Ministère de l'Enseignement Supérieur et de la recherche Scientifique
Université Mohamed Boudiaf M'sila
Faculté des Sciences
Département de Microbiologie & Biochimie

وزارة التعليم العالي و البحث العلمي
جامعة المسيلة
كلية العلوم
قسم الميكروبيولوجيا و البيوكيمياء

Année universitaire : 2019/2020

Matière : Validation des méthodes analytiques

Master 1 : NSA

Dr BOUAOUDIA-MADI

Chapitre 2 : Comparaison de deux moyennes

Plan du cours

I- Grands échantillons→ Test de l'écart réduit
II- Petits échantillons→ Test de Student
III- Comparaison de deux variances → Test F de Snedecor

1. Test Z (Ɛ) ou de l'écart réduit

Le test Z : comparer des paramètres en testant leurs différences
Utilisé pour comparer :
– Une moyenne observée à une moyenne théorique
– Deux moyennes observées

• Principe du test Z

Deux paramètres de 2 échantillons que l'on désire comparer :
– **H0 :** Les paramètres des populations d'où sont issus les 2 échantillons sont identiques
– **H1 :** Les paramètres sont différents

- On compare les 2 paramètres par leur différence Δ

• Δ est une variable aléatoire
• Si H0 est vrai alors Δ est proche de 0 :
– Si H0 est vrai et que l'échantillon est de taille suffisante
– La division de Δ par son écart type suit une loi Z normale centrée réduite de moyenne 0 et d'écart type 1

- **Le test Z consiste :**

– à estimer l'écart type de la différence sd
– à calculer l'écart réduit zo = |Δ| / sd
– à comparer cette valeur à la distribution théorique de la loi Z
– On utilise la table Z
• **Condition d'application :**
Effectif de chaque échantillon ≥ 30

• Interprétation du test Z

- **Au risque α = 5% :**

• Si la valeur observée zo < 1,96→ on ne rejette pas H0

→ On ne peut pas affirmer que les échantillons proviennent de populations différentes

→ la différence entre les paramètres n'est pas significative

• Si la valeur observée zo ≥ 1,96→ on rejette H0

→ On accepte H1 en affirmant que les échantillons proviennent de populations différentes

→ On affirme que la différence entre les paramètres est significative

- **Utilisation du test Z**

1- Comparaison d'une moyenne observée à une moyenne théorique

On compare une moyenne observée dans un échantillon à une moyenne connue dans la population de référence

• Variable quantitative

• Paramètre étudié moyenne

• Hypothèses

– H0 : M = μ

– H1 : M ≠ μ

μ : moyenne théorique connue de la population de référence

M : moyenne inconnue de la population d'où est issu l'échantillon

• Conditions d'applications:

Taille de l'échantillon ≥ 30

• Calcul de Z

$$z = \frac{|m - \mu|}{\frac{s}{\sqrt{n}}}$$

Formulation

– μ : moyenne théorique connue de la population de référence

– m : moyenne observée de l'échantillon

– s : écart type de l'échantillon

– n : effectif

Interprétation

Zo < 1,96 → Ho non rejetée → M n'est pas significativement différente de μ

Zo ≥ 1,96 → Ho est rejetée → M diffère significativement de μ

Exemple n°1:

Lors d'une enquête sur la durée de sommeil des enfants de 2 à 3 ans dans un département français, on a trouvé une moyenne du temps de sommeil par nuit de 10,2 heures dans un groupe de 40 enfants. L'écart type est 2,1 heures.

La moyenne du temps de sommeil est de 11,7 heures chez les enfants de cet âge.

- La durée de sommeil des enfants de ce département diffère-t-elle du temps de sommeil des enfants de cet âge?

Solution n°1:

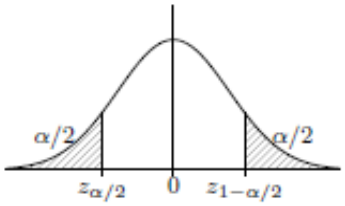
- **H0 :** les enfants de ce département dorment autant que ceux de la population
- **H1:** la durée de sommeil des enfants de ce département est différente
- **Z**o = (11,7 – 10,2)/(2,1/√ 40) = **4,5**
- 4,5 > 1,96 →**On rejette H0** → DS

La population des enfants examinés présente un temps de sommeil **significativement différent** de la population générale.

3° *Quantiles de la loi Normale (bis).* — Si Z est une variable aléatoire suivant la loi normale $\mathcal{N}(0,1)$, la table donne, pour α fixé, la valeur $z_{1-\alpha/2}$ telle que

$$\mathbb{P}\{|Z| \geqslant z_{1-\alpha/2}\} = \alpha.$$

Ainsi, $z_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,5758	2,3263	2,1701	2,0537	1,9600	1,8808	1,8119	1,7507	1,6954
0,1	1,6449	1,5982	1,5548	1,5141	1,4758	1,4395	1,4051	1,3722	1,3408	1,3106
0,2	1,2816	1,2536	1,2265	1,2004	1,1750	1,1503	1,1264	1,1031	1,0803	1,0581
0,3	1,0364	1,0152	0,9945	0,9741	0,9542	0,9346	0,9154	0,8965	0,8779	0,8596
0,4	0,8416	0,8239	0,8064	0,7892	0,7722	0,7554	0,7388	0,7225	0,7063	0,6903
0,5	0,6745	0,6588	0,6433	0,6280	0,6128	0,5978	0,5828	0,5681	0,5534	0,5388
0,6	0,5244	0,5101	0,4959	0,4817	0,4677	0,4538	0,4399	0,4261	0,4125	0,3989
0,7	0,3853	0,3719	0,3585	0,3451	0,3319	0,3186	0,3055	0,2924	0,2793	0,2663
0,8	0,2533	0,2404	0,2275	0,2147	0,2019	0,1891	0,1764	0,1637	0,1510	0,1383
0,9	0,1257	0,1130	0,1004	0,0878	0,0753	0,0627	0,0502	0,0376	0,0251	0,0125

α	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
$z_{1-\alpha/2}$	3,2905	3,8906	4,4172	4,8916	5,3267	5,7307	6,1094

Exemples. — Pour $\alpha = 0{,}5$, on trouve $z \approx 0{,}6745$; pour $\alpha = 0{,}25$, on trouve $z \approx 1{,}1503$; pour $\alpha = 10^{-6}$, on trouve $z \approx 4{,}8916$.

2- Comparaison de deux moyennes observées

On veut comparer **les moyennes observées** dans deux échantillons

• Paramètre étudié moyennes

• **Hypothèses**

– H0 : µ1=µ2

– H1 : µ1≠ µ2

µ1et µ2 : moyennes inconnues des deux populations d'où sont tirés les échantillons

- **Conditions d'application :**

Effectif de chaque échantillon ≥ 30

Calcul :

$$z = \frac{|m_1 - m_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Formulation

– m1 et m2 : moyennes observées des 2 échantillons
– s² 1 et s² 2 : variances des 2 échantillons
– n1 et n2 : effectifs des 2 échantillons

Interprétation

Zo < 1,96 → Ho non rejetée → μ_1 n'est pas significativement différent de μ_2
Zo ≥ 1,96 → Ho est rejetée → μ_1 diffère significativement de μ_2

Exemple n°2:

On désire comparer la pression artérielle diastolique d'un groupe de sujets sains et d'un groupe de sujets atteints de drépanocytose. Une étude donne les résultats suivants :

	Effectif (n)	Pression artérielle diastolique	Variance (s²)
Sujets sains	88	70,1	10,8
Sujets drépanocytaires	85	61,8	6,9

- La pression artérielle des sujets drépanocytaires diffère t- elle de celle des sujets sains ?

Solution n°2:

- H0 : les pressions artérielles sont identiques
- H1 : la pression artérielle est différente chez les sujets drépanocytaires
- Sd = √(10,8/88) + (6,9/85) = 0,45
- Zo = (|70,1 – 61,8|) / 0,45 = 18,4
- 18,4 > 1,96 : on rejette H0 → DS

La pression artérielle des sujets drépanocytaires est significativement différente de celle des sujets sains.

%§ § <W

2. Test de Student

Lorsque la taille des échantillons est faible (n<30) le rapport entre les différences de leurs moyennes et l'écart type ne suit pas une loi normale centrée réduite Z
On utilise alors le test T de Student

• Le test de Student sert à comparer :
– Une moyenne observée à une moyenne théorique
– Les moyennes observées de 2 petits échantillons

- Principe : idem à Z

• On calcule la différence Δ entre les moyennes
• On estime l'écart type sd de la différence Δ
• On calcule to= |Δ| / sd
• On compare cette valeur à la distribution théorique de la loi T de Student
• On utilise la table de la loi T

- **Conditions d'application :**

– Utilisable si petits effectifs
– Mais la distribution de la variable dans les populations doit être normale
– Et les populations doivent avoir des variances identiques
• Soit on le sait
• Soit on le teste (test F de comparaison de 2 variances)

- **Interprétation du test T**

Au risque α = 5% :
– to < Tα la différence entre les paramètres n'est pas significative
– to ≥ Tα la différence entre les paramètres est significative

- **Utilisation de la table T**

La table de T est plus difficile à utiliser que la table de Z
• Il y a autant de table de T que de degré de liberté
ddl c'est l'effectif d'un échantillon - 1
– Pour 1 échantillon : ddl = n-1
– Pour 2 échantillons : ddl = (n1 - 1) + (n2 - 1)
• En ligne les valeurs possibles de ddl
• En colonne les valeurs de α

Repérer la ligne correspondant au degré de liberté
• Repérer la valeur Tα dans cette ligne
– Si la valeur calculée to < à Tα → on ne rejette pas H0
– Si la valeur calculée to ≥ à Tα → on rejette H0 et on accepte H1

1- Comparaison d'une moyenne observée à une moyenne théorique

On compare une moyenne observée dans un échantillon de petite taille

à une moyenne connue dans une population de référence

• Variable quantitative
• Paramètre étudié moyenne
• Hypothèses:
  – H0 : M=μ
  – H1 : M≠ μ

μ : moyenne théorique connue de la population de référence
M : moyenne inconnue de la population d'où est issu l'échantillon

- **Condition d'application :**

La distribution de la variable doit être supposée normale dans la population d'où est issu l'échantillon

- **Calcul :**

$$t = \frac{|m - \mu|}{\frac{s}{\sqrt{n}}} \text{ avec ddl} = n - 1$$

- **Formulation**

– μ : moyenne théorique connue de la population de référence
– m : moyenne observée de l'échantillon
– s : écart type de l'échantillon
– n : effectif
– ddl : degré de liberté

Interprétation

to < tα → Ho non rejetée → M n'est pas significativement différente de μ
to ≥ tα → Ho est rejetée → M diffère significativement de μ

Exemple n°3:

Dans un échantillon de 18 sujets suspects d'être atteints de trypanosomiase, on mesure la quantité de protéines dans le liquide céphalorachidien. On trouve dans ce groupe une protéinorachie moyenne de 460 mg/l avec un écart type de 280 mg/l.
Dans la population générale, la protéinorachie est en moyenne de 300 mg/l.

- On se demande si ce groupe de sujet présente une protéinorachie différente de la normale ?

Solution n°3:

- **H0:** la protéinorachie des sujets atteints de trypanosomiase ne diffère pas de celle de la population générale

• **H1:** la protéinorachie des sujets atteints de trypanosomiase est différente de celle de la population
• **n < 30** : Test de T

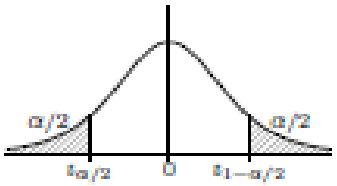
• **Condition d'application** : on suppose que la protéinorachie est distribuée normalement chez les sujets atteints de trypanosomiase
• **to** = (460 – 300) / (280 / √18) = **2,4**
• **ddl** =**17**
• **Tα** pour 17 ddl = **2,11**
• 2,4 > 2,11 : on rejette H0 → DS
La protéinorachie des sujets atteints de trypanosomiase est **significativement différente** de celle de la population.

A.3. Lois de Student

Si T est une variable aléatoire suivant la loi de Student à ν degrés de liberté, la table donne, pour α fixé, la valeur $t_{1-\alpha/2}$ telle que

$$\mathbb{P}\{|T| \geqslant t_{1-\alpha/2}\} = \alpha.$$

Ainsi, $t_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à ν degrés de liberté.



ν \ α	0,900	0,500	0,300	0,200	0,100	0,050	0,020	0,010	0,001
1	0,1584	1,0000	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6193
2	0,1421	0,8165	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,1366	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,1338	0,7407	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,1322	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,1311	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,1303	0,7111	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,1297	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,1293	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
11	0,1286	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370
12	0,1283	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
13	0,1281	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2208
14	0,1280	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	4,1405
15	0,1278	0,6912	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	4,0728
16	0,1277	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0150
17	0,1276	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,1274	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9216
19	0,1274	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,8834
20	0,1273	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,8495
21	0,1272	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,1271	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7921
23	0,1271	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,1270	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,7454
25	0,1269	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,1269	0,6840	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,7066
27	0,1268	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6896
28	0,1268	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
29	0,1268	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,6594
30	0,1267	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460
40	0,1265	0,6807	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	3,5510
60	0,1262	0,6786	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	3,4602
80	0,1261	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	3,4163
120	0,1259	0,6765	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	3,3735
∞	0,1257	0,6745	1,0364	1,2816	1,6449	1,9600	2,3263	2,5758	3,2905

Lorsque $\nu = \infty$, $t_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$.

2- Comparaison de 2 moyennes observées

On veut comparer les moyennes dans 2 échantillons de petite taille

• Paramètre étudié moyennes

• Taille de l'échantillon au moins un inférieur à 30

• **Hypothèses:**
– H0 : μ1=μ2
– H1 : μ1≠ μ2

μ_1 et μ_2 : moyennes inconnues des deux populations d'où sont tirés les échantillons

• **Conditions d'application :**

– Les distributions de la variable dans les populations d'où sont tirés les échantillons doivent être normales
– Les variances des deux populations d'où sont tirés les échantillons doivent être égales

Calcul de t :

Estimation de la variance commune aux deux échantillons:

$$S^2 = [(n_1-1).S^2_1 + (n_2-1).S^2_2]/ (n_1 + n_2)-2$$

Ecart type de la différence $\Delta = \mu1 - \mu2$ par:

$$Sd = \sqrt{(S^2/ n_1)+(S^2/ n_2)}$$

Test T de Student:

$$t_0 = |\mu1-\mu2| / Sd$$

Formulation
– m1 et m2 : moyennes observées des 2 échantillons
– s^21 et s^22 : variances des 2 échantillons
– n1 et n2 : effectifs des 2 échantillons
– ddl : degré de liberté

Interprétation

to < tα → Ho non rejetée → μ1 n'est pas significativement différent de μ2
to ≥ tα → Ho est rejetée → μ1 diffère significativement de μ2

Exemple n°4:

On a mesuré un marqueur biologique chez 2 séries de sujets, l'une composée de sujets sains, l'autre de sujets atteints d'hépatite alcoolique. L'étude a trouvé les résultats suivants:

	Effectif (n)	Moyenne du marqueur (g/l)	Ecart type
Sujets sains	15	1,6	0,19
Sujets alcooliques	12	1,4	0,21

- On veut comparer les 2 populations.

Solution n°4:

- **H0:** la valeur moyenne du marqueur est identique dans les 2 populations
- **H1:** la valeur moyenne du marqueur est différente chez les sujets atteints d'hépatite alcoolique
- **n < 30** : test de T
- **Condition d'application** : on suppose que :
  – le marqueur se distribue normalement dans les 2 populations
  – Les variances des 2 populations sont égales
- $S^2 = [(15-1) \times (0,19)^2 + (12-1) \times (0,21)^2] / (15 + 12 - 2) =$ **0,04**
- $Sd = \sqrt{(0,04 / 15) + (0,04 / 12)} =$ **0,077**
- $to = (1,6 – 1,4) / 0,077 =$ **2,60**

- **ddl** = 15+12-2 = 25

- Tα pour 25 ddl = **2,06**
- 2,6 > 2,06 : on rejette H0 → DS

Les malades atteints d'hépatite alcoolique présentent une valeur du marqueur **significativement différente** de celle des sujets sains.

3- Comparaison de deux variances

Prenons l'exemple suivant:

- On dose l'hémoglobine sanguine de 20 garçons et de 12 filles pris au hasard dans une population de jeunes âgés de 13 à 15 ans.
- Pour **les garçons**,on trouve

14,00 14,71 14,02 14 20 14 00 14,30 14,70 15,10 15,00 15,60 16,20 16,40
15,40 15,52 16,50 16,10 16,70 17,17 16,41 15,75

- Pour **les filles**,on trouve

12,12 12,10 11,90 13,20 13,10 13,50 13,40 14,80 13,50 13,88 14,00 14,60

- Peut-on accepter l'hypothèse selon laquelle **l'hémoglobinémie moyenne** est la même chez les garçons et chez les filles?

- On supposera que les teneurs en Hb, X et Y,(respectivement chez les garçons et chez les filles) sont des variables normales.

Il s'agit ici d'échantillon de petite taille. $\mathbf{n_1 = 20}$ et $\mathbf{n_2 = 12}$
Il faut vérifier, avant d'appliquer le test de **Student** , **l'égalité des variances.**

Comment comparer les variances?

- On teste l'hypothèse **H0 :** $\mathbf{\sigma^2_1 = \sigma^2_2}$

$\mathbf{\sigma^2_1}$ et $\mathbf{\sigma^2_2}$ étant les variances (vraies) respectives des variables.

- Pour comparer 2 variances calculées sur des échantillons indépendants, on utilise **le test F de Snedecor**.
- On forme le rapport **F0 =** $\mathbf{S^2_1 / S^2_2}$ en mettant la plus grande variance au numérateur.
- Ce rapport F0 est comparé à la valeur **Fα** lue sur **la table de Fisher** à ddl1 =**k1** =(n_1 - 1) et ddl2 = **k2** = (n_2 – 1) au **point 2,5%**

- Si **F0 ≥ Fα** → H0 est rejetée , les variances sont significativement différentes.
- Si **F0 < Fα** → H0 n'est pas rejetée, on peut supposer l'égalité des variances.

Dans notre exemple

- $\mathbf{m_1}$ =15,40 et $\mathbf{m_2}$ =13,34

$\mathbf{S^2_1}$ = 0,968 et $\mathbf{S^2_2}$ = 0,870

- Donc **F0** = 0,968/0,870 = 1,11
- La lecture sur la table de Fisher à **(19,11) ddl** donne au **point 2,5%**

Fα = 3,33 pour **(15,11) ddl**
Fα = 3,23 pour **(20,11) ddl**
(19,11) étant **intermédiaire** entre (15,11) et (20,11) et **1,11 < 3,23**

- Nous pouvons affirmer au seuil α=5% que l'hypothèse $\sigma^2_1 = \sigma^2_2$ **ne peut être rejetée**.

La comparaison des moyennes est alors possible en utilisant le test de **Student**

Les références:

- *Schwartz D. Méthodes statistiques.1992*
- *Ancelle T. Statistique Épidémiologie. Édition 2002*
- *Bayat S. Introduction à la Biostatistique. Université de Rennes.2009-2010*
- Bezzaoucha A. Tests statistiques en science médicales.édition 2004
- Abrouk S. Biostatistique. INSP octobre 2005