Université Saad Dahlab Blida1
Faculté de Médecine
Département de Médecine
6ème année de Médecine

Module d'Epidémiologie

Mesures d'association statistique, principes des tests d'hypothèse (Cours à l'usage des étudiants en sciences médicales)

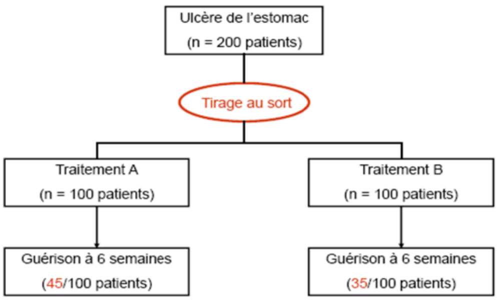
Chef de département de Médecine CPRS

Mesures d'association et Principes des tests d'hypothèse

Introduction :

- Lorsqu'on effectue une comparaison entre deux ou plusieurs séries de données on observe toujours une différence entre les paramètres mesurés.
- Le but du test est de déterminer si la différence observée est simplement due au hasard, (fluctuations d'échantillonnage), ou au contraire la différence observée est bien réelle.

Exemple : **Lequel de deux traitements est efficace ?**



Que conclure ?

Les tests statistiques d'hypothèse permettent de se fixer une règle de décision objective

- Un test n'a de sens que s'il teste une hypothèse **préalablement posée** afin de répondre **à une question**.
- Tout test statistique doit donc avoir pour objectif de **vérifier une hypothèse justifiée.**

Observation → hypothèse → test

Principe des tests :

- Servent à comparer des séries de données entre elles ; deux situations :

-comparer un échantillon observé à une population de référence.

-comparer deux ou plusieurs échantillons entre eux.

- Le principe général d'un test est de regarder si la différence qu'on observe est due au hasard ou non.

Quelle que soit la nature d'un test son principe et sa chronologie sont toujours les mêmes

1- Etablir l'hypothèse nulle(H0):

Proposer H0 c'est de supposer que la différence observée provient seulement **des fluctuation d'échantillonnage** donc la **différence n'est pas significative**.

2- Proposer une hypothèse alternative :

On appelle hypothèse **alternativeH1** l'hypothèse qui sera retenue au cas où les résultats du test aboutirait **à rejeter** l'hypothèse nulle H0, donc lorsque **la différence est significative.**

3-calcul statistique :
Calculer une quantité mathématique X exprimant l'écart entre les paramètres ou les distributions (chaque test est sa formule).
Confronter cette quantité à un modèle de distribution théorique **X`** (**table statistique de chaque test**).

4- résultat d'un test
deux situations :

- la valeur de **X**< à **X` (table)** .On en conclut que la différence observée entre les paramètres étudiées **n'est pas significative** (fluctuations d'échantillonnage).

H0 est pas retenue

-la valeur de **X** ≥ à **X`, donc** la différence observée entre les paramètres étudiées est significative.
on rejette l'hypothèse nulle H0 et on accepte l'hypothèse H1.

5-Choix du risque d'erreur :
a) **le risque α :**
c'est le risque de se tromper en rejetant H0.
On l'appelle risque de première espèce ou risque α
=**probabilité de rejeter H0 si H0 est vraie**
on lui assigne communément la valeur **5%**
b) **le risque β** :
risque β de deuxième espèce
= probabilité de ne pas rejeter H0,si H1 est vraie.

Université Saad Dahlab Blida1
Faculté de Médecine
Département de Médecine

6ème année de Médecine

Module d'Epidémiologie

PARAMETRES DE REDUCTION

(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA . S

Paramètres de réduction

Introduction :

Afin de résumer une série statistique d'une façon simple tout en en conservant au mieux le contenu informationnel en limitant au maximum la perte d'information, on utilise:

A- Les paramètres de tendance centrale (de position):

Elles permettent d'obtenir une idée de l'ordre de grandeur des valeurs de la série et indiquent la position où semble se rassembler les valeurs de la série.

B- Les paramètres de dispersion:

Elles quantifient les fluctuations des valeurs observées et leur étalement.

Les paramètres de tendance centrale :

a-mode:

Le mode est la variable qui a l'effectif le plus grand:

1-Données non groupées :

Ex: série : 3,5,7,15,**16,16,16**,17,17,30 Mode = 16

2-Données groupées: (Variable quantitative discrète)

Le mode correspond à l'effectif maximal. Mode =4

Nombre d'enfant	Nombre de famille
0	4
1	5
2	10
3	16
4	18
5	14
6	7

3-Données groupées: (variable quantitative continue):

La classe modale correspond à l'effectif le plus élevé. Le mode correspond au centre de la classe modale. La classe modale : 3.5- 4

Le mode = (3.5+4)/2 = 3.75 Mode = 3.75 kg

POIDS en KG	effectif
2 - 2.5	2
2.5 - 3	4
3 - 3.5	6
3.5 - 4	30
4 - 4.5	8

B-médiane:

La médiane est une valeur de variable qui divise l'ensemble des observations en 2 parties égales ,50% de l'effectif se situe en dessous de la médiane et 50% de l'effectif se situe au dessus.

1-Données non groupées:

Nombre d'observation est impair :

Ex: 7, 9,13 , 45, 70, 101,115 Médiane = 45

Nombre d'observation est pair :

Ex: 2, 5, 9 , 10, 12, 14, 20, 22 Médiane = (10+12)/2=11

2-Données groupées :(variable quantitative discontinue),

La médiane est la valeur de la variable qui occupe le (n/2) éme rang

N=80 le (n/2) éme rang = 40

Nombre d'enfant	Nombre de famille	Effectif cumulé
0	4	4
1	5	9
2	10	19
3	16	35
4	18	53
5	14	67
6	7	74
7	6	80

Université Saad Dahlab Blida1
Faculté de Médecine
Département de Médecine
6ème année de Médecine

Module d'Epidémiologie

COMPARAISON DES DISTRIBUTIONS DE CARACTERES QUALITATIFS LE TEST DU CHI-DEUX (χ^2)

(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA. S

COMPARAISON DES DISTRIBUTIONS DE CARACTERES QUALITATIFS LE TEST DU CHI-DEUX (χ^2)

A-Définition :
Formulations équivalentes : Test de chi- deux = Test de chi- carré = Test de Pearson
Le principe général d'un test est de regarder si la différence qu'on observe est due au hasard ou si au contraire cette différence est telle qu'il est fort peu probable de l'observer par hasard

Quelle que soit la nature d'un test, son principe et sa chronologie sont toujours les

Mêmes.

B-Principe du test :
Le test de χ^2 permet de tester la liaison entre deux ou plusieurs distributions de caractères qualitatifs

C-Categories du test :
1-Comparaison d'une distribution observée et théorique
Il sert à comparer une distribution observée sur un échantillon à une distribution connue dans la population ou à une distribution théorique

Il s'agit d'un test statistique qui étudie l'écart entre la distribution théorique et la distribution observée
Exemple :
Dans une maternité, sur 100 naissances, on a observé 44 garçons et 56 filles ; cette observation est- elle compatible avec la statistique nationale donnant les proportions de naissances de garçons et de filles de respectivement 53% et 47% ?
Tableau de contingence:

	Garçon	Fille	Total
Statistique nationale	0.53	0.47	1
Effectif théorique (Ci/Ti)	53	47	100
Effectif observé	44	56	100

Démarche à suivre :

1- **Ho** :la distribution par sexe observée à la maternité est conforme à la distribution nationale

2-choix du test ➔ un test de χ^2

3-vérification des conditions d'application :

Tous les $T_i \geq 5$ (53 et 47 sont > 5)

4-Calcul de la statistique (somme des écarts) :

$$\chi^2 = \sum \frac{(Ci - Oi)^2}{Ci}$$

χ^2 = (53-44)² / 53 + (47-56)² / 47 =3.25

5-on fixe le seuil de signification au risque **α = 5%**

6-on compare le χ^2 calculé au χ^2 de la table au risque α = 5% et avec un nombre de degré de liberté **ddl = (l – 1)(c – 1)**

(k = nombre de modalités de la variable étudiée) ➔ k= 2-1=1

7-conclusion :

χ^2 calculé < **χ^2** de la table (3.25 < 3.84)

➔ **Ho n'est pas rejetée au risque** α = 5%

➔ La distribution observée est **conforme** à la répartition nationale par sexe des naissances

2-Comparaison entre plusieurs distributions observées :
Il sert à comparer deux ou plusieurs distributions observées sur des échantillons

Exemple :
Deux médicaments (A et B) ont été testés sur deux groupes de malades.
A l'issue de l'essai, on a observé les résultats suivants :

	Disparition des symptomes	Persistance des symptomes	aggravation	Réaction secondaire	Total
A	100	40	20	30	190
B	220	80	70	40	410
Total	320	120	90	70	600

Peut-on dire que ces deux traitements ont les mêmes effets ?
Correction :
1-H0 :
les deux traitements ont les mêmes effets ➔
il n'y a pas de différence entre les deux distributions ➔ les différences observées résulteraient des seules fluctuations d'échantillonnage.
2-tableau de contingence :

	Disparition des symptômes	Persistance des symptômes	aggravation	Réaction secondaire	Total
A	100 /**101.33**	40/**38**	20/**28.5**	30/**22.16**	190
B	220/**218.66**	80/**82**	70/**61.5**	40/**47.83**	410
Total	320	120	90	70	600

3-choix du test ➔ un χ^2

4-vérification des conditions d'application :

Tous les Ti ≥ 5

Calcul des Ti :

Ti = (total de la ligne × total de la colonne) / total general

(101.33- 218.66- 38- 82 - 28.5- 61.5- 22.16 et 47.83) sont tous > 5

5-Calcul de la statistique (somme des écarts) :

$$\chi^2 = \sum \frac{(C_i - O_i)^2}{C_i}$$

χ^2 = **(100 – 101,33) ² / 101,33 + (220 – 218,66) ² / 218,33 + (40 – 38) ² / 38 + (80 – 82) ² / 82 + (20 – 28,5) ² / 28,5 + (70 – 61,5) ² / 61,5 + (30 – 22,16) ² / 22,16 + (40 – 47,83) ² / 47,83**

χ^2 = **0,0175 + 0,0082 + 0,105 + 0,048 + 2,535 + 1,174 + 2,773 + 1,28**

$$\chi^2 = 7{,}94$$

6-on fixe le seuil de signification au risque **α = 5%**

7-on compare le χ^2 calculé au χ^2 de la table au risque α = 5% et avec un nombre de degré de liberté

ddl = (C-1) × (L – 1)

C= nombre de colonnes

L= nombre de lignes ddl = (4-1) × (2-1) = 3

8-conclusion:

χ^2 calculé > χ^2 de la table (7.94 > 7.81)

- ➔ Ho est rejetée
- ➔ La différence est significative au risque α = 5%
- ➔ les deux distributions **ne sont pas homogènes**. Ou

Les deux traitements ont des effets différents au risque de 5%.

Avec **α = 5 %** et **d.d.l = 3** χ^2 **= 7,815**

3-Etude de la liaison entre les distributions de deux variables :

Il sert à étudier sur un même échantillon la liaison entre les distributions de deux variables

Exemple :

Un épisode d'intoxication alimentaire collective (TIAC) est survenu parmi les stagiaires d'un atelier, Le docteur chargé de l'enquête a dressé le tableau suivant croisant l'information sur la consommation de glace au chocolat, l'un des desserts proposés au cours du dernier repas pris en commun par les stagiaires, et le statut malade / non malade.

1-

Tableau de contingence :

Ho : il n'y a pas de liaison entre la consommation de glace au chocolat et la survenue de la gastro- entérite.

	Malade	Sain	Total
Glace au chocolat	270 (206.5)	7 (70.5)	277
Pas de glace	26 (89.5)	94 (30.5)	120
Total	296	101	397

Le principe et le calcul du test sont identiques

χ^2 calculé > χ^2 de la table (253.7 > 3.84)

- ➔ Ho est rejetée
- ➔ La différence est significative au risque $\alpha = 5\%$
- ➔ Il existe une liaison statistique entre la consommation de glace au chocolat et la survenue de la gastro- entérite au risque de 5%.

Table de la loi du X^2

Cette table donne la probabilité p d'observer une valeur supérieure à une valeur de X^2 en fonction du nombre de degrés de liberté (dl)

dl / p	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.001
1	0.000	0.001	0.004	0.016	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	13.82
3	0.12	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	20.52
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	22.46
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.47	24.32
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	26.13
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	27.88
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	29.59
11	3.05	3.82	4.57	5.58	17.27	19.67	21.92	24.72	31.26
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	32.91
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	34.53
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	36.12
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	37.70
16	5.81	6.91	7.96	9.31	23.54	26.30	28.84	32.00	39.25
17	6.41	7.56	8.67	10.08	24.77	27.59	30.19	33.41	40.79
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.80	42.31
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	43.82
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	45.32
21	8.90	10.28	11.59	13.24	29.61	32.67	35.48	38.93	46.80
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	48.27
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	49.73
24	10.86	12.40	13.85	15.66	33.20	36.41	39.37	42.98	51.18
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	52.62
26	12.20	12.84	15.38	17.29	35.56	38.88	41.92	45.64	54.05
27	12.88	13.57	16.15	18.11	36.74	40.11	43.19	46.96	55.48
28	13.57	15.31	16.93	18.94	37.92	41.34	44.46	48.28	56.89
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	58.30
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	59.70