Exemple de sondage

1

La popularité de XX et YY chute. Selon un sondage BVA diffusé ce vendredi, les deux hommes perdent 5 points. Avec respectivement 34% et 38% d'opinions positives, XX et YY sont à leur plus bas niveau dans ce baromètre depuis leur entrée en fonction.

A 34%, la cote de XX reste légèrement supérieure à celle de ZZ (32%) au même moment de son mandat, mais très en-dessous de celle de ZZZ (37%).

Commentaire absurde du point de vue statistique! Il est possible que les IC de ces pourcentages se recouvrent très largement, et que donc, finalement ces pourcentages soient considérés comme identiques!

Exemple de sondage

- **Le Monde 02/09/2018**

La côte de popularité de XXX s'établit à 11%, d'opinions positives, selon le sondage de l'institut Louis Harris. Cette popularité est en baisse de trois points par rapport à juillet dernier où elle était mesurée à 14%.

Sondage réalisé par téléphone du 12 au 28 Aout 2018, auprès d'un échantillon de 1059 personnes, représentatif de la population française, âgées de 18 ans et plus, constitué selon la méthode des quotas.

En fait il manque l'IC seule façon statistiquement inattaquable de présenter des résultats de sondage !

- **Sondage Ifop pour Sud Radio :** *réalisé en ligne du 31 aout au 4 septembre 2017, auprès d'un échantillon de 937 personnes inscrites sur les listes électorales (méthode des quotas). L'Ifop prévient que la marge d'erreur est de 1,8 point pour un score de 10% ; 2,5 points pour un score de 20% ; 2,8 points pour un score de 30%.*

Commission des sondages

Communiqué du 5 septembre 2016

À la suite de la diffusion, à compter du 2 septembre 2016, par les journaux L'Opinion et Valeurs actuelles des résultats d'un sondage d'intentions de votes relatif à l'élection présidentielle de 2017, **la commission des sondages**, agissant en application de la loi du 19 juillet 1977, a entendu, le vendredi 5 septembre, successivement, les représentants de ces deux organes de presse qui n'ont pas été en mesure de lui fournir les informations relatives à ce sondage électoral.

En l'absence de tout élément permettant d'établir, de manière certaine, l'existence de ce sondage électoral et, a fortiori, de contrôler la qualité des méthodes retenues pour obtenir les résultats diffusés, la commission des sondages appelle l'attention de l'opinion publique sur l'absence de fiabilité de ces résultats.

« Les internes en médecine, "bouche-trous" de l'hôpital en crise »
Le Monde

4

Le Monde consacre une page aux *« internes en médecine, "bouche-trous" de l'hôpital en crise ». Tous assument déjà, souvent dans la plus grande illégalité, le travail et les responsabilités d'un médecin diplômé, poussés à bout par un système hospitalier en sous-effectif chronique, saturé par la demande ».*

Le Monde rappelle qu'*« en septembre, le principal syndicat d'internes, l'Isnih, a publié les* ***résultats d'une vaste enquête menée dans les hôpitaux français****. Le rapport montrait ainsi que* ***85%*** *des internes […] travaillaient bien au-delà des 48 heures hebdomadaires réglementaires, avec* ***une moyenne de 60 heures par semaine****. De même, le "repos de sécurité", imposé depuis 2002 après chaque garde de nuit pour empêcher un interne de travailler plus de 24 heures consécutives, n'est pas respecté* ***dans 21% des cas*** *».* Le quotidien poursuit : *« Selon le syndicat, "aucune région ne respecte aujourd'hui la législation", et ces entorses au règlement ne sont pas sans conséquences.* ***15% des étudiants*** *affirment avoir commis des erreurs de prescription, de diagnostic ou d'acte opératoire en lendemain de garde, alors que* ***39% déclarent*** *en avoir "probablement réalisé".*

2.2 Estimation Statistique

Comparaison estimation ponctuelle / estimation par intervalle

Soit un groupe de 220 patients, représentatif d'une population rhumatismale (R).
On observe 167 cas de rhumatismes inflammatoires.

Quel pourcentage de rhumatismes inflammatoires dans la population R?

1) Estimation ponctuelle p=167/220 = 0,76 soit **76%**

2) Estimation par intervalle

Nous choisissons le risque α **= 5%,** donc calcul de $IC_{0,95}$

p = 0,76 donc q = 0,24

$$IC_{0,95} = 0,76 \pm 1,96 \sqrt{\frac{0,76 \times 0,24}{220}}$$

$$\mathbf{IC_{0,95} = [\ 0,70\ ;\ 0,82\ ]}$$

L'estimation par intervalle semble moins précise. Mais si l'on refait ce calcul sur un autre échantillon, cette nouvelle estimation recouvrira la première. Ce ne sera pas forcément vrai avec l'estimation ponctuelle.

Précision

Soit P la population des ouvriers travaillant dans une usine

Nous voulons estimer le pourcentage **p** d'hommes dans cette population.

Considérons un échantillon TAS de 10 ouvriers : 7 hommes, soit $\mathbf{p_o=70\%}$

Estimation au niveau de P, au risque α=1% ?

$s = \sqrt{\frac{0,7 \times 0,3}{10}} = 0,144$ $\quad IC_{99\%} = [0,7 \pm 2,6 \times 0,144] = [\ 33,6\% ; 100\ \%]$

Considérons un échantillon de 1000 ouvriers : même % d'hommes $\mathbf{p_0 = 70\%}$

$s = \sqrt{\frac{0,7 \times 0,3}{1000}} = 0,014$ $\quad IC_{99\%} = [0,7 \pm 2,6 \times 0,013\ ] = [\ 66,2\% ; 73,8\%]$

Effectif n augmente ⇒ IC se resserre ⇒ Précision augmente

ESTIMATION DE VALEURS
Important !

- Taille de l'échantillon
- Représentativité de l'échantillon
- Absence de biais lors de la sélection

La précision déterminée par l'intervalle de confiance.

s petit ➔ Intervalle de confiance diminue ➔ Précision augmente

La précision est fonction de $\sigma = \sqrt{\frac{p_0 q_0}{n}}$

⇨ (n multiplié par 100 ➔ σ divisé par 10 ➔ précision augmente facteur 10)

Afin de réduire l'intervalle de confiance, il est important que la taille de l'échantillon soit grande (il existe des formules). Les valeurs estimées sont alors plus précises.

Un chirurgien écrit à 1000 de ses patients afin de connaitre leurs suites chirurgicales ➔ sur 100 réponses : 75 vont très bien, 25 ont des séquelles handicapantes.

Le chirurgien s'intéresse aux mauvaises suites chirurgicales, et veut estimer le % au niveau de l'ensemble de ses 1000 patients:

$$IC_{95\%} = [\ 0{,}25 \pm 1{,}96 \times \sqrt{\frac{0{,}25 \times 0{,}75}{100}}\ ]$$

Soit $IC_{95\%}$ = 25% ± 8% de mauvais résultats soit **[17% ; 33%]**

Résultat : calcul statistique correct, mais non utilisable : il est faux !

1) Il y a eu 900 non-réponses. On ne peut pas préjuger de l'état de ces 900 patients. Ils sont peut être décédés des suites opératoires, ou bien très mécontents du chirurgien, ou tout au contraire sont très satisfaits et ne jugent pas utile de répondre .. Cet échantillon est BIAISÉ

2) Si on se trouve dans ce dernier cas, les échecs ne représentent que 25/1000 = 2,5%. La conclusion est toute autre !!

a) Données quantitatives

$$m = \frac{\sum_{1}^{n} x_i}{n}$$

m = moyenne calculée sur l'échantillon

$$s = \sqrt{\frac{\sum_{1}^{n} (x_i - m)^2}{n - 1}}$$

s = écart type calculé sur l'échantillon

Estimation de la moyenne inconnue dans la population cible

$$\mu \in \left[m \pm \varepsilon \frac{s}{\sqrt{n}} \right]$$

ε lu dans la table ⇒ risque d'erreur accepté

b) Données qualitatives

% au niveau de l'échantillon = p_o et $s = \sqrt{\frac{p_0 q_0}{n}}$ (avec $q_0 = 1 - p_0$)

Estimation du % inconnu dans la population cible

$$p \in [p_0 \pm \varepsilon s]$$

ε lu dans la table ⇒ risque d'erreur accepté

10

On compare la diminution du taux de cholestérol produite par un nouveau tt N à celle produite par le tt de référence R.

Après tt dans le groupe N (n_T=400) $IC_{95\%}$ = [1,10 ; 1,20] m_T= 1,15 mmol/l
Après tt dans le groupe R (n_R=400) $IC_{95\%}$ = [1,30 ;1,50] m_R= 1,4 mmol/l

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

A) Le nouveau traitement N est plus efficace que celui de référence R
B) Différence statistiquement significative car $m_R \notin IC_{95\%}$ (m_N)
C) Les intervalles de confiance permettent de conclure que $m_N \neq m_R$
D) Différence significative entre les 2 traitements.
E) On ne peut rien conclure, car on ne sait pas si la distribution des valeurs est normale.

Réponses : B, C, D Les 2 IC ne se recouvrent pas. Les 2 moyennes sont donc significativement différentes au risque 5%.
N est moins efficace que R pour la baisse des taux de cholestérol (1,15<1,4) (A)
Avec des échantillons de 400 sujets (n>30), on ne se pose pas la question (E)

Avant une épidémie de grippe, on décide de vacciner 1800 personnes > 60 ans.

1) On les convoque 3 mois après cette vaccination. 900 viennent et 90 déclarent avoir eu la grippe.

Quelle est la proportion de sujets vaccinés qui ont eu la grippe?

Cette proportion sera estimée sur les 900 venus : p = 90/900 = 10%. Il s'agit d'une estimation ponctuelle :

Intervalle de confiance au risque α =5%

$$IC_{95\%} = [0{,}10 \pm 1{,}96 \sqrt{\frac{0{,}1 \times 0{,}9}{900}}] = 0{,}10 \pm 0{,}02$$

$$\mathbf{IC_{95\%} = [0{,}08 ; 0{,}12]}$$

2) Estimation correcte? Peut on obtenir une estimation plus satisfaisante?

Les 900 personnes qui se sont présentées ne sont pas représentatives des 1800 personnes vaccinées. Les gens qui ne sont pas venus sont soit :

- Non atteints de la grippe >> n'ont pas jugé utile de se déplacer
>>50% de non réponses

- Atteints de la grippe >> malades ou mécontents de la vaccination

Dans ces 2 cas, estimation de p incorrecte

Il faudrait envisager d'obtenir des nouvelles des 900 non venus par courrier, téléphone, auprès du médecin traitant.

3) On a appliqué cette méthode, et on a obtenu une estimation = 8%. Que pensez vous de l'efficacité du vaccin ?

Supposons que l'information (grippé ou non grippé) soit obtenue pour les 1800 personnes, et que l'estimation de p soit 8% :

- Meilleure procédure, estimation plus satisfaisante

- Par contre on ne peut pas conclure sur l'efficacité du vaccin, car il n'y a pas de population de référence, comparable dans sa composition et non vaccinée.

Il aurait fallu comparer les % de sujets grippés dans ces 2 populations afin de juger de l'efficacité du vaccin.

PLAN GÉNÉRAL DU COURS

1 - Biostatistique

2 - Statistique Descriptive

3 - Statistique Déductive

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

- **Observations, mesures**
- **Conclusions**
- **Les tests**

Objectifs pédagogiques :

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

Tirer des conclusions à partir d'observations

Exemple :
Comparer 2 groupes pour un caractère donné.

2 hypothèses :

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

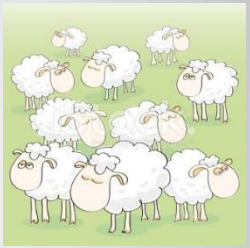
LES TESTS

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

Comparer 2 groupes pour un caractère donné.

Combien de litres de lait te donnent tes brebis chaque mois ? J'ai 65 brebis, et a peu près 1250 l

Moi c'est 1000 l pour 56 brebis



Ah j'ai beaucoup plus de lait que toi!!

Bien sur tu as plus de brebis!! Mais en proportion ?

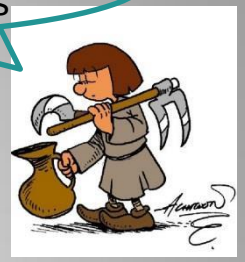
Quelle proportion ?
J'ai plus de lait que toi!!
C'est tout! Mauvais perdant!

Il m'énerve ! en moyenne 19 l par brebis pour lui et moi 17,8 l..
C'est presque pareil !!
Et si on refaisait l'expérience?

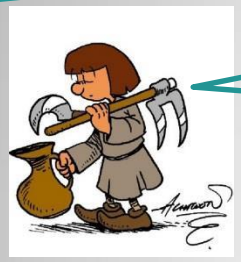
Un mois plus tard…

J'ai toujours 65 brebis, et 1210 l
Au lieu de 1250 mais c'est pareil..

Moi 1080 l au lieu de 1000,
et toujours 56 brebis



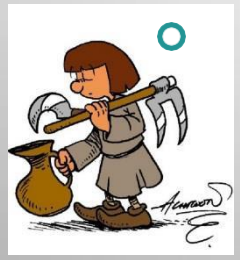
J'ai encore gagné!!



Mais non !!

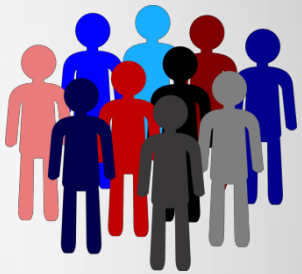
Il comprends rien ! en moyenne 18,6 l
par brebis pour lui et moi 19,3 l..
C'est bien ce que je pensais
Il y des fluctuations individuelles donc
la différence entre nos 2 moyennes
fluctue aussi

Je vais tenir compte des fluctuations
individuelles (xi – m)
par rapport à cette valeur moyenne.
Ca va m'aider à comprendre la
différence des 2 moyennes : H0 ou H1 ?
(J'ai fait Biostat en PACES..)



Population 1

Echantillon 1



Population 2

Echantillon 2

Si à **l'observation, pour un paramètre donné,** il semble exister une **différence ou au contraire une similitude entre les échantillons** :

Fluctuations d'échantillonnage ?

Différences entre populations ?

Hasard?

Test Statistique

Comparer 2 populations pour un paramètre donné qualitatif ou quantitatif

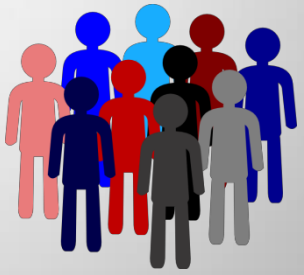
Hypothèse nulle (H0)

Paramètre population 1

Paramètre population 2

Echantillon 1

ESTIMATION



Echantillon 2

Paramètre échantillon 1

Paramètre échantillon 2

Comparer 2 populations pour un paramètre donné qualitatif ou quantitatif

Hypothèse nulle (H0)

Population 1

Population 2

Echantillon 1

Echantillon 2

Il n'y a pas de différence observée entre les 2 populations pour le paramètre étudié

Comparer 2 populations pour un paramètre donné qualitatif ou quantitatif

Hypothèse alternative (H1)

Paramètre population 1 ≠ Paramètre population 2

ESTIMATION

Echantillon 1

Echantillon 2

Paramètre échantillon 1 ≠ Paramètre échantillon 2

Comparer 2 populations pour un paramètre donné qualitatif ou quantitatif

Population 1

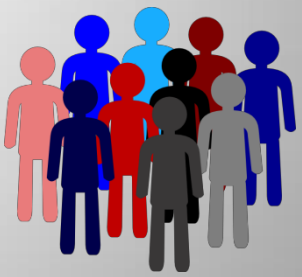
Hypothèse alternative (H1)

Population 2

Echantillon 1

Les 2 populations sont significativement différentes pour le paramètre étudié

Echantillon 2



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

Question simple à propos d'un problème médical.

- **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

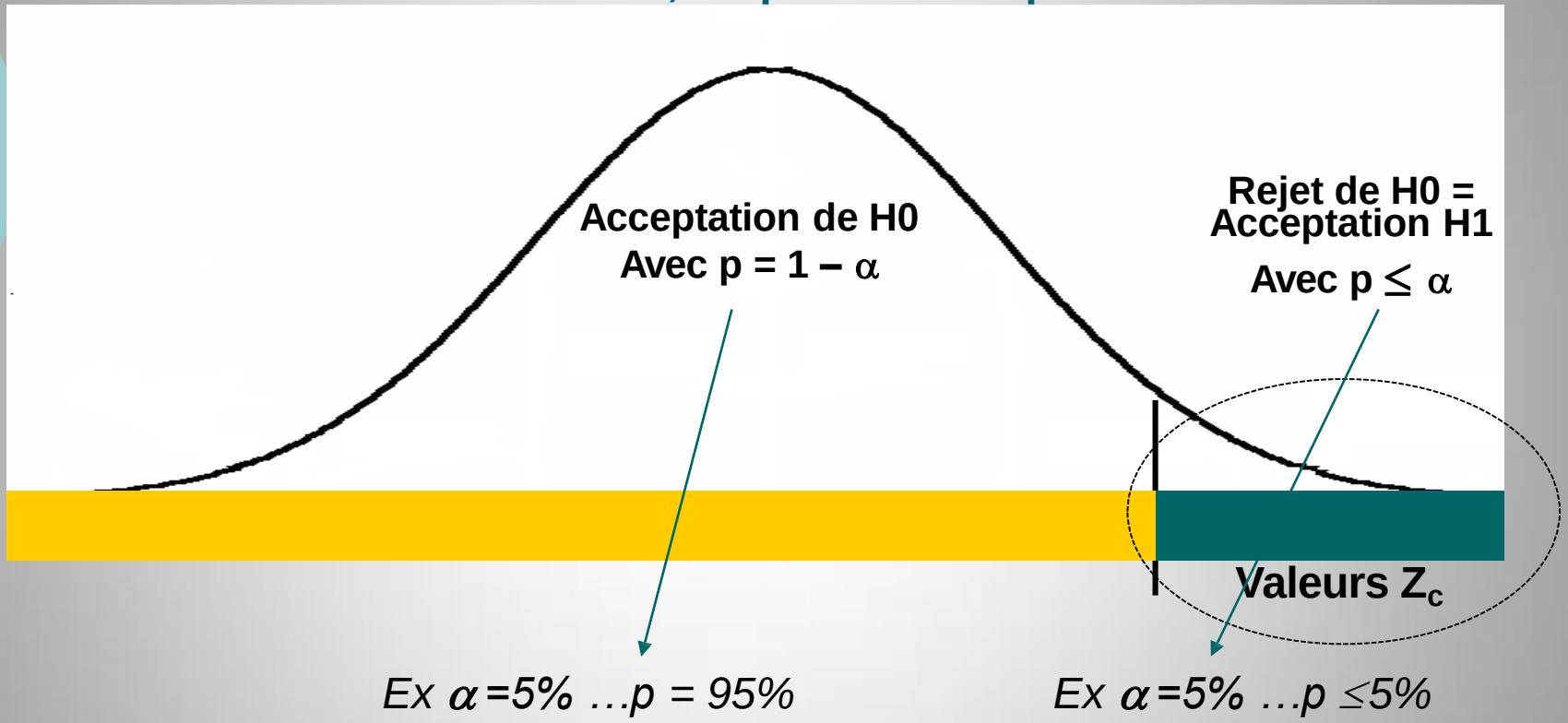
- **Etape 3** : Avant recueil des données on choisi **le risque** α (dans la pratique souvent 5%)

- **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. Fixation du risque d'erreur attaché à la conclusion..

- **Etape 5** : Interprétation des résultats.

Interprétation graphique du risque α

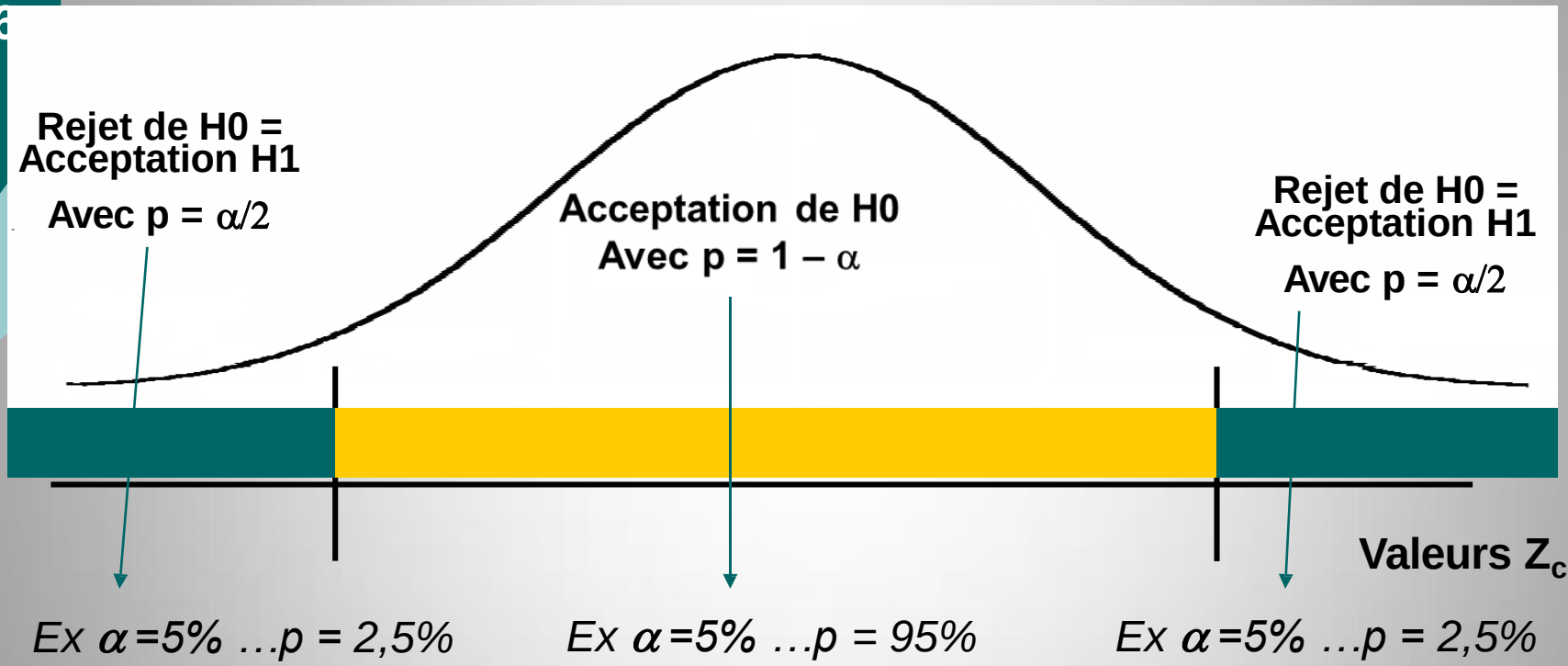
Le paramètre Zc résultat du test suit une distribution probabiliste en forme de **courbe de Gauss**. **Soit α, risque de 1ère espèce choisi = 5 %**



Situation unilatérale : **Les 2 situations observées sont elles différentes ?**

La seule réponse possible est OUI ou NON

Interprétation graphique du risque α



Ex α=5% …p = 2,5% *Ex α=5% …p = 95%* *Ex α=5% …p = 2,5%*

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

Si OUI, il est possible d'affirmer laquelle est la meilleure des 2.

Interprétation graphique du risque α

Situation unilatérale :

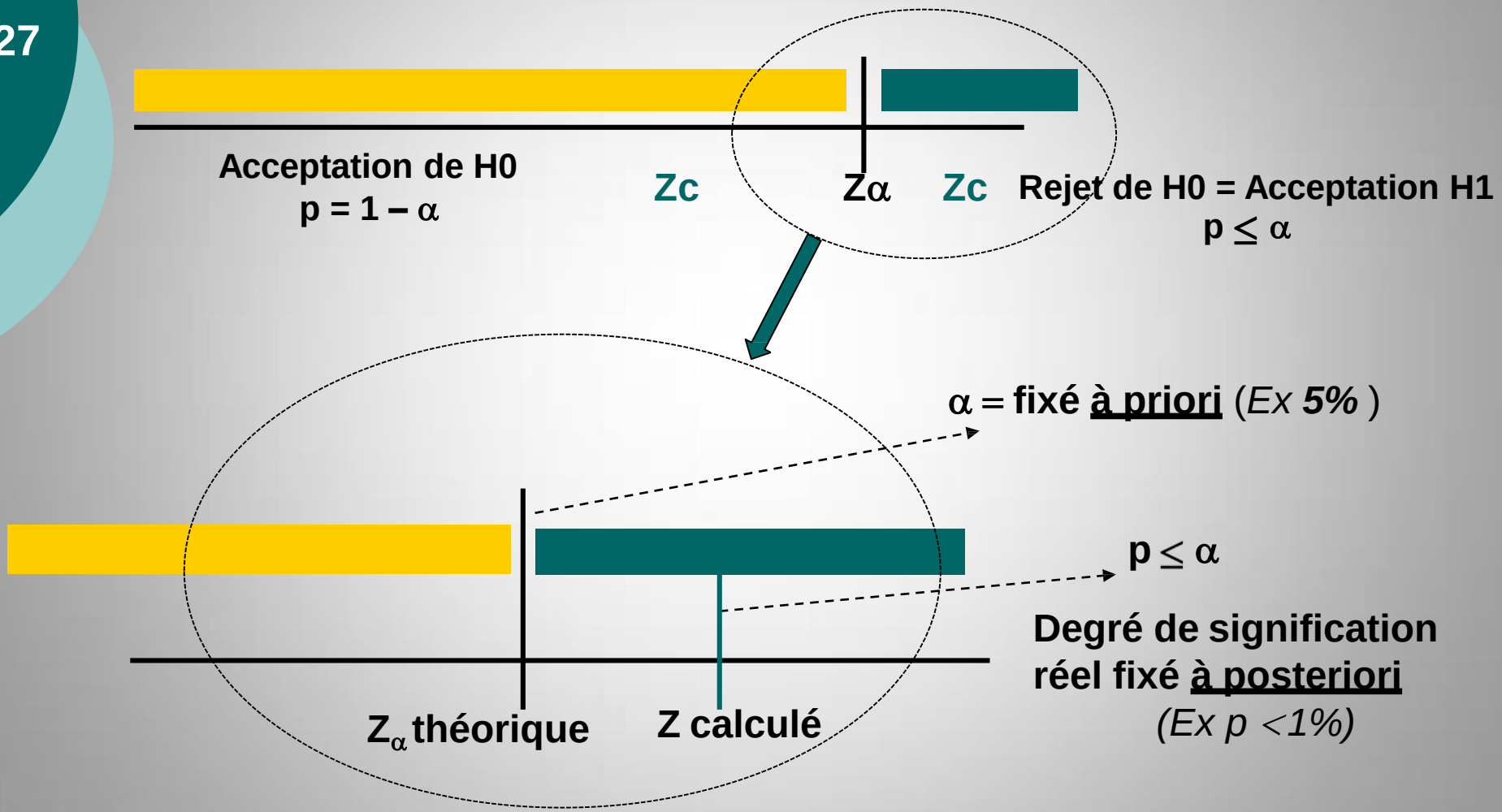


Table de l'écart réduit

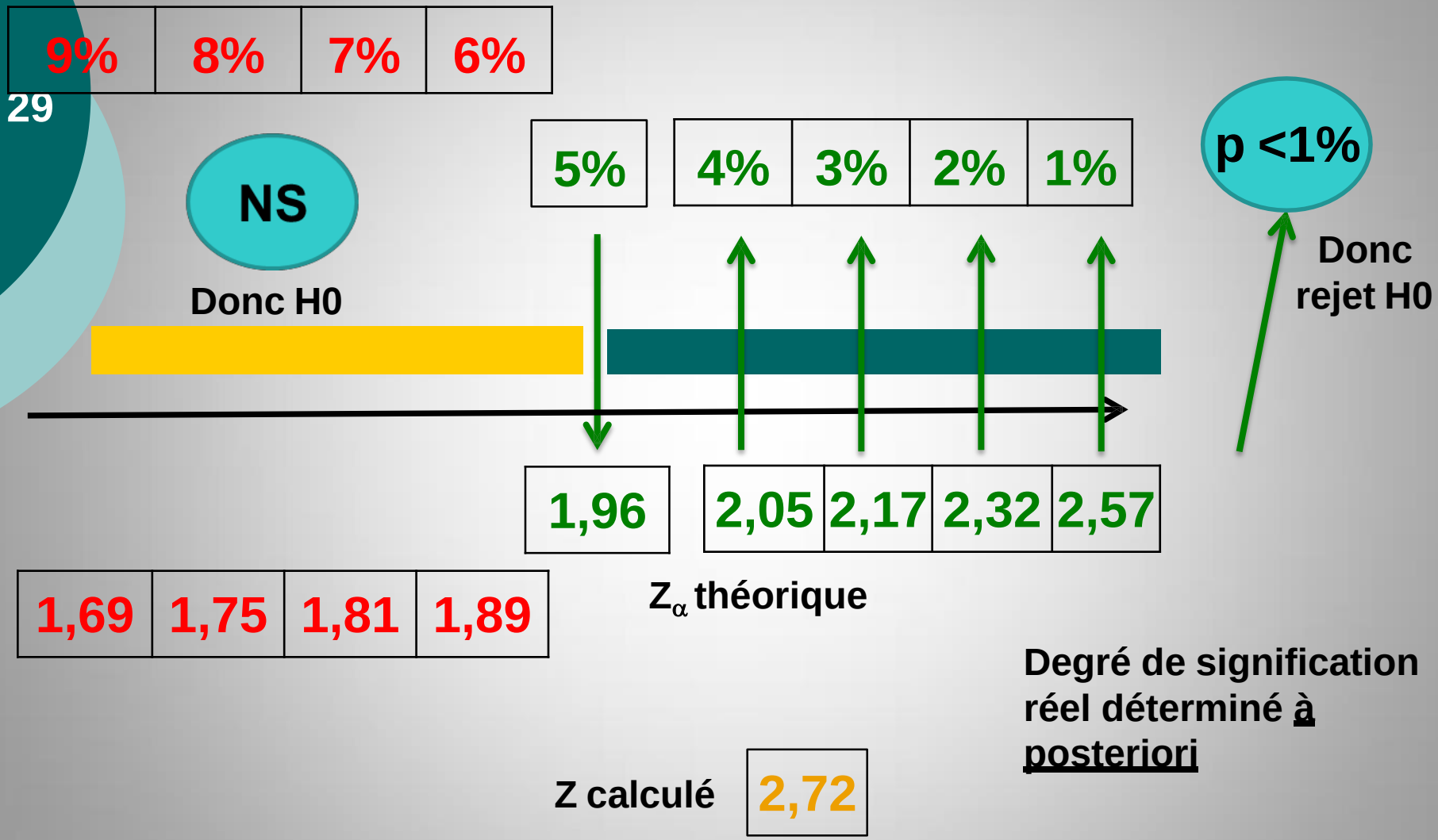
α

		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	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	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

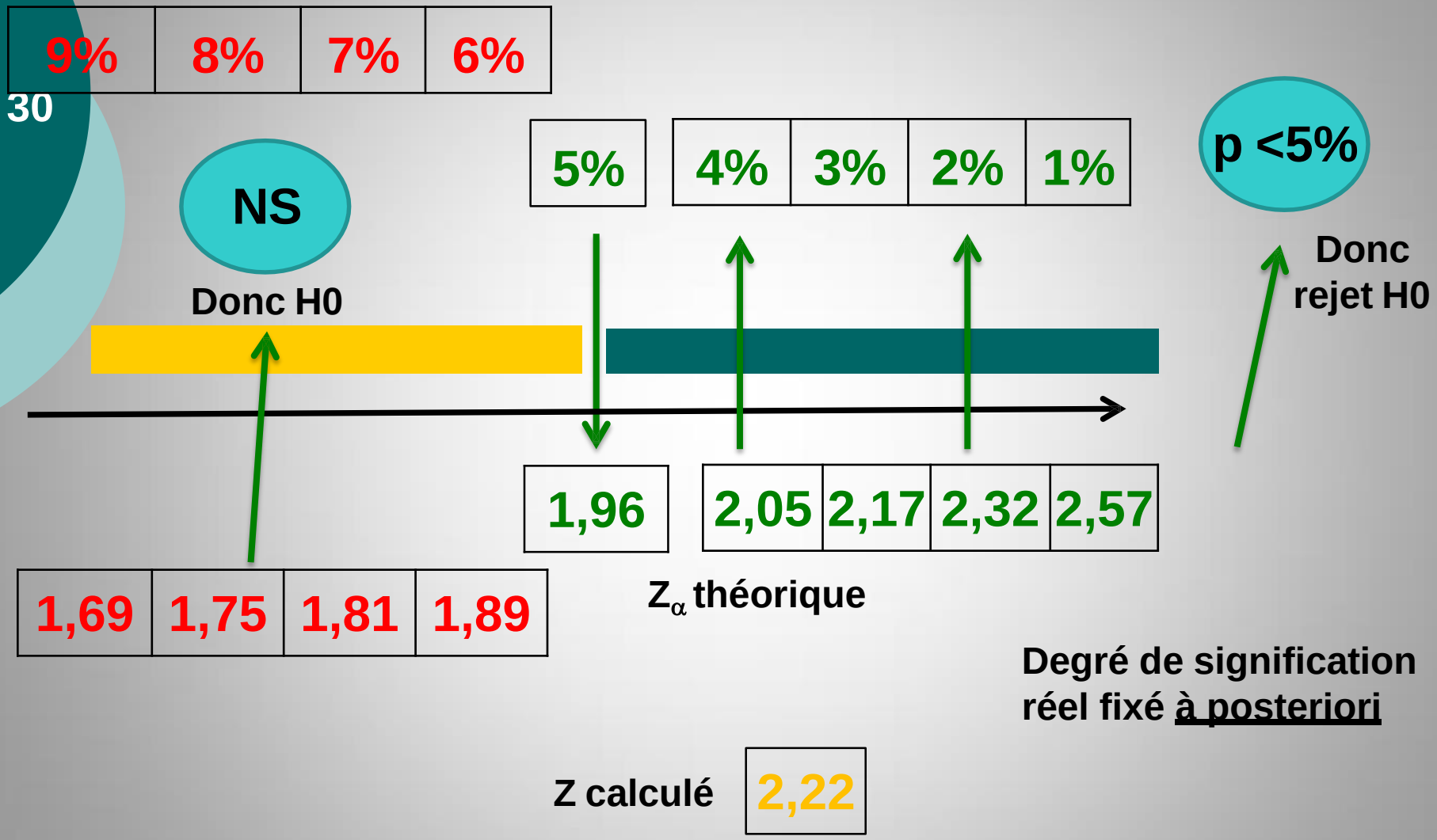
Interprétation graphique du risque α (2/3)



Degré de signification réel déterminé à posteriori

Z calculé 2,72

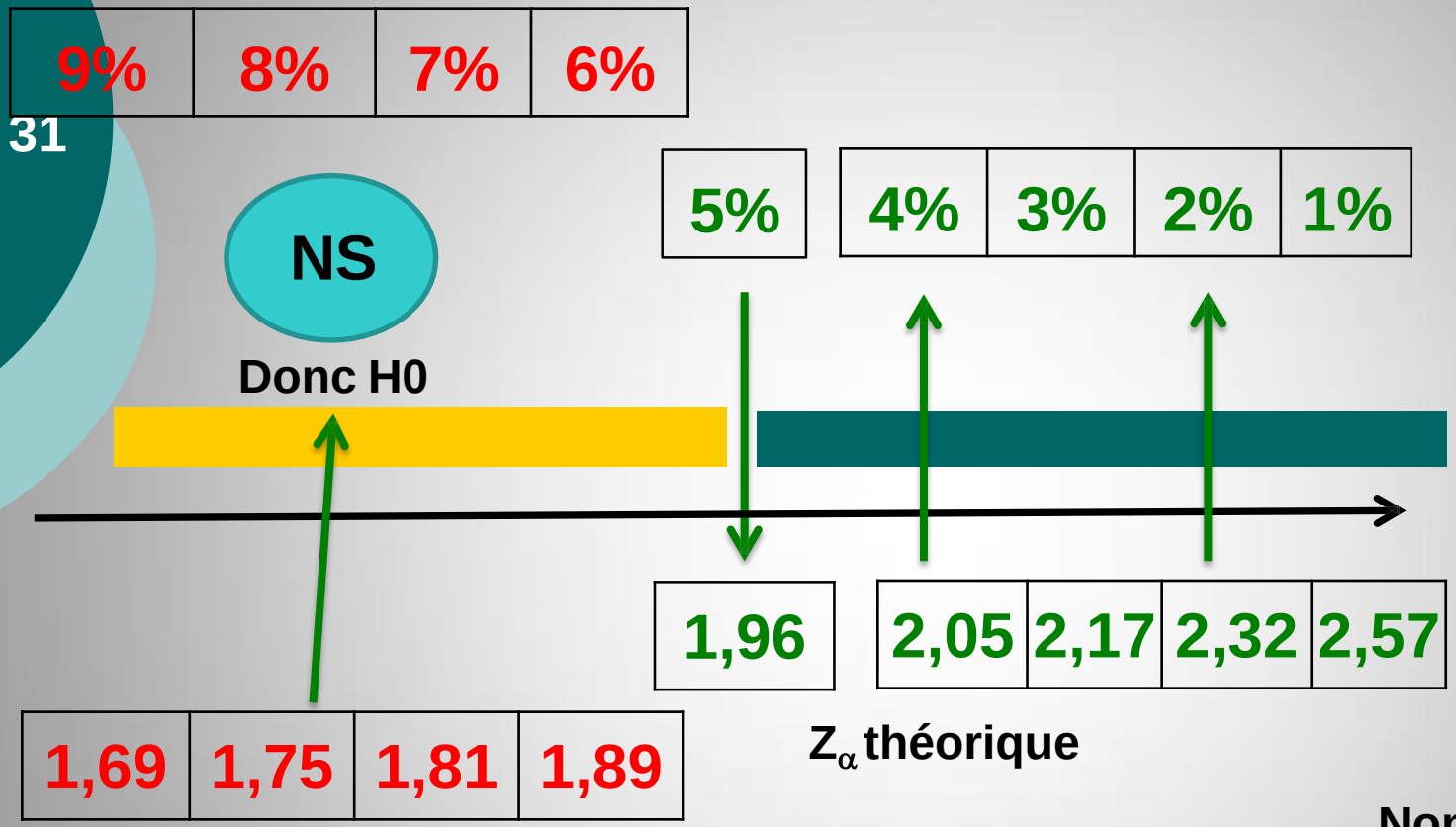
Interprétation graphique du risque α (3/3)



Degré de signification réel fixé <u>à posteriori</u>

Z calculé 2,22

Interprétation graphique du risque α (3/3)



Non significatif
<u>à posteriori</u>

Z calculé 1,3

Qu'appelle t on risque ?

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

compromis universel : $\alpha = 5\%$

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.

Les risques d'erreur

Réalité	Décision du statisticien : Rejet H0	Décision du statisticien : Non rejet H0
H0 Vraie	Erreur 1ère espèce α	1 - α
H1 Vraie	Puissance 1 - β	Erreur 2ème espèce β

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).

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.

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

36

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

« Le risque de seconde espèce »

A) Est noté β et vaut en général 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 A, C : C'est la définition de β

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

(D) La puissance est **1- β**

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

37

Pour une certaine maladie, un traitement ancien donnait de bons résultats dans 75% des cas. On teste un nouveau traitement N.
Sur 100 patients on mesure le % de succès de N.
On décide (bêtement !) de rejeter H0 qq soit ce %. α = ? Puissance = ?

α = Risque de rejeter H0 vraie. Rejet systématique de H0 donc α=100%
Puissance = proba de rejeter H0 fausse. Rejet systématique de H0 donc P =100%

On décide (tout aussi bêtement !) maintenant d'accepter H0 qq soit le %
α = ? Puissance = ?

On ne rejette jamais H0 ! En particulier quand elle est vraie donc α= 0%
Et on ne la rejette pas quand elle est fausse donc P =0%

BIG DATA (DONNEES MASSIVES)

Et si les <u>données</u> étaient le pétrole du 21ème siècle ?

Nous générons et détenons quantités d'info personnelles >>

Alimentation, achats, contributions réseaux sociaux, goûts, préférences, recherches sur Google, santé connectée,…

Données éparses mais captées par différents intervenants sur Internet.

Domaine de la santé :

Etudes épidémiologiques diverses lancées (pour le meilleur et pour le pire ?) : société privées (USA) analysent ces data et en tirent des conclusions : proposent à des femmes l'ablation des 2 seins car leur profil génétique comparé à celui de milliers d'autres femmes >> risque accru de K sein.

Les objets connectés (bracelets, balances, tee-shirts, fauteuils, iwatch,..). Suivre sa propre forme physique, la comparer à ce qu'elle devrait être (!). Mais alimentent aussi de manière continue ces fameuses Big Data.

L'utilisation de ces masses de données remet en cause certaines théories statistiques et la notion d'échantillonnage.

Jusqu'à aujourd'hui les données recueillies dans les études cliniques sont des données démographiques (sexe, âge), cliniques (poids, taille, diag, trait, dose, durée), biologiques,…Jamais de données de type psy, émotionnel, ..
Big Data : permettent de recouper et analyser TOUS ces types de données et de remettre en cause certaines conclusions ou décisions..

De plus : échantillon traditionnel = effectif de qq dizaines, au mieux qq centaines d'individus, représentant des populations cibles souvent de plusieurs centaines de milliers d'individus. ***Schéma le plus performant ?***

Grâce aux Big Data :
effectif de l'échantillon observé et étudié est de l'ordre de la population cible
Et çà, c'est tout de même un vrai bouleversement théorique !

PLAN GÉNÉRAL DU COURS

1 - Biostatistique

2 - Statistique Descriptive

3 - Statistique Déductive

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

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

- **Etape 1 : Avant recueil des données définir H0 et H1**
- **Etape 2 : définir le test en fonction du type des données (qualitatives, quantitatives). Soit Z le paramètre calculé**
- **Etape 3 : Avant recueil des données on choisi le risque a priori (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**
- **Fixation du risque d'erreur réel à posteriori**
- **Etape 5 : Interprétation des résultats.**

42

Etude de la liaison entre 2 caractères qualitatifs.

Question :

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

1) Comparaison de 2 pourcentages observés.

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

$\varepsilon = 1{,}96$ avec $\alpha = 5\ \%$

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)

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 ?

1. **H0 : Pas de différence entre les 2 modes de garde vis-à-vis des rhinopharyngites**
   H1 : Différence entre les 2 modes de garde

2. **Caractère qualitatif 1 : Garde en crèche ou à domicile**
   Caractère qualitatif 2 : Avoir une rhinopharyngite ou non
   Donc
   Test = Comparaison de pourcentages

3. **α = 5% défini à priori**

4. **p_A = 130/200 = 65%** **p_B = 96/200 = 48%**

$$\varepsilon = \frac{0,65 - 0,48}{\sqrt{\frac{0,65 \times 0,35}{n_A} + \frac{0,48 \times 0,52}{n_B}}} = 3,4$$

5. **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	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

Le test statistique vient de démontrer **sur cet échantillon**, que le risque de rhinopharyngites est supérieur chez les enfants gardés en crèche

(p <0,001) défini à posteriori

Conclusion :

Sur cet échantillon 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.