PLAN GÉNÉRAL DU COURS

1 - La méthode Statistique en Médecine

2 - Statistique Descriptive

3 - Statistique Déductive

- ***Liaisons entre caractères qualitatifs***
- ***Liaisons entre caractères qualitatifs et quantitatifs***
- ***Liaisons entre caractères quantitatifs***
- ***Tests non paramétriques***

3 - Statistique Déductive

- Observations, mesures
- Conclusions
- Les tests

Objectifs pédagogiques :

Notion d'hypothèses
Risque de première espèce
Choix du bon test
Interprétation statistique et médicale

3 - Statistique Déductive

Tirer des conclusions à partir d'observations

Exemple :

Comparer 2 groupes pour un caractère donné.

2 hypothèses :

- **H0 = Hypothèse nulle. Pas de différence observée.**
- **H1 = Hypothèse alternative. Différence significative entre les 2 groupes.**

LES TESTS

Techniques permettant de décider si on garde ou repousse H0, en ayant fixé le risque d'erreur accompagnant cette décision.

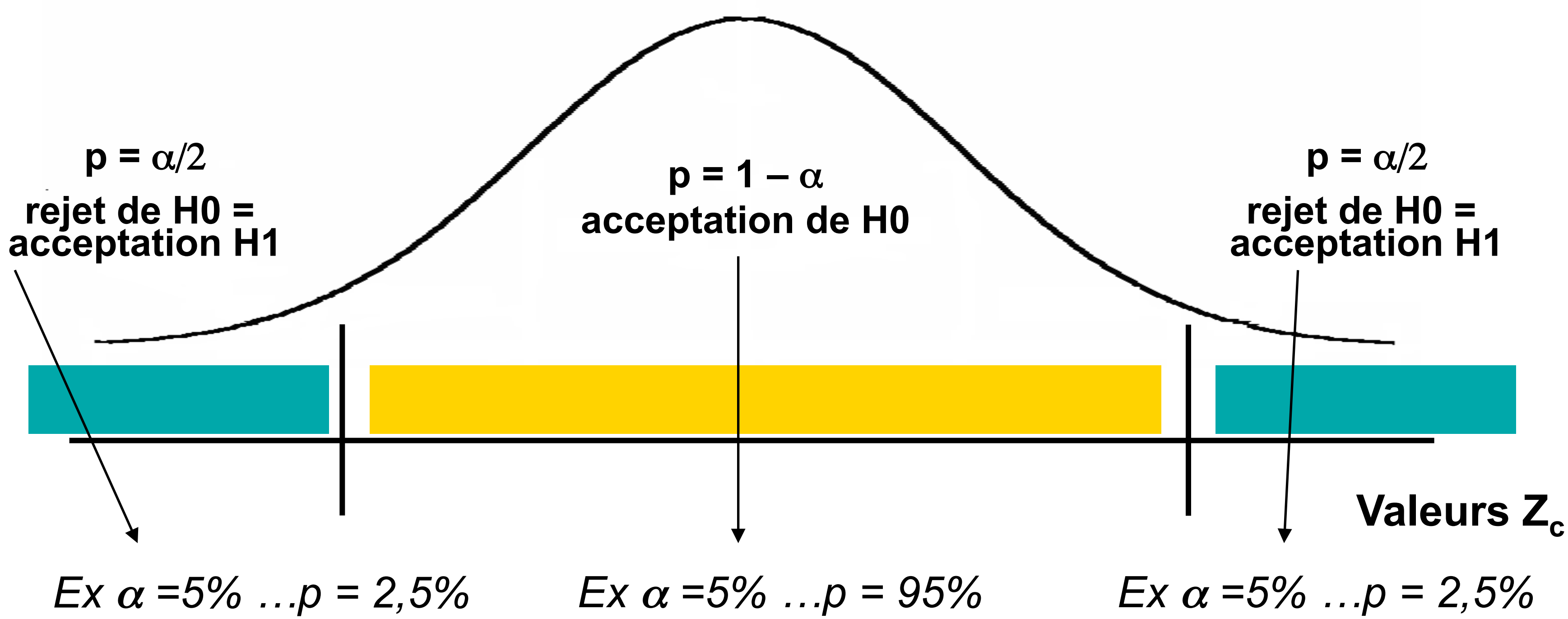
LES ÉTAPES DE MISE EN ŒUVRE D'UN TEST D'HYPOTHÈSE

A propos d'un problème médical, question simple qui sera testée

- **Etape 1** : Avant recueil des données définir H0 et H1. Les 2 hypothèses jouent des rôles **symétriques**
- **Etape 2** : Avant recueil des données **définir le test en fonction du type des données (qualitatives, quantitatives)**. Soit **Z** le paramètre qui sera calculé
- **Etape 3** : Avant recueil des données on choisi **le risque** α (dans la pratique souvent 5%)
- **Etape 4** : Recueil des données.
  Calcul de Z.
  Règle de décision : examiner la position de cette valeur Z, par rapport à un modèle théorique dont on connait la distribution.
- **Etape 5** : Interprétation des résultats.

Interprétation graphique du risque α

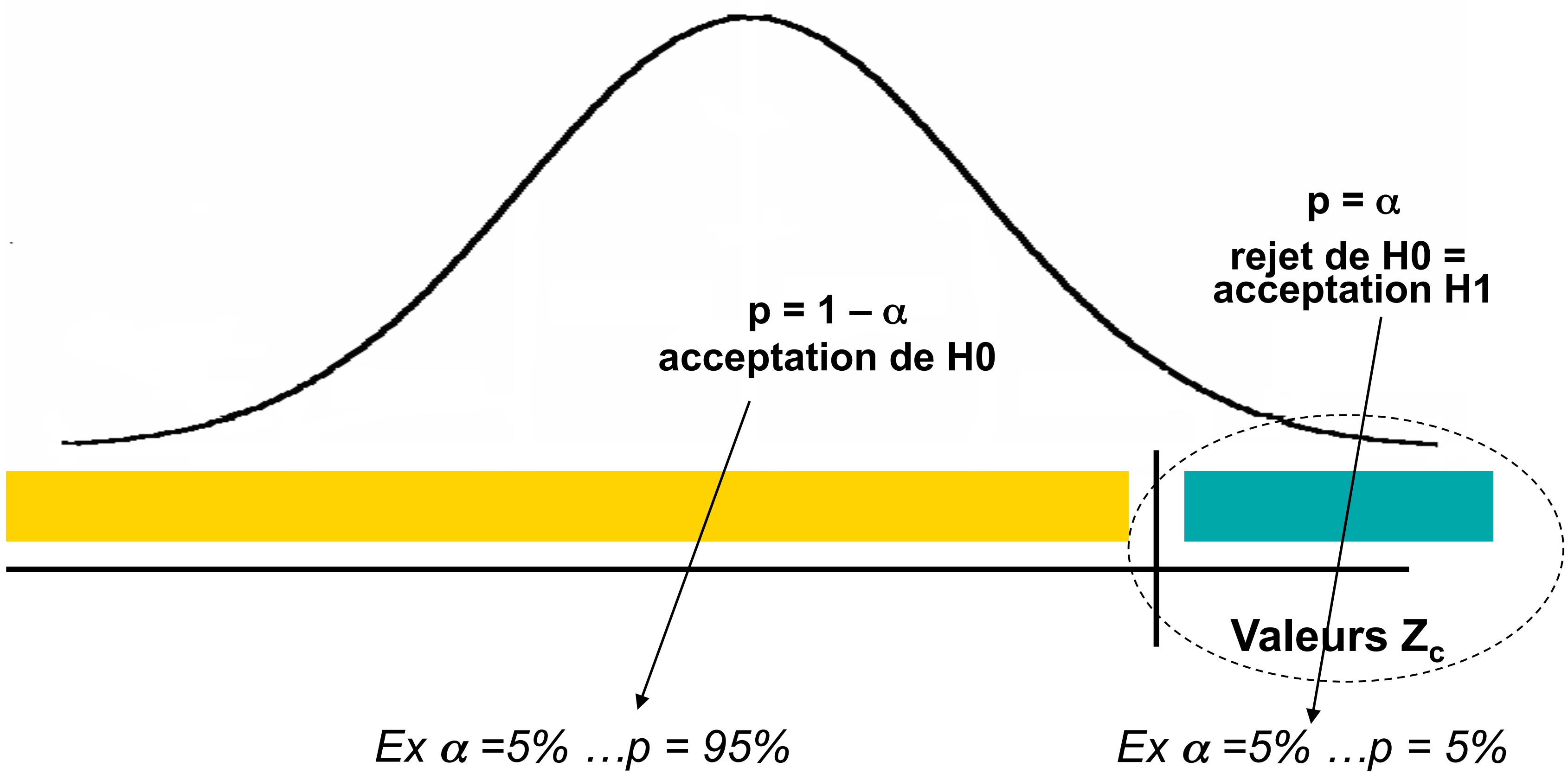
Le paramètre Zc que nous allons apprendre à calculer, suit une distribution probabiliste en forme de **courbe de Gauss**.



Situation bilatérale : **les 2 situations observées sont elles différentes?**

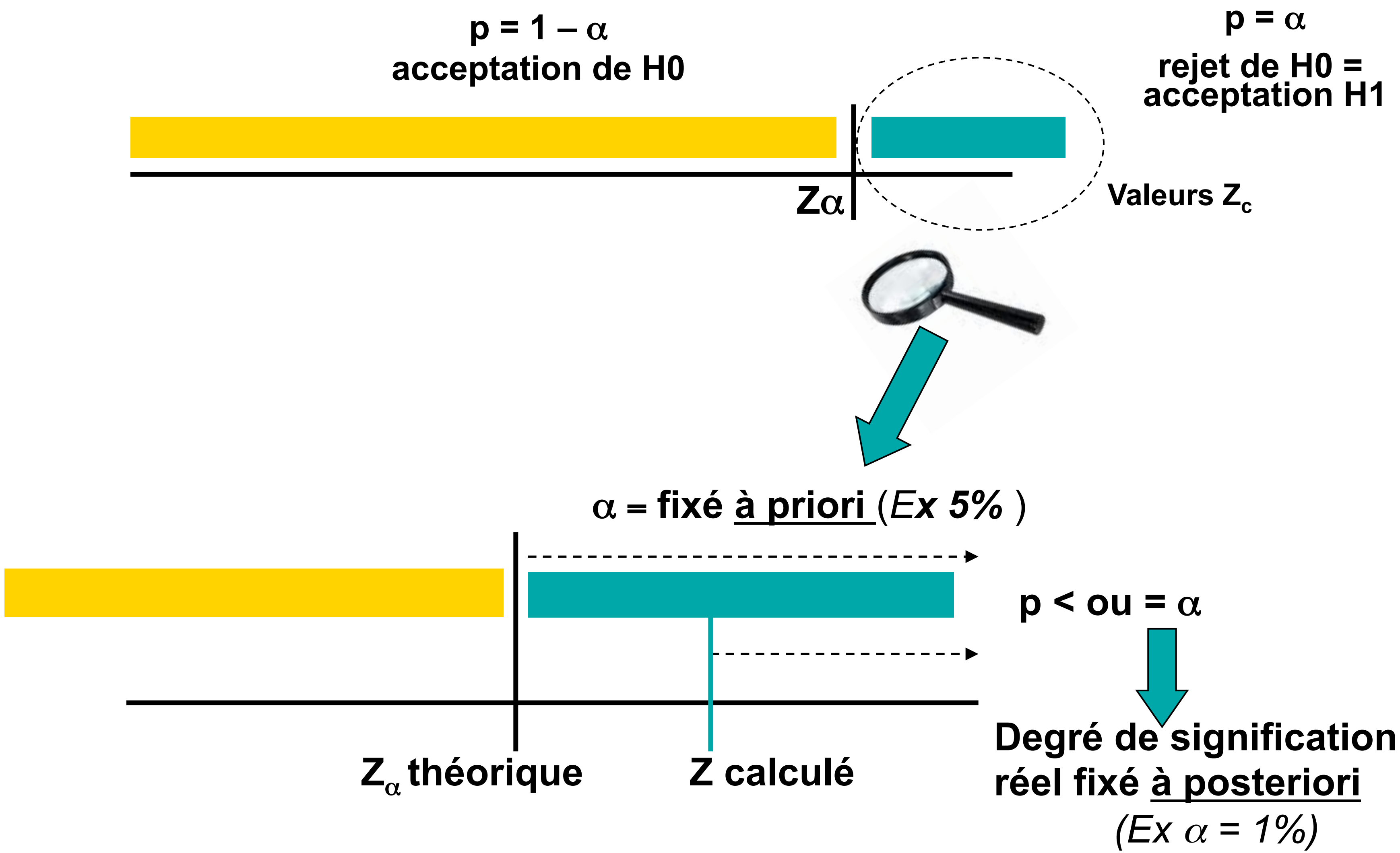
Si OUI, il est possible de classer ces 2 situations (laquelle est > ou <,..)

Interprétation graphique du risque α



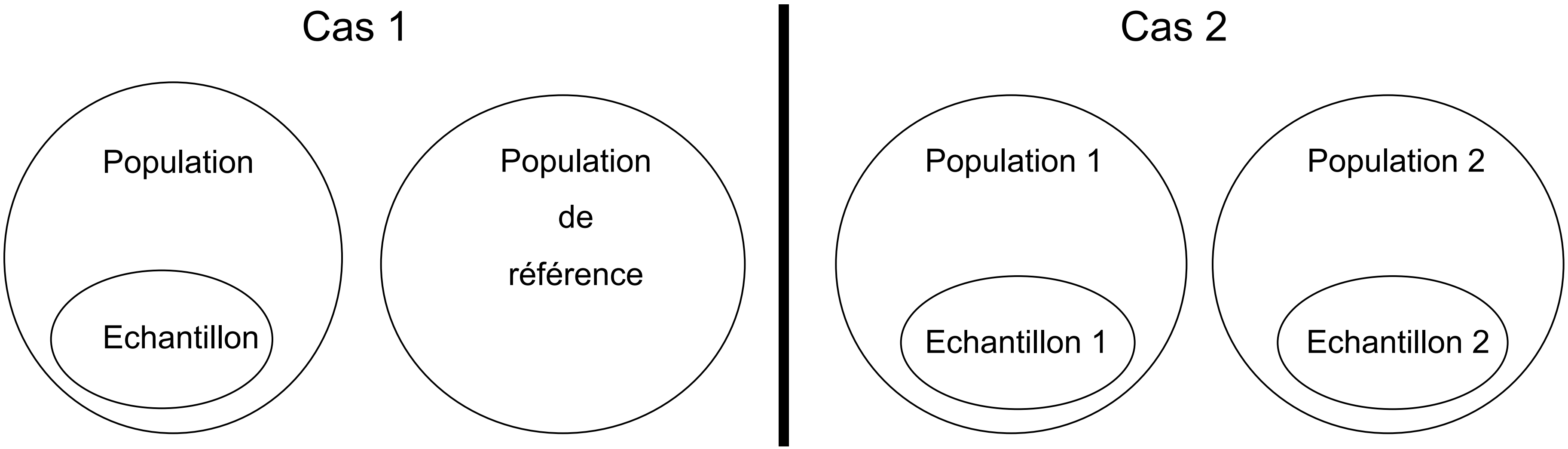
<u>Situation unilatérale</u> : les 2 situations observées sont elles différentes ?

Interprétation graphique du risque α



3 - Statistique Déductive

Les tests de comparaison : 2 cas



Différence observée :

Fluctuations d'échantillonnage ?

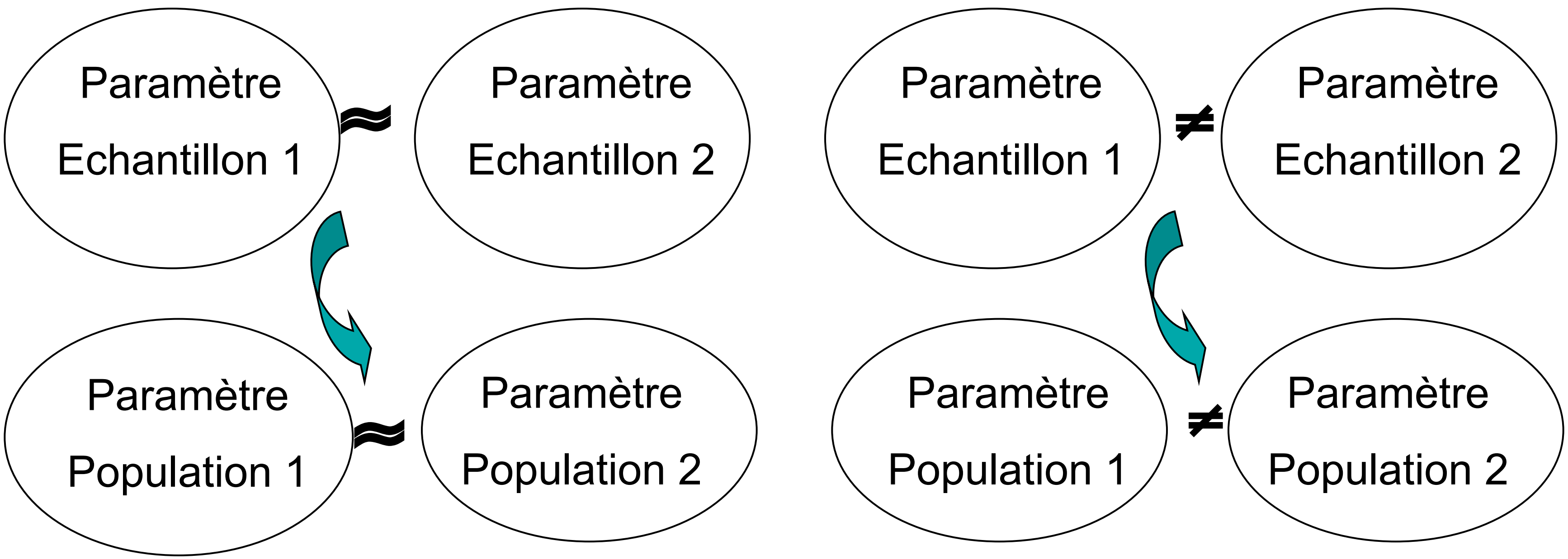
Différences réelles entre populations ?

Hasard?

Test Statistique

3 - Statistique Déductive

Formulation des hypothèses



Hypothèse nulle (H0) **Hypothèse alternative (H1)**

3 - Statistique Déductive

Qu'appelle t on risque ?

Risque de première espèce : α **Probabilité de rejeter H0 si H0 vraie.**

compromis universel : α = 5%

Risque de seconde espèce : β **Probabilité d'accepter H0, si H0 fausse**

Puissance d'un test : $1 - \beta$ **Probabilité de rejeter H0 si H0 fausse.**

Il se peut que le risque de deuxième espèce β soit assez important. L'erreur α est celle qu'on choisit de maîtriser, quitte à ignorer β. Cela induit une dissymétrie dans le traitement des deux hypothèses.

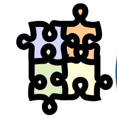
La règle de rejet du test est définie uniquement à partir de α **et H0.** Entre deux alternatives, on choisira pour H0 l'hypothèse qu'il serait le plus grave de rejeter à tort.

3 - Statistique Déductive

Les risques d'erreur

Décision du statisticien

Réalité	Rejet H0	Non rejet H0
H0 Vraie	Erreur 1ère espèce α	$1 - \alpha$
H1 Vraie	Puissance $1 - \beta$	Erreur 2ème espèce β



Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).

« Le risque de seconde espèce »

A) Est noté β et vaut environ 20%

B) Est défini à priori

C) Correspond au risque de rejeter a tort l'hypothèse alternative.

D) Correspond à la puissance

E) Est le risque d'accepter à tort l'hypothèse alternative

Réponse C : C'est la définition de β

Le risque de seconde espèce n'est jamais défini à priori (A, B)

La puissance est **1- β (D)**

Le risque d'accepter à tort H1 est **α (E)**

Exercice réflexion

Dans un procès on demande au jury de décider entre H0 « accusé innocent » et H1 « accusé coupable ». Pour chaque question suivante, préciser quelle(s) réponse(s) est (sont) exacte(s).

1- l'erreur de première espèce (risque α) correspond à :

A) Accusé innocent mais condamné

B) Accusé innocent et relâché

C) Accusé relâché faute de preuves

Réponse A) C'est la définition même du risque **α : rejeter à tort H0**.

B) Peut se traduire par : accepter H0.

C) Aucun test n'est effectué : il manque des données. L'étude doit se poursuivre.

Exercice réflexion

Dans un procès on demande au jury de décider entre H0 « accusé innocent » et H1 « accusé coupable ».

A quel risque correspond l'affirmation suivante ?

Accusé coupable mais déclaré innocent et relâché

C'est la définition du risque **β : accepter H0**, alors qu'elle est fausse

PLAN GÉNÉRAL DU COURS

1 - La méthode Statistique en Médecine

2 - Statistique Descriptive

3 - Statistique Déductive

- ***<u>Liaisons entre caractères qualitatifs</u>***
- ***Liaisons entre caractères quantitatifs***
- ***Liaisons entre caractères qualitatifs et quantitatifs***
- ***Tests non paramétriques***

3 - Statistique Déductive

Etude de la liaison entre 2 caractères qualitatifs.

Question :

Le pourcentage d'un certain type d'individus dans un groupe *A* coïncide-t-il avec le pourcentage du même type d'individus dans un autre groupe *B* ?

1) Comparaison de 2 pourcentages observés.

$$\varepsilon = \frac{p_A - p_B}{\sqrt{\dfrac{p_A q_A}{n_A} + \dfrac{p_B q_B}{n_B}}}$$

ε < 1,96 avec α <5 %

q = probabilité complémentaire de p = 1 – p

2) Test du χ^2

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

χ^2 tabulé

Nb ddl = (nb lignes-1)(nb colonnes-1)

Exercice : Comparaison de pourcentages observés

On cherche à savoir si le mode de garde (crèche ou domicile) modifie le risque de rhinopharyngite des enfants. On fait une étude sur 2 groupes de 200 enfants:

Crèche	n_A=200	Nb rhino = 130
Domicile	n_B=200	Nb rhino = 96

Le mode de garde influe-t-il sur le risque d'avoir une rhinopharyngite?

Quel test statistique, et conclusion ?

LES ÉTAPES DE MISE EN ŒUVRE D'UN TEST D'HYPOTHÈSE

- Etape 1 : Avant recueil des données définir H0 et H1 (1)
- Etape 2 : définir le test en fonction du type des données (qualitatives, quantitatives). Soit Z le paramètre calculé (2, 3, 4)
- Etape 3 : Avant recueil des données on choisi le risque α (dans la pratique souvent 5%)
- Etape 4 : Recueil des données.
- Calcul de Z.
- Règle de décision : examiner la position de cette valeur Z, par rapport à un modèle théorique dont on connait la distribution (5, 6, 7)

- Etape 5 : Interprétation des résultats.

Exercice Comparaison de pourcentages observés

1. H0 : il n'y a pas de différence entre les 2 modes de garde vis-à-vis des rhinopharyngites
2. Caractère qualitatif 1 : Garde en crèche ou à domicile
3. Caractère qualitatif 2 : Avoir une rhinopharyngite ou non
4. Test = Comparaison de pourcentages
5. $p_A = 130/200 = 65\%$ $p_B = 96/200 = 48\%$
6. $$\varepsilon = \frac{0,65 - 0,48}{\sqrt{\dfrac{0,65 \times 0,35}{n_A} + \dfrac{0,48 \times 0,52}{n_B}}} = 3,4$$
7. Table de l'écart réduit → $\varepsilon > 3,3$ ($p < 1\ {}^{0}/_{00}$)

Table de l'écart réduit

α

		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	∞	2,576	2,326	2,17	2,054	1,96	1,881	1,812	1,751	1,695
0,1	1,645	1,598	1,555	1,514	1,476	1,44	1,405	1,372	1,341	1,311
0,2	1,282	1,254	1,227	1,2	1,175	1,15	1,126	1,103	1,08	1,058
0,3	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,86
0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,69
0,5 ε	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,6	0,524	0,51	0,496	0,482	0,468	0,454	0,44	0,426	0,412	0,399
0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,8	0,253	0,24	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,9	0,126	0,113	0,1	0,088	0,075	0,063	0,05	0,038	0,025	0,013

Table pour les petites valeurs de la probabilité

0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
3,2905	3,8905	4,41717	4,89164	5,32672	5,73073	6,10941

Comparaison de pourcentages observés (suite)

Le test statistique vient de démontrer **<u>sur cet échantillon</u>**, que le risque de rhinopharyngites est supérieur chez les enfants gardés en crèche (p <0,001)

On peut conclure **<u>sur cet échantillon</u>** : le mode de garde est cause de cette différence

On ne pourra pas généraliser cette conclusion au niveau de tous les enfants en âge d'être gardés en France ou ailleurs, car :

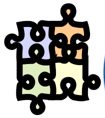
Il n'y a pas eu TAS. On ne sait rien des enfants, ni des lieux de garde. Les familles n'ont peut être pas les mêmes revenus, donc l'accès aux soins n'est pas forcément le même..

Nous distinguerons toujours les 2 aspects à discuter :

a) L'aspect statistique et ses conclusions

b) L'aspect médical et ses conclusions qui peuvent être différentes.

Exercice : Test de comparaison de pourcentages



On veut étudier l'efficacité d'un nouveau traitement (T) contre la leucémie.

On administre T à 50 souris et le traitement de référence (R) à 50 autres souris de la même espèce. On note au bout d'un mois, 33 morts dans le groupe T et 44 morts dans le groupe R.

Peut on conclure à la supériorité de ce nouveau traitement?

1. On dispose de 2 groupes indépendants traités T ou R : variables qualitatives
2. Dénombrements des morts dans chaque groupe (mort ou non) : variables qualitatives
3. Comparaison de %
4. H0 : Il n'y a pas de différence significative entre les 2 traitements.

Exercice : Test de comparaison de pourcentages

groupe T % DC = 33/50=66 %

groupe R %DC = 44/50=88%

On peut calculer ε

$$\varepsilon = \frac{0{,}88 - 0{,}66}{\sqrt{\dfrac{0{,}88 \text{x} 0{,}12}{50} + \dfrac{0{,}66 \text{x} 0{,}34}{50}}} = 2{,}83$$

ε calculé = 2,83 > ε théorique lu dans la table (α=5%) = 1,96

On rejette H0. Il existe une diff significative entre les 2 groupes (H1), α<0,001

On conclut que le % DC traitées par T < à celui des souris traitées par R

T est meilleur que R **sur cet échantillon.**

On ne sait rien des groupes, de l'état de santé des souris, et même de l'étude.

On ne peut pas généraliser ce résultat.

3 - Statistique Déductive

Etude de la liaison entre 2 caractères qualitatifs.

1 - Comparaison de 2 pourcentages observés.

$$\varepsilon = \frac{p_A - p_B}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}}$$

ε < 1,96 avec α <5 %

q = probabilité complémentaire de p = 1 – p

2 - Test du χ^2

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

χ^2 tabulé

Nb ddl = (nb lignes-1)(nb colonnes-1)

CAR 1

CAR 2

	A	B	Total
C	n_1	n_3	Tot C
D	n_2	n_4	TOt D
Total	Tot A	Tot B	T

Exemple 2 lignes, 2 colonnes : ddl = 1

Exercice : Test du χ^2

On cherche à savoir si l'exposition professionnelle au benzène peut entrainer une leucémie. On lance une étude dans une grande entreprise, on dénombre les salariés exposés au benzène, et ceux qui ne le sont pas. Au bout de 12 ans, on fait le bilan des leucémies apparues.

	Leucémies	Non Leucémies	Total
Expo	15	485	500
Non Expo	20	980	1000
Total	35	1465	1500

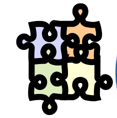
Existe-t-il une relation entre exposition au benzène et leucémies ?

Exercice : Test du $\chi 2$

1. H0 : il n' existe pas de lien entre expo benzène et leucémie
2. Variables qualitatives 1 : Malades leucémie ou non malades leucémie
3. Variables qualitatives 2 : Exposition au benzène ou non exposition
4. Test du $\chi 2$ permet de prendre en compte tous les cas de figure (expo-malades, expo-non malades, non expo-malades, non expo-non malades) et pas seulement deux %.
5. Si répartition au hasard les nb de leucémies seraient à peu près identiques dans les 2 groupes Expo et Non expo. Nous allons donc construire ce modèle et comparer la situation réelle à ce modèle théorique.

Université Nice Sophia Antipolis

Exercice : Test du $\chi 2$ (suite)



35/1500 = 2,33% malades, expo ou non, et 1465/1500 = 97,66% non malades.

Appliquons ces % aux salariés expo et non expo : **modèle théorique**.

2,33 % de 500 = 11,65 salariés, chiffre théorique de malades chez les expo.

2,33 % de 1000 = 23,35 salariés, chiffre théorique de malades chez les non expo.

	Leucémies	Non Leucémies	Total
Expo	15 **(11,65)**	485 **(488,3)**	500
Non Expo	20 **(23,35)**	980 **(976,7)**	1000
Total	35	1465	1500
%	2,33	97,66	

Les chiffres calculés (rouge), forment le modèle théorique. Nous allons les comparer aux chiffres observés (noir), à l'aide de la formule ci-dessous :

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

Test du $\chi 2$ (suite)

	Leucémies	Non Leucémies	Total
Expo	15 11,65	485 488,3	500
Non Expo	20 23,35	980 976,7	1000
Total	35	1465	1500

$$\chi 2 = (15-11{,}65)^2/11{,}65 + (20-23{,}35)^2/23{,}35 + (485-488{,}3)^2/488{,}3 + (980-976{,}7)^2/976{,}7$$

$$\chi^2 = 1{,}42$$

Nb de degrés de liberté = (nb lignes-1)(nb colonnes – 1) = 1

La table du χ^2 indique que χ^2 calculé < χ^2 théorique (α = 5%, soit 3,84)

Nous acceptons H0 :

Il n'existe pas de relation entre expo benzène et apparition des leucémies.

Table du $\chi 2$ (extrait)

ddl \ α	0,9	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,016	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,21	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,61	4,351	6,064	7,289	9,236	11,07	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,49	7,344	9,524	11,03	13,362	15,507	18,168	20,09	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,34	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,34	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,79	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,79
…									