

Objectifs du cours

- Exprimer une comparaison en termes statistiques
- Formuler une hypothèse à tester
- Interpréter les résultats d'un test
- Discuter le degré de signification d'un test

Plan

1. Hypothèses nulle et alternative
2. Le petit p et son interprétation

Mode de raisonnement :

1) Observation



2) Hypothèse



3) Test

Mode de raisonnement d'une comparaison

1) Observation

Tout part d'une observation de mesures qu'on a pratiqué sur plusieurs séries d'individus dans le cadre des études épidémiologiques

MESURES DU VOLUME GLOBULAIRE (VGM)
dans le personnel d'une usine chimique

1) groupe de 30 employés de bureau
moyenne VGM = $94,7 \mu\text{m}^3$

2) groupe de 30 ouvriers travaillant dans l'usine
moyenne VGM = $92,5 \mu\text{m}^3$

Problème ?

La moyenne du VGM est-elle réellement différente
entre employés de bureau et ouvriers ?

Exemple 2

Epidémie de gastro-entérite

	Malades	Sains	Taux d'attaque
Tomates +	14	38	26,9 %
Tomates -	4	17	19,0%

Problème ?

Avait-t-on vraiment plus de risques de contracter
une gastro-entérite après avoir mangé des tomates?

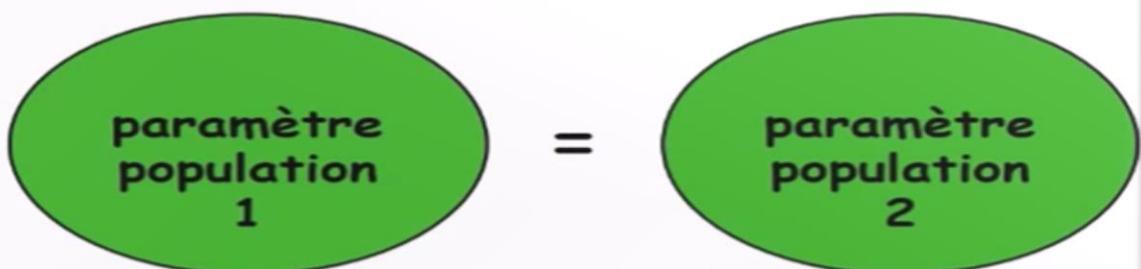
Ce que vous voulez savoir :

- 1) La moyenne du VGM est-elle réellement différente entre employés de bureau et ouvriers ?
- 2) Le taux d'attaque de gastro-entérite est-il plus élevé chez les mangeurs de tomates?

Y a -t-il vraiment une différence
ou est-ce simplement dû au hasard?

☒ Tests statistiques

Hypothèse nulle H_0



quelque soit le test , la démarche est toujours la même : on commence par formuler :

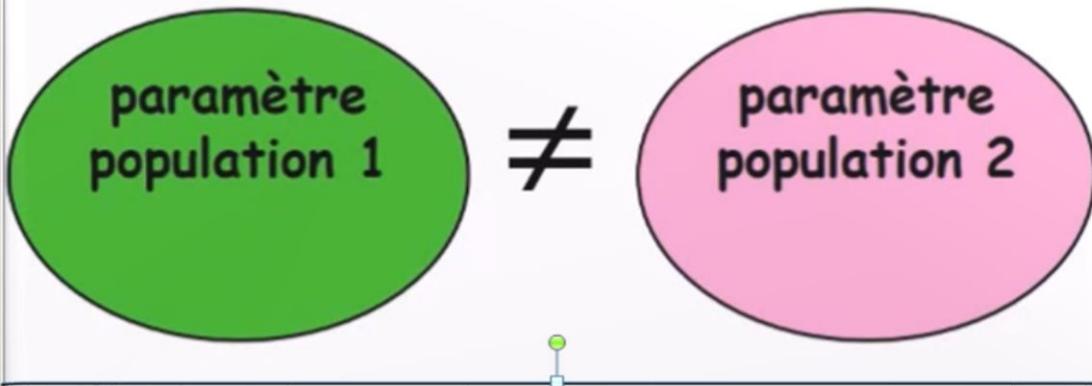
-l'hypothèse nulle : H_0

n' y a aucune différence non pas entre les échantillons mais entre les 2 populations dont ils ont issus ,

Les paramètres sont identiques dans les 2 populations .

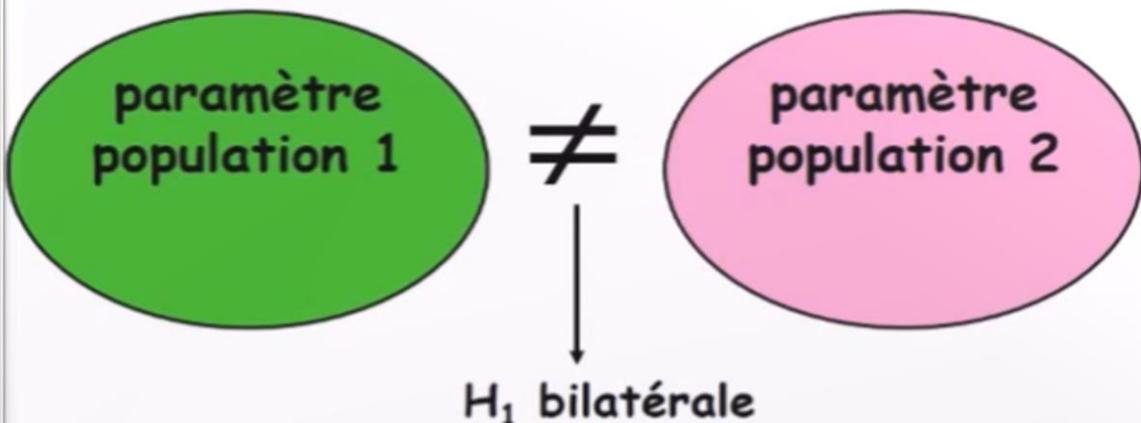
Les 2 échantillons proviennent de la même population

Hypothèse alternative H_1



-l'hypothèse alternative : **H1**
les paramètres des populations dont sont issus les échantillons sont différents : Ils proviennent de 2 populations différentes

Hypothèse alternative H_1



On ne sait pas à priori dans quel sens bascule la différence, d'où de préférence garder **H1 bilatérale**

Principe général d'un test statistique

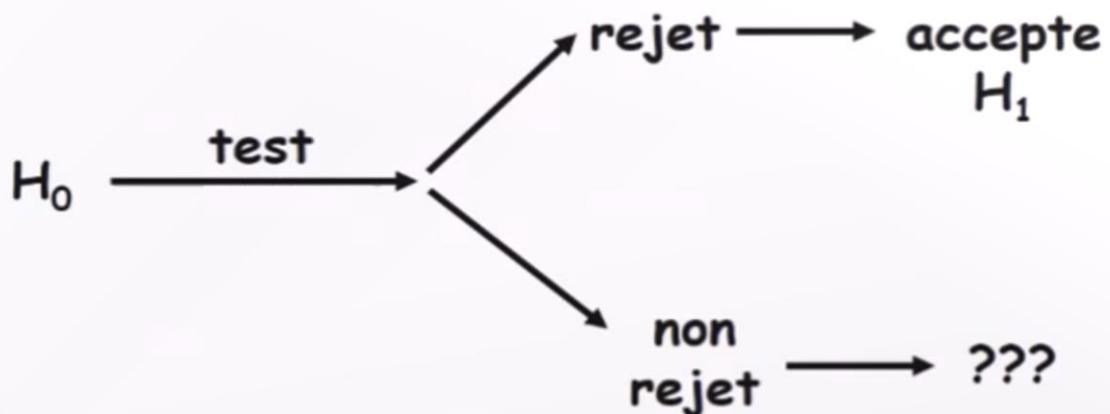
1. C'est l'hypothèse nulle H_0 qu'on teste
2. Si H_0 n'est pas vérifiée,
 - a) on la rejette
 - b) on accepte l'hypothèse alternative H_1

Importance d'écrire clairement l' H_1 , car son acceptation devient automatique

3. Sinon, on ne peut pas conclure.

(on n'accepte jamais une hypothèse nulle)

Principe général d'un test statistique



Mode de raisonnement :

1) Observation



2) Hypothèse



3) Test

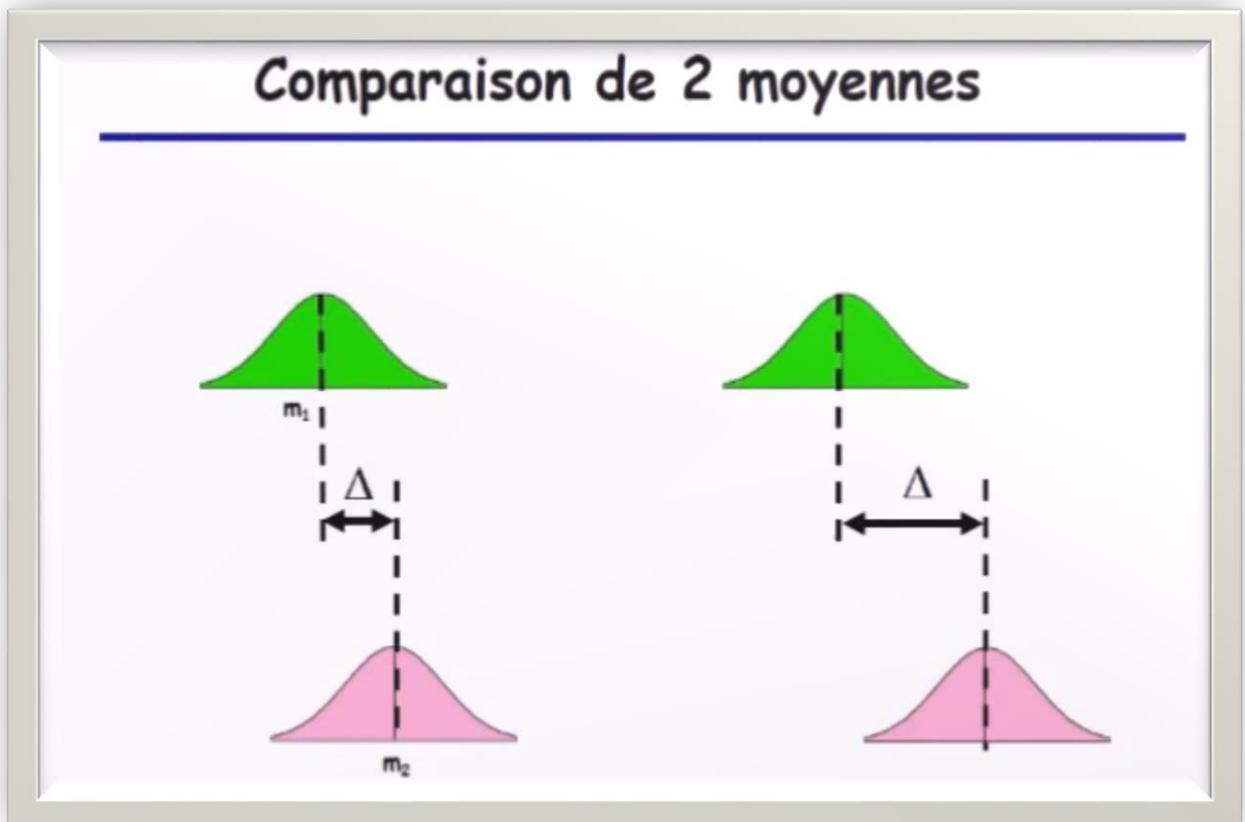
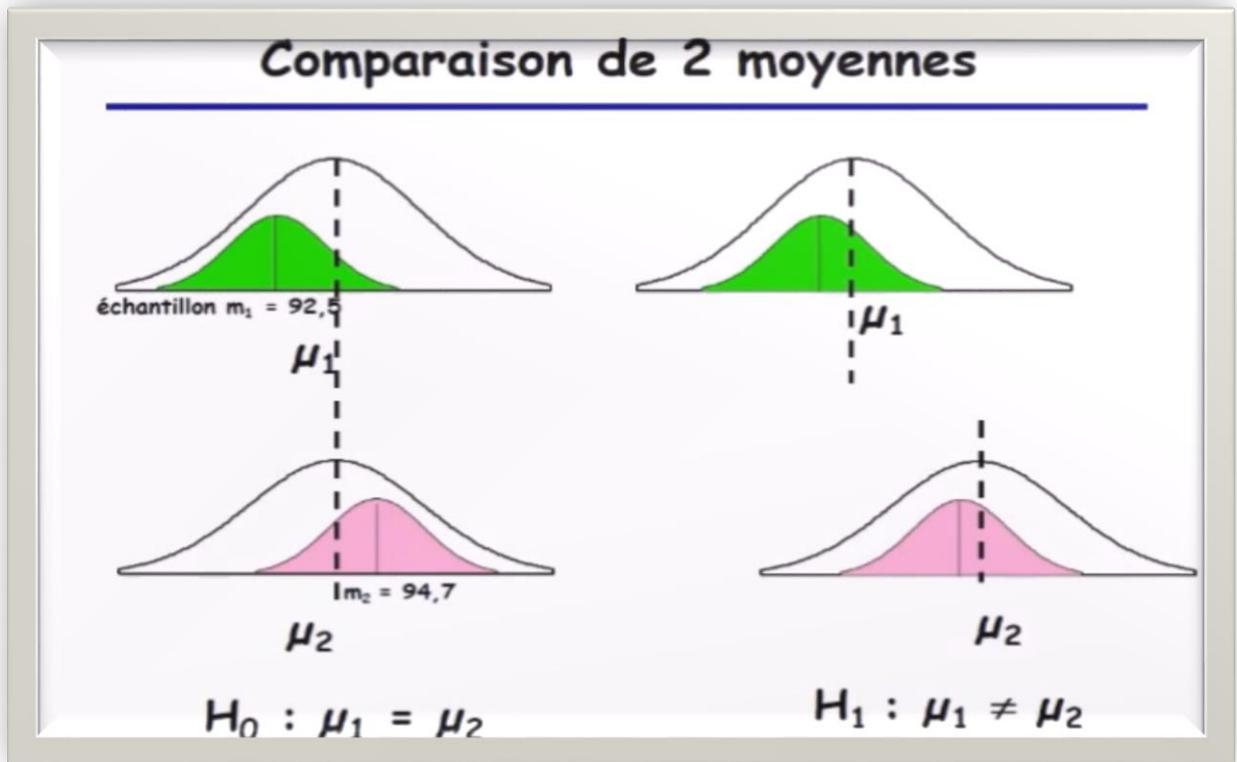
Exemple 1

Le VGM diffère-t-il
entre employés et ouvriers ?

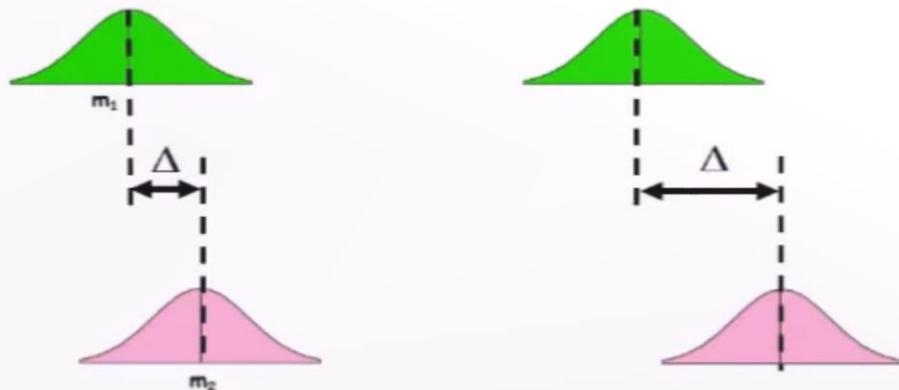
Employés	VGM	$m_1 = 94,7$
Ouvriers	VGM	$m_2 = 92,5$

H_0 : VGM_{population employés} = VGM_{population ouvriers}

H_1 : VGM_{population employés} ≠ VGM_{population ouvriers}



Comparaison de 2 moyennes



Si H_0 vraie : $\Delta = m_1 - m_2 \approx 0$

Exemple 1

Le VGM diffère-t-il
entre employés et ouvriers ?

Employés	VGM	$m_1 = 94,7$
Ouvriers	VGM	$m_2 = 92,5$

H_0 : VGM_{population employés} = VGM_{population ouvriers}

H_1 : VGM_{population employés} \neq VGM_{population ouvriers}

$$\Delta = 94,7 - 92,5 = 2,2$$

Est-ce que 2,2 est suffisamment éloigné de zéro
pour qu'on rejette H_0 ?

Exemple 2

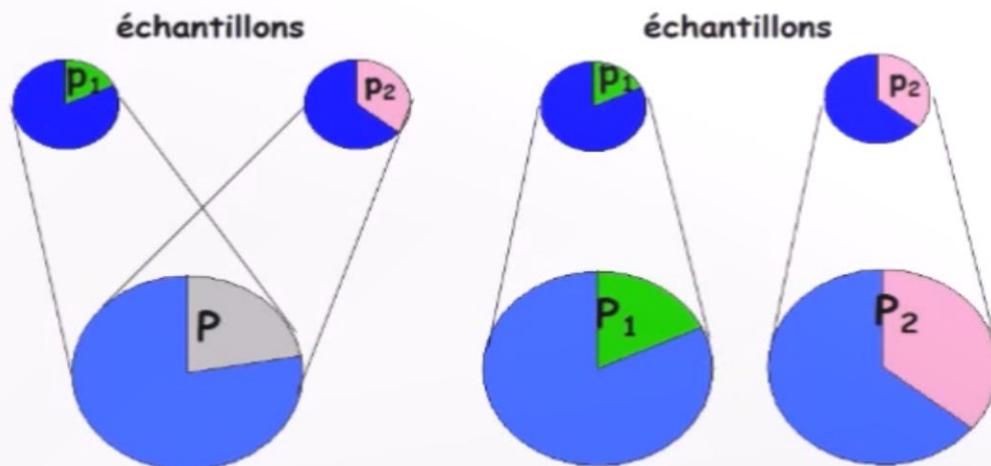
La fréquence de gastro-entérite est-elle plus élevée chez les mangeurs de tomates ?

Mangeurs de tomates : taux d'attaque = 26,9 %
Non-mangeurs : taux d'attaque = 19,0 %

H_0 : taux d'attaque_{mangeurs} = taux d'attaque_{non-mangeurs}

H_1 : taux d'attaque_{mangeurs} \neq taux d'attaque_{non-mangeurs}

Comparaison de 2 pourcentages



$H_0 : P_1 = P_2$

$H_1 : P_1 \neq P_2$

exemple 5

La fréquence de gastro-entérite est-elle plus élevée chez les mangeurs de tomates ?

Mangeurs de tomates : taux d'attaque = 26,9 %
Non-mangeurs : taux d'attaque = 19,0 %

H_0 : taux d'attaque_{mangeurs} = taux d'attaque_{non-mangeurs}

H_1 : taux d'attaque_{mangeurs} \neq taux d'attaque_{non-mangeurs}

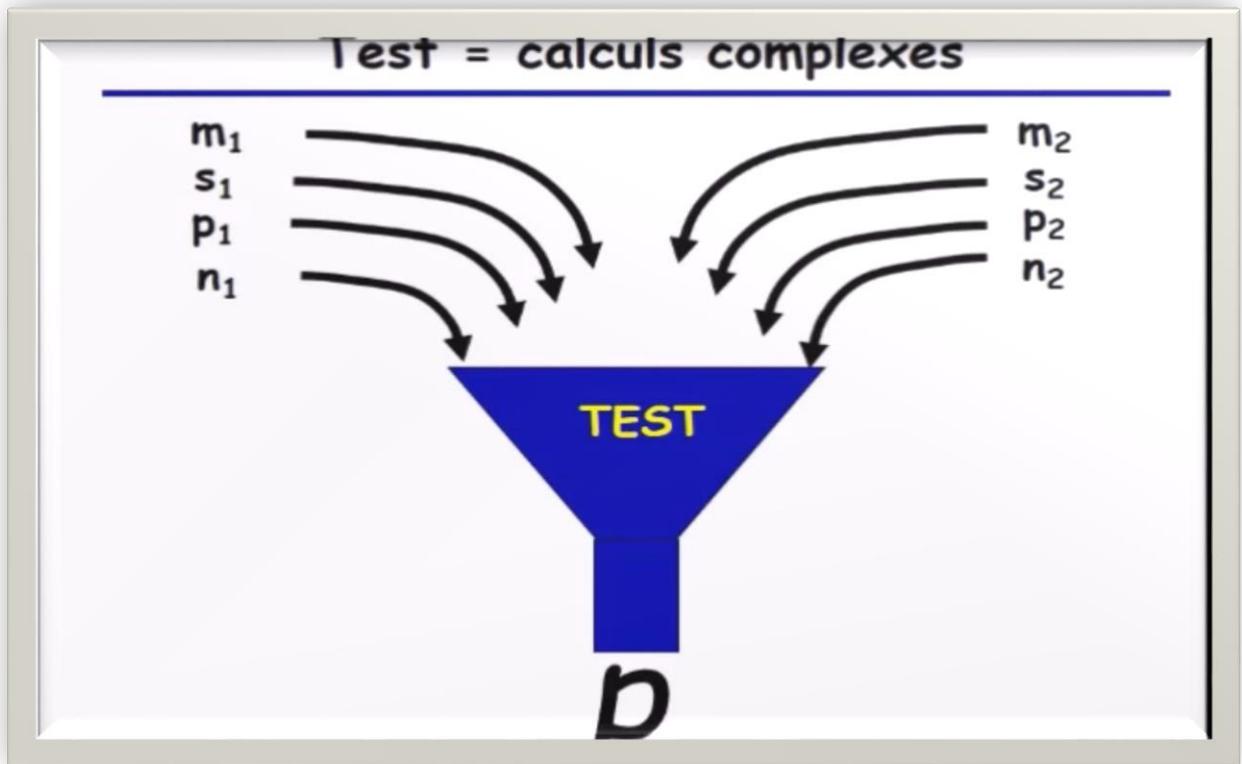
$$\Delta = 26,9 \% - 19,0 \% = 7,9 \%$$

Est-ce que 7,9 est suffisamment éloigné de zéro pour qu'on rejette H_0 ?

Tous les tests statistiques, quels qu'il soient, calculent

la probabilité de se tromper en rejetant H_0

le petit p



Exemple 1

Le VGM diffère-t-il
entre employés et ouvriers ?

Employés $m_1 = 94,7 \mu\text{m}^3$

Ouvriers $m_2 = 92,5 \mu\text{m}^3$

Test de comparaison de 2 moyennes

→ $p = 0,0042$

p : On a une probabilité de 0,0042 de se tromper si on rejette l'hypothèse nulle

Exemple 2

La fréquence de gastro-entérite est-elle plus élevée chez les mangeurs de tomates ?

Mangeurs de tomates : taux d'attaque = 26,9 %
Non-mangeurs : taux d'attaque = 19,0 %

Test de comparaison de 2 pourcentages

→ $p = 0,56$

p : On a une probabilité de 0,56 (56 chances /100) de se tromper si on rejette l'hypothèse nulle

CONCLUSION

Signification de p

On le compare au risque limite
qu'on se donne *a priori*

Risque α

Son seuil habituel est fixé à

5 %

α : Risque maximum de se tromper qu'on accepte lorsqu'on rejette l' H_0

Comparaison de p à la valeur seuil 5%

1) Si p est inférieur à 5%

- on rejette H_0
- on accepte H_1

On affirme que la différence observée
est significative

Exemple 1

Le VGM diffère-t-il
entre employés et ouvriers ?
Conclusion du test statistique

Employés VGM $m_1 = 94,7$

Ouvriers VGM $m_2 = 92,5$ ($p < 0,0042$)

Il existe une différence significative
entre le VGM des ouvriers et des employés

$p < 1$ pour cent

Comparaison de p à la valeur seuil 5%

2) Si p est supérieur à 5%

- on ne rejette pas H_0

On ne conclut pas mais on dit que
la différence observée n'est pas significative

Exemple 2

La fréquence de gastro-entérite est-elle plus élevée chez les mangeurs de tomates ?

Mangeurs de tomates : taux d'attaque = 26,9 %
Non-mangeurs : taux d'attaque = 19,0 %

$$p = 0,56$$

La différence entre les deux taux d'attaque
n'est pas significative.

On ne peut pas incriminer les tomates

Différentes expressions de p

- $p < 0,05$

- $p < 5 \%$

- $p < 0,001$

- $p < 10E-3$

- $p < 10E-7$

- $p < 0,049$

On peut rejeter l' H_0

- $p = 0,45$

- $p < 0,45$

- $p : NS$

- $p = 0,051$

Si $p > 5\%$, on met directement $p : NS$