TESTS STATISTIQUES

1

Plan du cours

- **Comment répondre au questions médicales avec des tests statistiques**
  - **Les hypothèses statistiques**
- **Types de tests statistiques**
  - Echantillons dépendants/ appariées ou indépendants
- **Les étapes d'une test statistique**
  - **le test t pour comparer 2 échantillons indépendants avec variances inégales**
- **Les erreurs d'un test statistique**
- **La relation entre la puissance et autres caractéristiques des tests statistiques**
- **Exemples depuis des articles scientifiques** – avec explications
- **Exemples des questions pour l'examen** – avec explications

2

Comment répondre au questions médicales avec des tests statistiques

- Dans le domaine médical
  - on a beaucoup des questions a répondre
    - L'enalapril est diffèrent de propanolol pour réduire la tension artérielle systolique?
    - Le virus papillome (HPV) est un facteur de risque pour le cancer de poumon?
    - La survie des patients atteint d'un cancer du poumon avec adénopathie (ganglions tuméfiées) est différente de ceux sans adénopathie?

3

Comment répondre au questions médicales avec des tests statistiques

Pour répondre a chaque question:

- On doit
  - réaliser des études
    - Préparer les études (protocole d'étude)
      - **transformer** les **questions médicales** dans **hypothèses** statistiques
      - **identifier** le type des **caractéristiques** /variables d'intérêt
      - **choisir** le type de **test statistique** approprie a la situation et aux données
    - Réalisation effective de l'étude
      - **collecter** les **données**, concernant les variables d'intérêt
      - **analyse** des données en utilisant des tests statistiques.
        - choisir l'hypothèse qui est bonne
      - **répondre** a la **question** médicale

4

Les hypothèses statistiques

- La création des hypothèses:
- Certains **questions médicales** ont deux réponses opposées
  - on force beaucoup des questions dans ce format
- Les réponses correspond aux deux **modèles** possibles de la réalité
- Ce deux modèles sont nommées: **hypothèses**
  - L' hypothèse **nulle**: H_0
    - **Il n'y a pas** un différence statistiquement significative entre 2/>=2 groups (ex. un traitement [ibuprofène vs. placebo] ou un facteur de risque [présent vs. absent]) en ce qui concerne la moyenne/ médiane/ variance/ fréquence … d'une caractéristique (ex résultat du traitement: la fréquence du guérison [oui vs. non] ou la moyenne de la température)
    - **Il n'y a pas** un relation/lien/association/dépendance/corrélation statistiquement significative entre 2 caractéristiques/variables: (ex. un Facteur de risque [présent vs. absent] – une maladie [oui vs. non] , ou un traitement [ibuprofène vs. placebo] – le résultat du traitement (ex: la fréquence du guérison [oui vs. non] ou la moyenne de la température))
  - L' hypothèse **alternative**: H_1 (négation du H_0)
    - **Il y a** un différence statistiquement significative entre 2/>=2 groups en ce qui concerne la moyenne/ médiane/ variance/ fréquence … d'une caractéristique
    - **Il y a** un relation/lien/association/dépendance/corrélation statistiquement significative entre 2 caractéristiques/variables

Les tests statistiques nous permet de faire la chois entre les deux possibilités (H_1 / H_0)

5

Qu'est qu' on compare avec les tests statistiques

- On utilise les tests statistiques pour comparer ce qui se passe dans des populations depuis lesquelles on a extrait les échantillons.
- On ne compare pas des échantillons. On fait une inférence/une généralisation pour savoir qu' est qui se passe dans la population, ou dans la ~ vérité.
- Quand on présente ce qu'on fait avec un test statistique, ou avec des hypothèses des tests statistiques on peut utiliser le nom « groupes », parce que ca se réfère aussi a les échantillons que a les populations correspondantes, ou on peut éviter d'utiliser des mots pour identifier échantillons/groupes.

6

Qu'est qu' on compare avec les tests statistiques

- Pour la vitesse, brévité, pour simplifier, pour être compris par le publique, dans la vie d'un médecin ou même d'un statisticien, on peut dire qu'on compare des échantillons – mais ca est fausse (surtout les échantillons - même s'il le font ils sachent ca, et il sachent qu'ils font des généralisations – ils comparent des populations).
- Vous devez toujours savoir que en réalité on compare des populations.

7

La création des bonnes hypothèses statistiques

- **C'est très important de préciser:**
  - **Ce que vous cherchez**: de faire un **comparaison** ou vous cherchez une **lien / relation / association / corrélation / dépendance**
  - **l'expression** «statistiquement significative »
  - **quel statistique vous comparez**: des moyennes, des variances, des fréquences. Ca doit être précisé dans l'énonce des hypothèses statistiques, et aussi dans l' interprétation du résultat du teste statistique
  - **les groups** a comparer (si il y a)
  - **la caractéristique que vous comparez** (ex. cholesterol), Ou **les noms des variables/caracterstiques** entre lesquelles vous voulez chercher une **lien / relation / association / corrélation / dépendance**

8

Types de tests statistiques

- Selon le **nombre échantillons**
- 2, ou >= 2
- Selon le **type des échantillons**
  - Dépendants (ou appariées) ou indépendants
- Selon le **types de variables**
  - Qualitatives ou quantitatives ou de survie
- Selon les **statistiques comparées**
  - Moyennes, médianes, variances, fréquences

9

Types des tests statistiques

- **Selon la constitution des échantillons**
- Pour **échantillons indépendants** (deux ou plusieurs)
  - les observations sont indépendantes à l'intérieur des groupes et d'un groupe à l'autre
    - Ex. on compare des sujets qui on reçu un traitement hypocholestérolémiant et des sujets qui ont reçu un placebo
- Pour **échantillons appariés/ dépendants** (deux ou plusieurs)
  - la situation des **mesures répétées** sur les mêmes sujets.
    - ex. on mesure la température d'un patient avant et après la prise d'un médicament qui baisse la température: ibuprofène
  - on compare des **jumeaux identiques**
  - on compare des **échantillons appariées** (pour chaque sujet dans un group on trouve un sujet dans l'autre groupe avec caractéristiques similaires)
  - on compare plusieurs observations faites sur certaines parties du corps du même individu, qui conceptuellement ne sont pas différentes – ex: **partie gauche** avec la **partie droite** (ex. un test dermatologique avec une traitement pour diminuer la perte de cheveux). La comparaison ne doit pas être seulement entre la partie gauche et droite. On peut comparer la partie antérieure et postérieure du scalp.
  - on compare **des méthodes** pour **mesurer** une caractéristique (ou deux techniques diagnostiques [ex tension artérielle]) **sur les mêmes sujets**

10

Types des tests statistiques

- **Concernant la constitution des échantillons**
- **Pour toutes les tests statistiques qu'on apprend cet année**, chaque échantillon, n'importe si on compare des échantillons indépendants ou dépendants, **a l'intérieur de chaque échantillon on doit avoir des observations indépendantes**
  - il n'y a pas des observations faites sur des sujets/malades/patients qui sont extrêmes similaires a les autres
    - jumeaux identiques,
    - observations répétées sur les mêmes sujets,
    - mesures sur la main gauche et sur la main droite du même sujet, ...

11

Les étapes dans la réalisation des tests statistiques
(de la vérification d'hypothèses statistiques)

Tous **les test statistiques paramétriques ont 6 étapes**

On va **exemplifier, les étapes, avec le test t** (**Student**)

pour comparer 2 échantillons indépendants avec variances inégales.

- **Étape 1. Formuler** la problème de recherche comme **des hypothèses statistiques**
- **Étape 2. Décider** sur une statistique approprie du test (**paramètre du test**)
- **Étape 3.** Sélecter le niveau de signification pour le test statistique - la **valeur alpha**
- **Étape 4.** Déterminer la valeur critique de la statistique du test, et puis la **région de rejet**
- **Étape 5. Calculer** la valeur de la statistique / **paramètre du test**
- **Étape 6.** La **décision** en fonction de la **région critique**, **ou** la décision avec **p-value**

12

12

Le Test t (Student) des échantillons indépendants

- **Objectif:**
  - **comparaison des moyennes des variables quantitatives continues des 2 échantillons indépendants**
- **Conditions d'application:**
  - observations indépendants dans chaque échantillon (a l'intérieur de l'échantillon il n'y a pas des observations (faites sur des sujets/malades/patients) qui sont extrêmes similaires a les autres (jumeaux identiques, observations répétées sur les mêmes sujets, mesures sur la main gauche et sur la main droite du même sujet, ...))
  - deux échantillons indépendants
  - variable quantitative
  - les données sont normalement distribuées
    - (dans les deux populations)
  - les variances des populations sont inconnues
  - le nombre des observations < 30 ou >=30

13

Le Test t (Student) des échantillons indépendants

L' hypothèse **nulle**: H_0

- **Il n'y a pas de différence** statistiquement significative entre les moyennes des deux groups / populations (depuis lesquelles les deux échantillons ont été extraites)
  - (la différence entre les deux moyennes est égale à zéro)

L' hypothèse **alternative**: H_1 du test bilatéral

- **Il y a une différence** statistiquement significative entre les moyennes des deux groups / populations
  - (la différence entre deux moyennes n'est pas égale à zéro)

14

Problème: test t des échantillons indépendants avec variances inconnues et inégales

On a étudié un échantillon de 13 enfants (différents) depuis des zones urbaines pour lesquels on a obtenu le poids moyenne de naissance de 3314 +/- 603,29 grammes – déviation standard d' échantillonnage.

On a étudié un échantillon de 14 des enfants (différents) depuis des zones rurales pour lesquels on a obtenu le poids moyenne de naissance de 2240 +/- 470,81 grammes – déviation standard d' échantillonnage. Les deux échantillons ont eu des données normale distribuées. Un test F (voir cours suivant) a montre que les variances sont inégales.

Question: est ce qu'on peut affirmer qu'il y a une différence statistiquement significative entre les moyennes du poids a naissance pour les enfants des zones urbaines et les enfants des zones rurales? 15

15

Les étapes dans la vérification d'hypothèses statistiques. Le cas du test t des échantillons indépendants avec variances inconnues et inégales.

- **Étape 1.**
  - **Formuler** la problème de recherche dans la langage des **hypothèses statistiques**:
    - ***Hypothèse nulle H_0 :***
      - il **n'existe pas/ il n'y a pas une différence statistiquement signifiante**
        - entre la moyenne du poids des enfants des zones urbaines (la moyenne du poids des enfants dans la population urbaine des enfants) et la moyenne du poids enfants des zones rurales
    - ***Hypothèse alternative H_1 :***
      - contredit l'hypothèse nulle
      - il **existe/ il y a une différence statistiquement signifiante**
        - entre la moyenne du poids des enfants des zones urbaines (la moyenne du poids des enfants dans la population urbaine des enfants) et la moyenne du poids enfants des zones rurales 16

16

Les étapes dans la vérification d'hypothèses statistiques

- **H_0: $\mu_1 = \mu_2$ –** moyennes egales
-
- **H_1** possibilités:
- **H_1: $\mu_1 \neq \mu_2$ (pour le test non directionnel/ bidirectionnelle/ bilatérale)**
  - **<u>la plus utilise</u> H_1 dans le domaine médical**
  - utilisée quand on ne sait pas avant faire l'étude quel group a la moyenne la plus grande – la situation la plus fréquente
  - **pour ce cours, et les travaux pratiques <u>on va utiliser cette hypothèse</u>**
- H_1: $\mu_1 < \mu_2$ (pour le test unidirectionnel inférieure/unilatérale)
- H_1: $\mu_1 > \mu_2$ (pour le test unidirectionnel supérieure/unilatérale)

17

17

Les étapes dans la vérification d'hypothèses statistiques

- **Étape 2.**
- Décider sur une **statistique approprie** du **test**
  - (**paramètre du test**)
  - exprimant la différence entre les éléments comparées.
  - est une variable aléatoire
  - elle suivent sur l'hypothèse nulle une certain distribution de probabilité.
- Ici le scenario presente une comparaison des <u>moyennes d'une caractéristique quantitative continue</u> sur 2 échantillons, indépendants (il n'y a pas de dépendance entre les sujets des deux comparées), les variances (et implicite la déviation standard/ écart type) ne sont pas connues dans la population, mais on les observe dans les échantillons, et on peut évaluer s'ils sont égales ou inégales dans la population (voir le cours suivant – le test F)
- =>donc on peut utiliser une variable aléatoire qui suit **la lois t de probabilité**.

18

18

Les étapes dans la vérification d'hypothèses statistiques

La statistique du test qui suit une lois t de distribution de probabilité

- **n_1, n_2 = les tailles des échantillons**
- **m_1, m_2= moyennes des échantillons;**
- **S_1^2, S_2^2 = les variances d'échantillonage** (ne sont pas les variances descriptives)

$$t = \frac{m_1 - m_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

19

19

Les étapes dans la vérification d'hypothèses statistiques

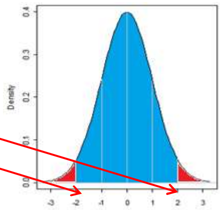
- **Étape 3.**
  - Sélecter le **niveau de signification** pour le test statistique
    - la valeur **alpha**.
- α est une probabilité de rejet incorrect de H_0 quand elle est vrais.
  - Valeurs traditionnelles:
    - α = **0,05**, (5% erreur) ! On va utiliser cet valeur pour toutes les tests dans notres cours et travaux pratiques.
    - α = 0,01,
    - α = 0,001

20

20

Les étapes dans la vérification d'hypothèses statistiques

- **Étape 4.**
  - Déterminer la/les ***valeur/s critique/s (v.c.)*** de la statistique du test
    - trouvée dans le tableau du distribution statistique t, Z, Chi, …
    - en fonction de alpha et du nombre de degrés de liberté (si il faut)
  - on détermine
    - une région *critique* ou **région de *rejet*** *(RR)* **(-∞ , -v.c.]** ∪ **[v.c. , + ∞)**
    - *Si la statistique du test se retrouve dans cet région on va rejeter l' H_0*
    - une région de non rejet RnonR=**(- valeur critique , + valeur critique)**
    - *Si la statistique du test se retrouve dans cet région on ne peut pas rejeter l' H_0*
  - Pour α = **0,05, la statistique t** a la valeur critique 2,06 (voir le cours avec variables aleatoires – la lois t)
    - **test non directionnel (H_1: $\mu_1 \neq \mu_2$):**
      - **RR = (-∞ , -v.c.]** ∪ **[v.c. , + ∞)**
      - **RR = (-∞ , -2,06]** ∪ **[2,06 , + ∞)**
      - **RnonR = (-2,06 , 2,06)**
      - **v.c. – valeur critique**



21

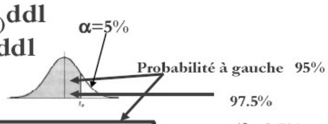
21

Avec la distribution t on trouve la valeur critique a l'aide du alpha, et les dégrées de liberté (d.d.l.)

Situation Unilatérale :$t_{(\alpha,n-1)}$ddl

Situation Bilatérale :$t_{(\alpha/2,n-1)}$ddl

VALEURS des CENTILES
pour la
DISTRIBUTION t de STUDENT
en fonction du nombre ν de degrés de liberté
(aire en grisé = p)



ddl

ν	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$	$t_{0,80}$	$t_{0,75}$	$t_{0,70}$	$t_{0,60}$	$t_{0,55}$
1	63,66	31,82	12,71	6,31	3,08	1,376	1,000	0,727	0,325	0,158
2	9,92	6,96	4,30	2,92	1,89	1,061	0,816	0,617	0,289	0,142
3	5,84	4,54	3,18	2,35	1,64	0,978	0,765	0,584	0,277	0,137
4	4,60	3,75	2,78	2,13	1,53	0,941	0,741	0,569	0,271	0,134
5	4,03	3,36	2,57	2,02	1,48	0,920	0,727	0,559	0,267	0,132
6	3,71	3,14	2,45	1,94	1,44	0,906	0,718	0,553	0,265	0,131
7	3,50	3,00	2,36	1,90	1,42	0,896	0,711	0,549	0,263	0,130
8	3,36	2,90	2,31	1,86	1,40	0,889	0,706	0,546	0,262	0,130
9	3,25	2,82	2,26	1,83	1,38	0,883	0,703	0,543	0,261	0,129
10	3,17	2,76	2,23	1,81	1,37	0,879	0,700	0,542	0,260	0,129
11	3,11	2,72	2,20	1,80	1,36	0,876	0,697	0,540	0,260	0,129
12	3,06	2,68	2,18	1,78	1,36	0,873	0,695	0,539	0,259	0,128
13	3,01	2,65	2,16	1,77	1,35	0,870	0,694	0,538	0,259	0,128
14	2,98	2,62	2,14	1,76	1,34	0,868	0,692	0,537	0,258	0,128
15	2,95	2,60	2,13	1,75	1,34	0,866	0,691	0,536	0,258	0,128
16	2,92	2,58	2,12	1,75	1,34	0,865	0,690	0,535	0,258	0,128
17	2,90	2,57	2,11	1,74	1,33	0,863	0,689	0,534	0,257	0,128
18	2,88	2,55	2,10	1,73	1,33	0,862	0,688	0,534	0,257	0,127
19	2,86	2,54	2,09	1,73	1,33	0,861	0,688	0,533	0,257	0,127
20	2,84	2,53	2,09	1,72	1,32	0,860	0,687	0,533	0,257	0,127
21	2,83	2,52	2,08	1,72	1,32	0,859	0,686	0,532	0,257	0,127
22	2,82	2,51	2,07	1,72	1,32	0,858	0,686	0,532	0,256	0,127
23	2,81	2,50	2,07	1,71	1,32	0,858	0,685	0,532	0,256	0,127
24	2,80	2,49	2,06	1,71	1,32	0,857	0,685	0,531	0,256	0,127
25	2,79	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
26	2,78	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
27	2,77	2,47	2,05	1,70	1,31	0,855	0,684	0,531	0,256	0,127
28	2,76	2,47	2,05	1,70	1,31	0,855	0,683	0,530	0,256	0,127
29	2,76	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
30	2,75	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
40	2,70	2,42	2,02	1,68	1,30	0,851	0,681	0,529	0,255	0,126
60	2,66	2,39	2,00	1,67	1,30	0,848	0,679	0,527	0,254	0,126
120	2,62	2,36	1,98	1,66	1,29	0,845	0,677	0,526	0,254	0,126
∞	2,58	2,33	1,96	1,645	1,28	0,842	0,674	0,524	0,253	0,126

D'après R.A. Fisher et F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, (5e éd.), tableau III, Oliver and Boyd Ltd., Edinburgh.

22

Les étapes dans la vérification d'hypothèses statistiques

- **Étape 5.**
  - **Calculer la valeur de la statistique / paramètre du test**

$$t_0 = \frac{m_1 - m_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{3314 - 2240}{\sqrt{\frac{603,29^2}{13} + \frac{470,81^2}{14}}} = 5,13$$

23

23

Les étapes du test statistique dans la vérification d'hypothèses

Étape 6.

la **décision statistique** en **fonction de la région critique** :

- Si **t_0 est dans RR (région du rejet/critique)** = (-inf , -v.c.] U [v.c. , + inf),
  - (la valeur critique – v.c. vous est donne dans l'énonce pour l'examen écrit, inf = infini)
  - **on rejet H_0 on accepte H_1 et**
  - **En général pour les tests - Il y a une différence/ Il y a une relation –statistiquement significative**
    - Pour le test t - il **existe** **une différence statistiquement significative entre les moyennes des poids a la naissance pour les enfants des zones urbaines et les enfants des zones rurales** .
- Si **t_0 est en dehors de la région du rejet**
  - **on ne peut pas rejeter H_0**
    - **En général pour les tests - On ne peut pas dire qu' Il y a une différence/ Il y a une relation –statistiquement significative**
    - Pour le test t - **On ne peut pas dire** qu'il **existe une différence statistiquement significative entre les moyennes des poids a la naissance pour les enfants des zones urbaines et les enfants des zones rurales** .

24

24

Les étapes du test statistique dans la vérification d'hypothèses

Étape 6.

Pour la problème:

$t_0 = 5{,}13. \in RR = (-\infty , -2{,}06] \cup [2{,}06 , + \infty)$,

- = > on accepte H_1

<u>La conclusion statistique</u>: il existe une différence statistiquement significative entre les moyennes des poids a la naissance pour les enfants des zones urbaines et les enfants des zones rurales .

Les résultats indique de se qui se passe dans les populations!!! Pas dans les échantillons!!!

25

25

Les étapes du test statistique dans la vérification d'hypothèses

- **Étape 6'.**
  - **La décision avec p-value.**

p-value est la probabilité d'obtenir, quand H_0 est supposée vrai (quant on suppose que il n'y a pas une différence dans la population – ici on suppose que les moyennes sont égales dans la population), un résultat pour la statistique du test plus extrême que le résultat observe.

Ici le résultat signifie la différence entre la moyenne du poids (observée et puis généralisée sur la population) du milieu urbain et rural.

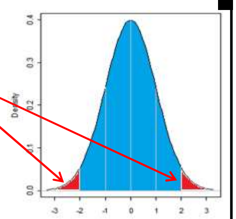
3314 g − 2240 g = 1074 g (la différence observée entre les moyennes)

p-value est l'équivalent du t_0 calculée

= la probabilité qui correspond a des valeurs $|>= t_0|$ si H_0 est vrai

- peut être trouve:
  - dans des tableaux des probabilités
  - avec des logiciels

p



26

Les étapes du test statistique dans la vérification d'hypothèses

- **Étape 6'.**
  - **La décision avec p-value.**
  - Si **p-value ≤ α (=0,05) => on rejet H_0 et on accepte H_1**
    - **En général pour les tests - Il y a une différence/ Il y a une relation –statistiquement significative**
    - Pour le test t - il **existe une différence statistiquement significative entre les moyennes des poids a la naissance pour les enfants des zones urbaines et les enfants des zones rurales .**
  - Si **p-value > α (=0,05) => on ne peut pas rejeter H_0**
    - **En général pour les tests - On ne peut pas dire qu' Il y a une différence/ Il y a une relation –statistiquement significative**
    - Pour le test t - **On ne peut pas dire** qu'il **existe une différence statistiquement significative entre les moyennes des poids a la naissance pour les enfants des zones urbaines et les enfants des zones rurales .**

27

27

Les étapes du test statistique dans la vérification d'hypothèses

- **Étape 6'.**
  - **La décision avec p-value.**

Pour la problème: p-value obtenu a l'aide des tableaux t (voir le cours des variables aléatoires – la lois t)

p-value = 0,0001384522 < 0,05 **=>** rejet H_0 et accepte H_1.

La **conclusion** statistique: il **existe une différence statistiquement significative entre les moyennes des poids a la naissance pour les enfants des zones urbaines et les enfants des zones rurales .**

28

28

Les étapes du test statistique dans la vérification d'hypothèses

- **Étape 6'.**

  Erreurs a éviter!!!
  - L'affirmation « p > 0.05 => on rejeté l' hypothèse alternative" – c'est faux
  - L'affirmation « p > 0.05 => il n'y a pas de différence"
  - L'affirmation « p > 0.05 => l' hypothèse nulle et vraie" – c'est faux – c'est très difficile de prouver l'absence de la différence. Un test statistique plus puissant (avec plus des sujets) peut parfois trouver une différence.

- Les résultats indique de se qui se passe dans les populations!!! Pas dans les échantillons!!!

P-value

Est **calcule** par les **logiciels statistiques** a partir de la valeur calculée du paramètre du test (t, Z, Chi, F…) avec (pour t, Chi, F) ou sans (pour Z) le nombre des degrés de liberté

On peut l'identifier dans des **tableaux des distributions de probabilités** a partir de la valeur calculée du paramètre du test avec (pour t, Chi, F) ou sans (pour Z) le nombre des degrés de liberté

Pour le comprendre:

Imaginez que deux solution de bain de bouche ont le même effet dans la réalité (la réduction du pourcentage des dents cariées). Donc **la différence d** de la réduction du pourcentage des dents cariées est 0 %.

P-value

Pour le comprendre:

On extrait un échantillon et on utilise les deux traitements pour voir la différence. Par la chance (malchance) on observe un différence d=20 % en faveur d'un des deux traitements.

La probabilité d'observer la différence d, ou une différence plus grande que d, si dans la réalité il n y a aucune différence est la valeur p.

Si la probabilité d'observer la différence d est très petite (<0,05/ < 5 %), on peut dire que la différence est en fait plutôt réelle. La différence est peut probable due a la malchance du chois des sujets dans l' échantillon étudié (le hasard de l'échantillonnage)

31

31

L'importance clinique

Un test statistique ne vous donne pas l'importance clinique de la différence. Il vous donne seulement la probabilité d'observer une différence d'une certaine magnitude si l' hypothèse nulle est vrai – et rien d'autre

Pour l'importance clinique, si le résultat est statistiquement significatif, on va regarder **la différence entre les choses comparées par le test statistique** (la **différence entre les moyennes, ou la différence entre les pourcentages**).

Pour notre exemple l' importance clinique est suggérée par la différence entre la moyenne du poids (observée et puis généralisée sur la population) du milieu urbain et rural.

3314 g − 2240 g = 1074 g

Si le résultat n'est pas statistiquement significatif c'est pas bien d' interpréter la différence clinique parce que la différence observée peut être du au hasard.

Ici la différence est statistiquement significative, donc on peut faire une interprétation clinique. 1 kg différence - pour leur poids (2-4kg) semble être importante cliniquement.

32

Les erreurs des tests statistiques

Situations possibles pour la chois d'un hypothèse:

- Correctes:
  - l'acceptation de l'hypothèse nulle alors qu'elle est vraie
  - le rejet de l'hypothèse nulle alors qu'elle est fausse.
- Erreurs:
  - le rejet de l'hypothèse nulle alors qu'elle est vraie,
    - (erreur de première espèce - alpha)
  - l'acceptation de l'hypothèse nulle alors qu'elle est fausse,
    - (erreur de deuxième espèce - beta)

		Situation réelle	
		Il y a une différence (H1)	Il n y a pas une différence (H0)
Notre décision a l'aide du test statistique	Existe une différence (le rejet de H0)	Puissance (1 – beta)	**Erreur alfa (erreur I)**
	La différence n'existe pas (on reste avec H0)	**Erreur beta (erreur II)**	

33

33

Les erreurs des tests statistiques

Alpha – on le décide avec le niveaux de signification alpha (ex. 0,05)

Beta – on doit le calculer avant le test. Il dépend de la puissance du test.

La puissance d'un test statistique (plus difficile a calculer):

- la probabilité d'observer un différence **d** ou plus grande si la différence existe.
- Il dépend de la taille d'échantillon, du type du test, de la différence d désirée d'étre observe, alpha, …
- Plus un échantillon est large, plus la puissance est bonne.

Beta = 1 – la puissance.

- Donc plus un échantillon sera large, le risque d'erreur beta sera plus petit.

La diminution du risque alpha, augmente le risque bêta

34

34