



Quelques notions de Statistique

Sloane ROLLIER
Interne en Santé Publique



Plan

- Les variables
- Généralités sur les tests comparatifs
- Risques d'erreur dans les tests comparatifs
- Degré de signification p
- Interprétation d'un test comparatif
- Exemple de test statistique : le test de chi-
2



Les variables

- **Différents types de variables**
- Variables quantitatives (valeurs numériques)
 - Continue (ex : âge, poids)
 - Discrète (ex : nombre d'enfant)
- Variables qualitatives
 - Dichotomique (ex : le sexe)
 - Ordinale (ex : le niveau d'études)
 - Nominale (ex : les groupes sanguins)



Les variables quantitatives (1)

- Variables quantitatives décrites par **des indice de position et de dispersion**
 - Exemple : les notes d'une classe
 - 11 11 12 13 14 16 16 18 19



Les variables quantitatives (2)

■ Indices de position :

- **Moyenne** : rapport entre la somme des valeurs et le nombre de valeurs
 - Total des notes / effectif total = $130/9=14,44$
- **Médiane** : valeur séparant les sujets en 2 groupes de taille identique (50/50)
 - 11 11 12 13 14 16 16 18 19 (nombre impair)
 - 10 11 11 12 13 14 16 16 18 19 (nombre pair)
- **Quartiles / déciles / percentiles**



Les variables quantitatives (3)

■ Indices de dispersion

- **Extrêmes** : 2 valeurs extrêmes de la distribution, valeurs minimum et maximum
- **Variance** : indice se basant sur la distance des valeurs par rapport à la moyenne (plus elle est élevée, plus les valeurs sont variables)
- **Écart-type** : racine carrée de la variance
- **Étendue** : différence entre la plus grande et la plus petite valeur



Les variables qualitatives

- **Pourcentage** : proportion d'une modalité multipliée par 100 (%)

$$P = \frac{n}{N} \times 100$$

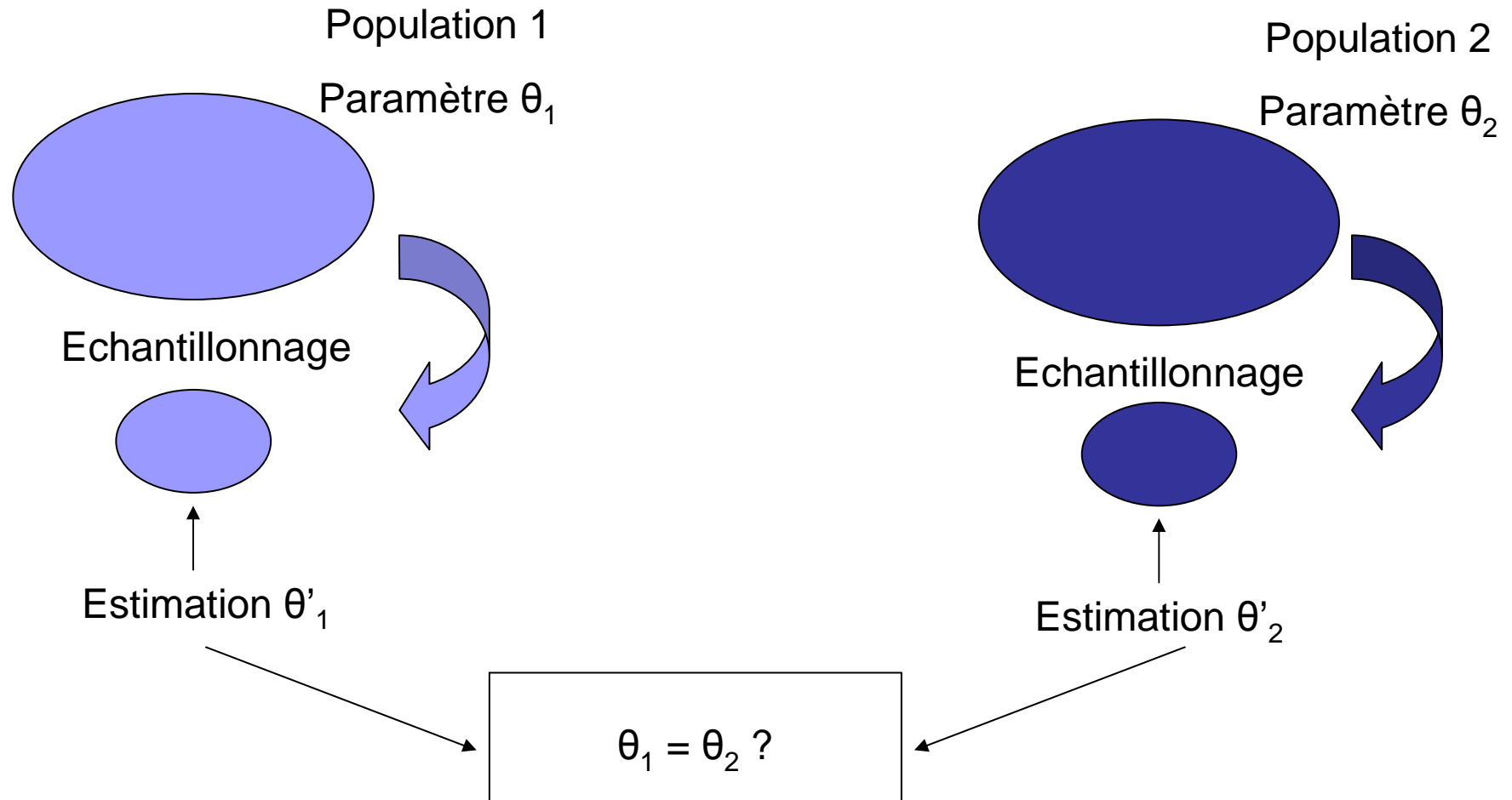
- **Variance**
- **Ecart-type**



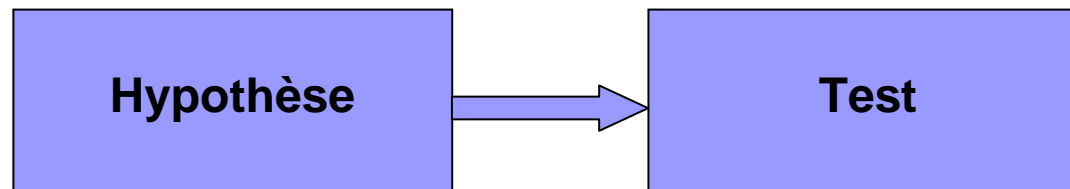
Généralités sur les tests comparatifs (1)

- Lorsqu'on compare 2 ou plusieurs séries de données, on observe une différence, plus ou moins grande entre les paramètres mesurés. Le but des tests est de déterminer si la différence est :
 - Due au hasard (i.e fluctuations d'échantillonnage)
 - Ou réelle
- Ils servent à extrapoler les résultats observés sur des échantillons à l'ensemble des populations dont ils sont issus

Généralités sur les tests comparatifs (2)



Généralités sur les tests comparatifs (3)



- **Hypothèse nulle (H_0)** : hypothèse que l'on cherche à réfuter ($H_0 : \theta_1 = \theta_2$)
- **Hypothèse alternative (H_1)** : hypothèse contraire de l'hypothèse nulle, celle que l'on cherche à démontrer (hypothèse retenue quand H_0 est rejeté)



Risques d'erreur dans les tests statistiques (1)

■ Risque d'erreur de 1^{ère} espèce α

- Probabilité de trouver une différence statistiquement significative alors qu'il n'y en a pas
- probabilité de rejeter à tort H_0 alors que H_0 est vrai
- α quasiment toujours fixé à 5%
- Exemple : considérer efficace un traitement qui ne l'est pas



Risques d'erreur dans les tests statistiques (2)

■ ***Risque d'erreur de 2^{ème} espèce β***

- *β = probabilité de rejeter à tort H_1 alors que H_1 est vrai*
- *Exemple : ne pas conclure alors que le traitement est efficace*

■ ***Puissance $1-\beta$***

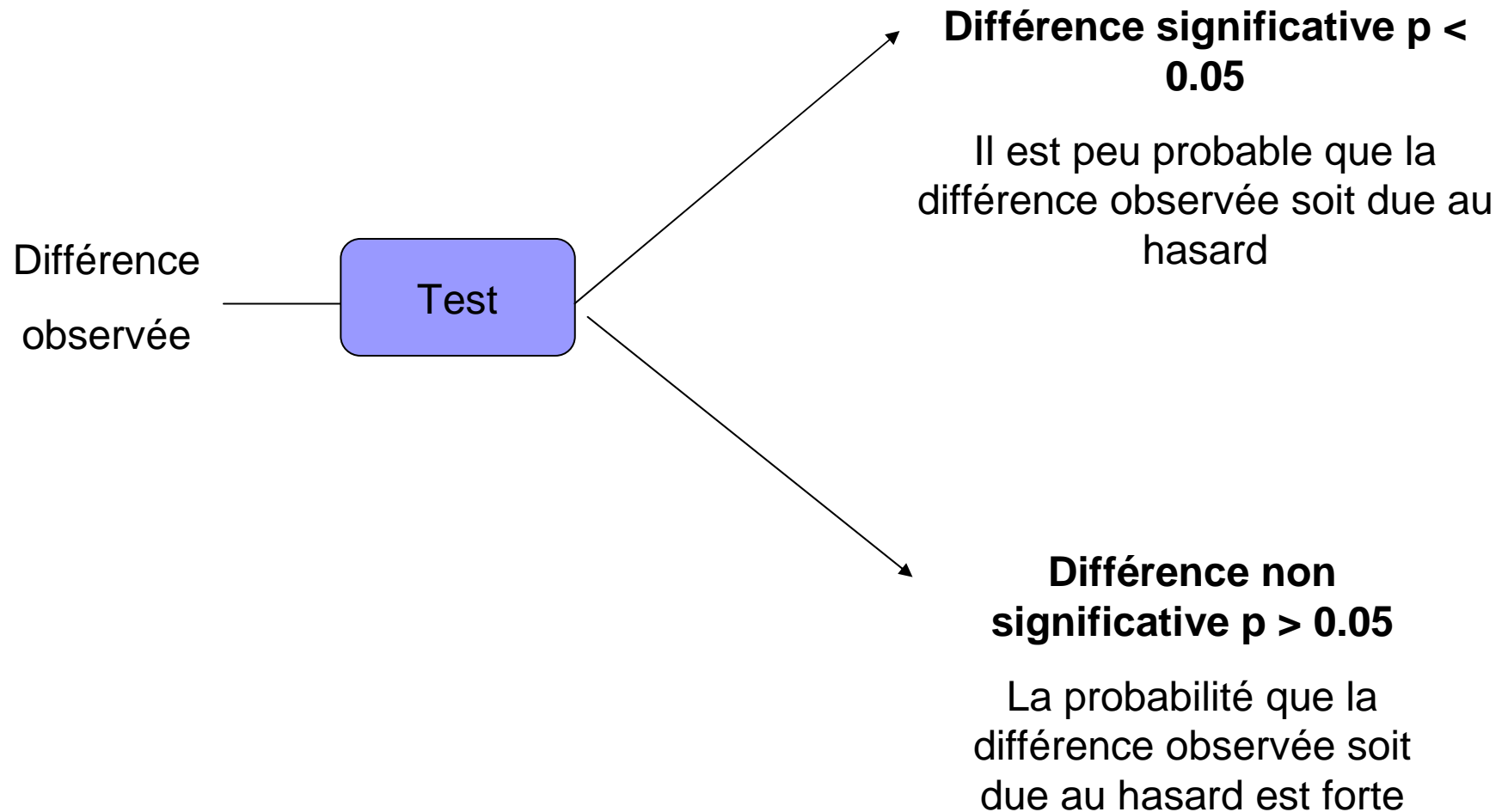
- *Probabilité de mettre en évidence une différence qui existe réellement*
- *Exemple : montrer l'efficacité d'un traitement réellement efficace*

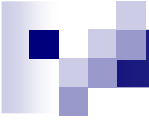


Degré de signification p (1)

- p quantifie le degré de désaccord entre l'hypothèse nulle et ce qui a été observé
- p diminue quand l'écart entre la réalité et H_0 est grand

Degré de signification p (2)





Interprétation d'un test comparatif

- Différence non significative
 - Absence de différence entre les 2 paramètres
 - Existence de différence mais qui n'a pas pu être découverte avec les données dont on dispose
- Différence significative
 - Existence de différence entre les paramètres (en prenant en compte le risque d'erreur)

Exemple de test statistique : le test de chi-2 (1)

- Il permet de comparer 2 ou plusieurs pourcentages

	Malades	Non malades	Total
Echantillon 1	O _a	O _b	N _{ech1}
Echantillon 2	O _c	O _d	N _{ech2}
Total	N _m	N _{nm}	N _{total}

- Question : le taux de prévalence de la maladie dans la population 1 est t-il différent de celui de la population 2?

Exemple de test statistique : le test de chi-2 (2)

■ Tableau des valeurs observées

	Malades	Non malades	Total
Echantillon 1	O_a	O_b	N_{ech1}
Echantillon 2	O_c	O_d	N_{ech2}
Total	N_m	N_{nm}	N_{total}

■ Tableau des valeurs calculées

	Malades	Non malades	Total
Echantillon 1	E_a	E_b	N_{ech1}
Echantillon 2	E_c	E_d	N_{ech2}
Total	N_m	N_{nm}	N_{total}

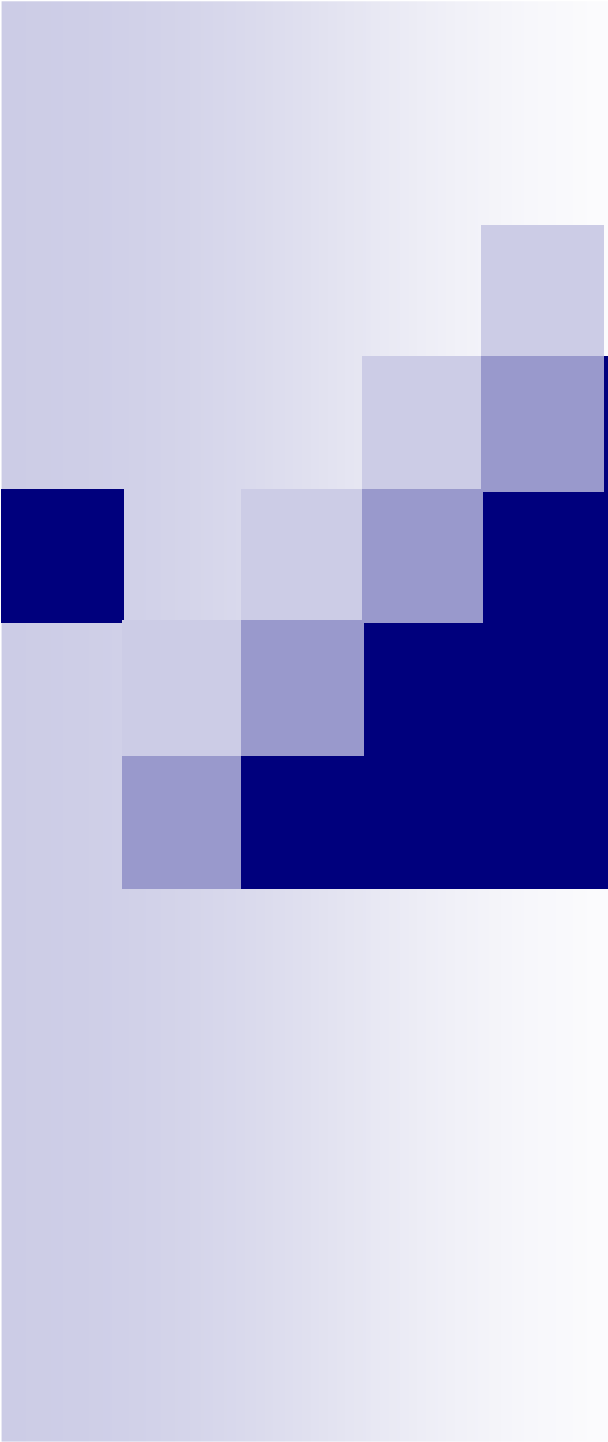
Exemple de test statistique : le test de chi-2 (2)

- Calcul de la différence entre les effectifs observés et attendus

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$



- (loi de chi-2 à 1 ddl)
- Comparaison de la différence à la valeur 3,84 qui correspond à α
 - Si $<3,84$: différence non significative
 - Si $>3,84$: différence significative



Représentativité et fluctuation



Représentativité

- Recueil exhaustif (population entière) systématisé limité à certains évènements/pathologies
- Données souvent « estimées » à partir d'un échantillon de la population d'origine ou source → « véritable » valeur inconnue
 - Risque : estimation biaisée



Estimation biaisée

- 2 mécanismes :
 - Des **fluctuations d'échantillonnage**
(variations aléatoires entre les échantillons)
 - Des **biais** (erreurs systématiques de mesure)



Biais

- **Biais de sélection** : erreur lors de la sélection des sujets
- **Biais de classement** : erreur de classement entre « malade/non malade » ou entre « exposés/non exposés »
- **Biais ou facteur de confusion** : variable liée à la maladie et au FdR pouvant être à l'origine d'une association statistique entre les 2.



Exemple

- **Healthy worker effect** : taux de mortalité des travailleurs moins élevé que dans la population générale
 - Population active → santé suffisante pour travailler
 - Affectation à un poste de W +/- exposé dépend de l'état de santé
- Type de biais ?
 - Risque de biais de sélection