Metformin versus Insulin in the Management of Pre-Gestational Diabetes Mellitus in Pregnancy and Gestational Diabetes Mellitus at the Korle Bu Teaching Hospital: A Randomized Clinical Trial. PLoS One. 2015 May 6;10(5):e0125712.

40

Exemples des questions pour l'examen

7) Lesquelles des réponses suivantes sont correctes :

a) La variable Age yr (year – années) est une variable quantitative continue

-> parce que les valeurs de la variable (âge en années) sont des nombres, elle est quantitative; parce que on peut mesurer les valeurs avec une précision qui nous donne des valeurs avec décimales, (même si on n'écrit pas les décimales, seulement s'ils sont possibles, en théorie), la variable est continue; au final est une variable quantitative continue. => vraie

b) La variable Parity (nombre des naissances) est une variable qualitative ordinale

-> parce que les valeurs de la variable (nuliparus [aucune naissance], P 1 – 4 [1 a 4 naissances], >=5 naissances) ne sont pas des nombres, mais du texte (en effet, une variable quantitative discret – nombre des naissances a été transformée dans une variable qualitative), elle est qualitative; parce que il y a plus de 2 catégories (3 en effet), elle n'est pas dichotomique, donc elle peut être soit nominale soit ordinale; entre les catégories on peut voir une ordre (une est plus grande que l'autre), donc elle est ordinale; au final est une variable qualitative ordinale. => vraie

c) La variable Weight on enrôlement (le poids au début de l'étude) est une variable quantitative continue

-> parce que les valeurs de la variable (<90 kg, et >=90kg) ne sont pas des nombres, mais du texte (en effet, une variable quantitative continue – le poids en kg, a été transformée dans une variable qualitative), elle est qualitative; parce que il y a seulement 2 catégories elle est dichotomique; au final est une variable qualitative dichotomique. => faux

d) La variable Obstetric history (histoire obstétrical) est une variable qualitative nominale

-> l'histoire obstétrical n'est pas une variable. Elle représente une section du tableau, dans lequel il y a plusieurs variables qui sont décrits. => faux

e) La variable Caesarean section (césarienne) est une variable qualitative nominale

-> les tableaux qui présentent les résultats, pour gagner d'espace présentent seulement le nom de la présence d'une signe, symptôme, pathologie, qui peut être présente ou absente. Parce que les valeurs de la variable césarienne (même s'il ne sont pas écrites, elle peuvent être devinées: oui/non, présente/absente, vrai/faux) ne sont pas des nombres, mais du texte, la variable est qualitative; parce que il y a seulement 2 catégories elle est dichotomique; au final est une variable qualitative dichotomique. => faux

Réponse: a, b

41

Exemples des questions pour l'examen

7) Regardez le table suivant depuis une article scientifique médical. Il étudie des patients avec insuffisance rénale. Répondez aux questions sur le diapositive suivante

Variable	Global	eGFR ≥60	eGFR 59–45	eGFR 44–30	eGFR <30	p-value ^a
Female, n (%)	12,520 (44.2)	10,363 (42.8)	1,478(51.2)	568 (53.9)	111 (52.6)	<0.001
Hypertension, n (%)	22,971 (81.0)	19,001 (78.5)	2,737 (94.8)	1,026 (97.4)	207 (98.1)	<0.001
Any cardiovascular disease ^b , n (%)	3,969 (14.0)	3,026 (12.5)	590 (20.4)	286 (27.2)	67 (31.8)	<0.001
Coronary Heart disease, n (%)	3,048 (10.8)	2,326 (9.6)	446 (15.5)	221 (21.0)	55 (26.1)	<0.001
Stroke, n (%)	1,148 (4.1)	854 (3.5)	186 (6.4)	90 (8.5)	18 (8.5)	<0.001
Insulin treatment, n (%)	4,737 (16.7)	3,781 (15.6)	567 (19.6)	302 (28.7)	87 (41.2)	<0.001
Age T2DM (years), mean (SD)	58.8(10.5)	57.5(10.2)	65.6 (9.0)	67.2 (9.7)	65.0(10.1)	<0.001
Age at retinography (years), mean (SD)	65.7(10.7)	64.3(10.4)	73.7(7.9)	76.1 (7.8)	74.6(8.9)	<0.001
Diabetes duration (years), mean (SD)	7.0 (5.2)	6.8 (5.0)	8.2 (5.6)	9.0 (6.0)	9.7(6.1)	<0.001
A1C (%), mean (SD)	7.4 (1.4)	7.4(1.4)	7.2 (1.2)	7.3 (1.4)	7.2 (1.3)	<0.001
Non-HDL cholesterol (mg/dL), mean (SD)	137.3(35.2)	137.8(35.2)	135.4(34.6)	132.5(35.2)	132.5(38.2)	<0.001
Hemoglobin (g/dL), mean (SD)	14.0 (1.5)	14.1 (1.4)	13.3 (1.6)	12.7 (1.6)	12.0 (1.5)	<0.001
SBP (mmHg), mean (SD)	134.9(12.6)	134.8(12.5)	136.1(13.1)	135.8(13.5)	133.9(14.0)	<0.001
DBP (mmHg), mean (SD)	76.4 (8.3)	76.9 (8.2)	74.2 (8.2)	71.9 (8.3)	70.4 (8.3)	<0.001
Heart Rate (bpm), mean (SD)	76.1 (12.1)	76.3 (12.0)	75 (12.5)	74.2 (12.5)	72.5 (12.0)	<0.001

^aP value for comparison of groups by glomerular filtration rate with Pearson's chi-square test for qualitative variables and t-test for quantitative ones.

^bAny cardiovascular disease includes coronary heart disease and/or stroke

A1C, glycated hemoglobin; bpm, beats per minute; DBP, Diastolic Blood Pressure; DR: diabetic retinopathy.

doi:10.1371/journal.pone.0149448.t001

Rodríguez-Poncelas A, Mundet-Tudurí X, Miravet-Jiménez S, Casellas A, Barrot-De la Puente JF, Franch-Nadal J, Coll-de Tuero G. Chronic Kidney Disease and Diabetic Retinopathy in Patients with Type 2 Diabetes. PLoS One. 2016 Feb 17;11(2):e0149448.

Lien vers l'article: <https://doi.org/10.1371/journal.pone.0149448>

42

Exemples des questions pour l'examen

7b) Lesquelles des réponses suivantes sont correctes :

a) La variable Haemoglobine (g/dL) (hémoglobine) est une variable quantitative continue

-> parce que les valeurs de la variable (hémoglobine en g/dL) sont des nombres, elle est quantitative; parce que on peut mesurer les valeurs avec une précision qui nous donne des valeurs avec décimales, (même si on n'écrit pas les décimales, seulement s'ils sont possibles, en théorie), la variable est continue; au final est une variable quantitative continue. => vraie

Une indice pour le fait que la variable est quantitative, est le fait qu'ils on calcule la moyenne - Haemoglobine (g/dL), mean

b) La variable Hypertension (hypertension) est une variable qualitative ordinale

-> les tableaux qui présentent les résultats, pour gagner d'espace présent seulement le nom de la présence d'une signe, symptôme, pathologie, qui peut être présente ou absente. Parce que les valeurs de la variable c hypertension (même s'il ne sont pas écrites, elle peuvent être devinées: oui/non, présente/absente, vrai/faux) ne sont pas des nombres, mais du texte, la variable est qualitative; parce que il y a seulement 2 catégories elle est dichotomique; au final est une variable qualitative dichotomique. => faux

Une indice pour le fait que la variable est qualitative, est le fait qu'ils on calcule le nombre des sujets et le pourcentage des sujets n (%) – Hypertension, n (%)

c) La variable A1c – glycated haemoglobin (%) (pourcentage du hémoglobine glyquée – A1c) est une variable quantitative continue

-> parce que les valeurs de la variable (pourcentage du hémoglobine glyquée – A1c) sont des nombres, elle est quantitative; parce que on peut mesurer les valeurs (ici – les pourcentages) avec une précision qui nous donne des valeurs avec décimales, (même si on n'écrit pas les décimales, seulement s'ils sont possibles, en théorie), la variable est continue; au final est une variable quantitative continue. => vraie

Une indice pour le fait que la variable est quantitative, est le fait qu'ils on calcule la moyenne - A1c – glycated haemoglobin (%), mean – ici ils on calcule la moyenne des pourcentages!

d) La variable Heart rate (bpm) (fréquence cardiaque – en battements par minute) est une variable quantitative

-> parce que les valeurs de la variable (fréquence cardiaque – en battements par minute) sont des nombres, elle est quantitative; C'est difficile de dire s'il est continue ou discrete. Ca dépend des chercheurs. Un étude peut mesurer un seul fois la fréquence cardiaque – dans ce cas, le nombre des battements sera quantitative discrete. Si on les chercheurs ont mesure trois fois la fréquence cardiaque a 5 minute intervalle, et ils ont fait la moyenne, la variable sera quantitative continue, parce que on peut avoir des decimales. La variable est néanmoins quantitative => vraie

Une indice pour le fait que la variable est quantitative, est le fait qu'ils on calcule la moyenne - Heart rate (bpm), mean

Réponse: a, c, d

43

Exemples des questions pour l'examen

8) Regardez le table suivant depuis une article scientifique médical. Il compare des sujets avec et sans diabète qui ont été soumis a une chirurgie cardiaque. Lesquelles des réponses suivantes représente la méthode d'échantillonnage utilisée:

- aléatoire simple
- probabiliste
- non probabiliste
- pratique
- Consécutives

Methods

Patient Selection

Prospectively collected data were extracted from our hospital database (Patient Analysis and Tracking Systems, Dentrite Clinical Systems, London, UK) on consecutive adult cardiac surgical patients operated on between April 1996 and March 2004. The data collection form

-> parce que les sujets avec chirurgie cardiaque consécutifs ont été introduits dans l' étude, ca indique une échantillonnage consécutives, un sou type d'échantillonnage non probabiliste

Réponse: c, e

Ascione R, Rogers CA, Rajakaruna C, Angelini GD. Inadequate blood glucose control is associated with in-hospital mortality and morbidity in diabetic and nondiabetic patients undergoing cardiac surgery. *Circulation*. 2008 Jul 8;118(2):113-23. doi: 10.1161/CIRCULATIONAHA.107.706416.

Lien vers l'article: <https://dx.doi.org/10.1161%2FCIRCULATIONAHA.107.706416>

44

Exemples des questions pour l'examen

9) Regardez le table suivant depuis une article scientifique médical. Il étudie des sujets avec diabete. Lesquelles des réponses suivantes représente la méthode d'échantillonnage utilisée pour la chois des hôpitaux :

- aléatoire simple
- probabiliste
- systématique
- stratifié
- en grappes (cluster)

2. Research design and methods

This study was an observational, cross-sectional, multicenter study conducted between December 2008 and December 2010 in 28 public clinics of the secondary and tertiary care level located in 20 cities in four Brazilian geographic regions (north/northeast, mid-west, southeast and south). All patients received health care from the National Brazilian Health Care System (NBHCS). To be eligible, the participant centers had to have a diabetes clinic with at least one endocrinologist. Overall, 31 public clinics were identified and invited to participate. A short questionnaire was sent to the clinics to acquire information on the characteristics of the patient population, especially the number of T1DM patients under routine care; three clinics did not send back the answered questionnaire until the implementation of the study and were not included. The number of patients enrolled in each region

Réponse: b, e

-> parce que la choisis d'échantillonnage initiale, n'est pas au niveau des sujets, mais au niveau des hôpitaux, ca indique une échantillonnage en grappes (en clusters). Parce que la choisis des hôpitaux n'a pas été décrite comme une choisis aléatoire depuis une ensemble des hôpitaux, ca indique plutôt un échantillonnage non probabiliste parce que la choisis des hôpitaux est par convenance, pratique.

Gomes MB, Coral M, Cobas RA, Dib SA, Canani LH, Nery M, de Freitas MC, Faria M, Felício JS, da Silva SC, Pedrosa H, Costa e Forti A, Rea RR, Pires AC, Montenegro Junior R, Oliveira JE, Rassi N, Negrato CA. Prevalence of adults with type 1 diabetes who meet the goals of care in daily clinical practice: a nationwide multicenter study in Brazil. *Diabetes Res Clin Pract.* 2012 Jul;97(1):63-70. doi: 10.1016/j.diabres.2012.02.008.

Lien vers l'article: <https://dx.doi.org/10.1016/j.diabres.2012.02.008>

45

Exemples des questions pour l'examen

10) Regardez le table suivant depuis une article scientifique médical. Il compare des sujets avec maladie rénale chronique dans des stages différents. Lesquelles des réponses suivantes représente la méthode d'échantillonnage utilisée:

- ce n'est pas sur
- probabiliste
- plus probable échantillonnage pratique/convenable
- moins probable échantillonnage consécutif
- en grappes (cluster)

Subjects and Methods

A total of 80 patients with CKD aged over 18, both male and female, were enrolled in this cross-sectional cohort study and divided into two groups according to the CKD stage. The first group comprised 40 patients with CKD stage V that had been treated with hemodialysis for at least three months (HD group). The second group comprised 40 patients with CKD stages III and IV (control group).

Réponse: a, c, d

-> Quand vous ne pouvez identifier quel type d'échantillonnage a été réalisée dans une étude (section Matériel et méthodes) clinique (dans des hôpitaux), il y a plusieurs possibilités:

Soit l'étude a utilise une échantillonnage pratique (convenable), et les auteurs ont cache cette information (ou parfois ils ont oublié, ou ils ne savent pas que cette information doit être écrite – surtout dans des articles plus anciennes) depuis l'article, parce que ils savent que méthodologiquement n'est pas la meilleure technique (variante plus probable) => réponse c) est correcte

Soit l'étude a utilise une échantillonnage consécutive (idéal dans des études cliniques), mais les auteurs ont oublié de le préciser. (variante moins probable) => réponse d) est correcte

On réalité on ne peut pas être sur, parce que ce n'est pas écrit. => réponse a) est correcte

Kopić V, Barbić J, Petrović S, Šahinović I, Mihaljević D, Kopić A, Bošnjak A. PERIODONTAL DISEASE IN DIFFERENT STAGES OF CHRONIC KIDNEY DISEASE. *Acta Clin Croat.* 2019 Dec;58(4):709-715. doi: 10.20471/acc.2019.58.04.18. PMID: 32595256; PMCID: PMC7314301.

Lien vers l'article: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314301/pdf/acc-58-709.pdf>

46

