tests statistiques sont des méthodes de la statistique inférentielle qui, comme l'estimation, permettent d'analyser des données obtenues par tirages au hasard. Ils consistent à généraliser les propriétés constatées sur des observations à la population d'où ces dernières sont extraites, et à répondre à des questions concernant par exemple la nature d'une loi de probabilité, la valeur d'un paramètre ou l'indépendance de deux variables aléatoires.

La plupart des techniques permettant de comparer deux moyennes ne peuvent être utilisées que si un certain nombre d'hypothèses sont vérifiées. Dans un premier temps, nous donnons des notations et nous précisons ces hypothèses.

Supposons que nous disposons de deux échantillons de taille respective n et p que nous noterons $X_1, X_2, ..., X_n$ et $Y_1, Y_2, ...Y_p$.

On est souvent amené à prendre des décisions sur la base d'un ensemble d'observations. On fera un choix entre deux propositions contradictoires : les hypothèses H_0 et H_1 (nulle et alternative, respectivement), qui dépendent des paramètres théoriques de la population.
Déroulement d'un test comme suit :

1. Choix de **H_0 et H_1**
2. Choix de la variable de décision **(statistique du test)**
(Choisie de façon à apporter le max d'info sur le problème posé et sa loi de probabilité sera différente selon que l'hypothèse nulle ou alternative est vraie.)
3. Choix de **α petit**(typiquement 1% ou 5%).
4. Calcul de la **région critique** en fonction de α et de la statistique du test.
5. Calcul de la valeur observée de la variable de décision.
6. Comparaison et conclusion : Rejet de l'hypothèse si valeur calculée ∈ région critique.

III.1. Tests statistiques

L'utilisation des tests paramétriques t et z permet de comparer les moyennes de deux échantillons. La méthode de calcul est différente en fonction de la nature des échantillons. Les tests t et z sont dits paramétriques car ils supposent que les échantillons sont distribués suivant des lois normales. Cette hypothèse pourra être testée à l'aide des tests de normalité.

Un test est dit paramétrique si son objet est de tester une hypothèse relative à un ou plusieurs paramètres d'une variable aléatoire qui suit la loi normale ou ayant un effectif important ($n > 30$).

III.1.1. Le test de Student

Ce test permet de comparer :

- **Une moyenne d'un échantillon à une valeur donnée.**
- **Les moyennes de deux échantillons indépendants.**
- **Les moyennes de deux échantillons appariés.**

L'emploi de ce test reste subordonné en général à deux conditions d'application importantes :

La normalité et le caractère aléatoire simple des échantillons

- La première condition n'est toutefois pas essentielle lorsque les échantillons ont des effectifs suffisants (en pratique, la valeur de 30 est souvent retenue) pour assurer la quasi-normalité des distributions d'échantillonnage des moyennes. En plus, de ces deux conditions, nous devrons supposer, dans certains tests relatifs aux moyennes, l'égalité des variances des échantillons considérées.

III.1.1.1.Comparaison d'une moyenne observée à une moyenne théorique

Le test de Student cas d'un seul échantillon est aussi appelé **test de conformité**.Ce test a pour but de vérifier si notre échantillon provient bien d'une population avec la moyenne spécifiée, μ_0, ou s'il y a une différence significative entre la moyenne de l'échantillon et la moyenne présumée de la population.

Les hypothèses à tester sont :

- **hypothèse nulle : $H_0 : \mu = \mu_0$**
- **hypothèse alternative :**

$H_1 : \mu > \mu_0$ (test unilatéral à droite).

$H_1 : \mu < \mu_0$ (test unilatéral à gauche).

- **Conditions d'application**

–Effectifs $n < 30$.

– La distribution de la variable dans les populations doit être normale ;

– Les populations doivent avoir des variances identiques.

- Soit on le sait.
- Soit on le teste (test F de comparaison de 2 variances)

➢ **Calcul (Eq 9)**

(Eq 9)

$$t = \frac{/m - \mu/}{\frac{s}{\sqrt{n}}}$$

Avec ddl= n-1

Où :

– 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**

On compare la valeur calculée de t (t_{obs}) avec la valeur critique appropriée de t avec n - 1 degrés de liberté. On rejette H_0 si la valeur absolue de (t_{obs}) est supérieure à cette valeur critique.

$t_{obs} < t_{\alpha}$ → H_o non rejetée → m n'est pas significativement différente de μ.

$t_{obs} \geq t_{\alpha}$ → H_o est rejetée → m diffère significativement de μ.

➢ **Utilisation de la table T**

- 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 = (n_1 - 1) + (n_2 - 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 T_α dans cette ligne :

– Si la valeur calculée $t_o < à T_\alpha \rightarrow$ on ne rejette pas H_0.

– Si la valeur calculée $t_o \geq à T_\alpha \rightarrow$ on rejette H_0 et on accepte H_1.

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 Pt $|T| \geq t_{1-\alpha/2} = \alpha$.

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

	γ										
k	0.25	0.20	0.15	0.10	0.05	0.025	0.010	0.005	0.0025	0.0010	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Lorsque ν =α, $t_{1-\alpha/2}$ est le quantile d'ordre 1 – α/2 de la loi normale N p 0, 1

III.1.1.2. Comparaison de deux moyennes observées

• Paramètre étudié : moyennes

• Taille de l'échantillon : ≤ 30

➢ **Hypothèses**

– H_0 : $\mu_1=\mu_2$

– H_1 : $\mu_1 \neq \mu_2$

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

➢ **Calcul**

Estimation de la variance commune aux deux échantillons (Eq 2) :

$$S^2 = [(n_1-1).S^2_1 + (n_2-1).S^2_2]/ (n_1 + n_2)-2 \qquad \textbf{(Eq2)}$$

Ecart type de la différence $\Delta = \mu_1 - \mu_2$ par :

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

Test T de Student (Eq3) :

$$t_0 = \left| \mu 1-\mu 2 \right| / Sd \qquad \textbf{(Eq3)}$$

➢ **Formulation**

– m_1 et m_2 : moyennes observées des 2 échantillons

– $s^2 1$ et $s^2 2$: variances des 2 échantillons

– n1 et n2 : effectifs des 2 échantillons

– ddl : degré de liberté

➢ **Interprétation**

$t_{obs} < t_\alpha \rightarrow H_o$ non rejetée → μ_1 n'est pas significativement différent de μ_2.

$t_{obs} \geq t_\alpha \rightarrow H_o$ est rejetée → μ_1 diffère significativement de μ_2.

Exemple :

La mesure de la glycémie sur deux échantillons indépendants a donné les résultats suivants :

$n_1 = 25$, $m_1 = 1{,}75$g/l, $S_1 = 0{,}07$g/l

$n_2 = 41$, $m_2 = 1{,}05$g/l, $S_2 = 0{,}05$g/l

1- Comparer les variances des deux échantillons ?
2- Comparer la moyenne des deux échantillons ?

I. Comparer les deux variances

1. **H0 :** il n'existe pas de différence significative entre les deux variances S1=S2

H1 : $S1 \neq S2$

2. Une comparaison de deux variances on applique le test **F de Snedecor**
3. On suppose que la glycémie suit la loi normale
4. Calcul de F : **$F=S_1^2/S_2^2$ Soit $S_1>S_2$**

$F= 0.07^2/0.05^2 =1.96$ à α = 5% et $ddl_1 = 25 - 1 = 25$

$ddl_2 = 41- 1 = 40$

$\boldsymbol{F^{24}_{40}}$de la table F à 2.5%= 2.01

5. F calculé < F tabulaire $\longrightarrow$ la différencen'est pas significative (**$S_1=S_2$)**

II. Comparer la moyenne deux échantillons

1. H0 : les deux moyennes de glycémie des deux échantillons sont égales (pas de différence significative) m1= m2

H1 : $m1\neq m2$

2. Il s'agit de comparer les deux moyennes on applique soit le test de student ou le test Z et puisque **n< 30** donc c'est le test T de Student
3. **Calculer le test t**

$$t = |m1 - m2| \, /Sd$$

On doit d abords calculé

- **La var**iance commune aux deux échantillons

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

$S^2 = \{24*(0.07)^2 + 40*(0.05)^{2\}}/ 25-41-2= 0.003$

- **Calculer l'écart type da la différence des deux moyennes**

Sd = √ (S²/ n_1) + (S²/ n_2)= $\sqrt{\frac{0.003}{25}} + \sqrt{\frac{0.003}{41}} = \mathbf{0.014}$

- **Calcul de t Student**

t = $|m1 - m2|\ /Sd$**= 1.75 – 1.05/** $\mathbf{0.014}$**= 50.4**

4. **Recherche de t tabulaire**

A α= 5% et ddl = 64 ⟹ le t tabulaire prend deux valeurs (t_{60}= 2.0003 et t_{80}= 1.9901)

t calculé > t tabulaire alors :

Les deux moyennes sont significativement différentes(P< 0.001)

III.1.2. Test Z (Ɛ) ou de l'écart réduit

Il est utilisé pour comparer :

– Une moyenne observée à une moyenne théorique.

– Deux moyennes observées.

III.1.2.1. Comparaison d'une moyenne observée à une moyenne théorique

- On suppose que l'on ait un échantillon qui suit une loi normale N (μ, σ_2) ou la variance est connue. On compare une moyenne observée dans un échantillon à une moyenne connue dans la population de référence.
- **Hypothèses**

On veut tester H_0 : $\mu = \mu_0$ contre H_1 : $\mu \neq \mu_0$, c'est le cas bilatéral.

- **Conditions d'applications :**

Taille de l'échantillon ≥ 30

- **Calcul de test Z (Eq 4)**

$$\mathbf{Z = \frac{/m - \mu/}{\frac{s}{\sqrt{n}}}} \quad \textbf{(Eq4)}$$

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 → H_0 non rejetée → m n'est pas significativement différente de μ.

Zo ≥ 1,96 → H_0 est rejetée → m diffère significativement de μ.

- **Utilisation de la table :**

Voici la table Z

Voici la table Z
3° Quantiles de la loi Normale (bis). – Si Z est une variable aléatoire suivant la loi normale N(0,1), la table donne, pour α fixé, la valeur telle que
$P(|Z| \geq z_{1-\alpha/2}) = \alpha$
Ainsi $z_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale 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

Exemple

pour α= 0.5, on trouve z≈0.6745, pour α=0.25 ; on trouve z≈ 1.1503 ; pour $\alpha=10^{-6}$ On trouve z≈ 4 .8916

Exemple :

La tension artérielle(TA) dans une population suit la loi normale, μ est égale à120,0 mm Hg et d'écart type δ est égale à 10,0 mm Hg

Dans un échantillon représentatif de 100 individus, on a trouvé une moyenne(m) de tension artérielle de 110,6 mm Hg et un écart type (S) de 11.3 mm Hg

Comparer les deux moyennes et commenter ?

C'est une comparaison d'une moyenne à une moyenne théorique tel que :

μ = 120.0 mm Hg ; δ= 10.0 mm Hg
M=110.6 mm Hg; S = 11.3 mm Hg

- **H_0 :**il n'existe pas une différence significative entre la moyenne de TA de l'echantillon et celle de la population

- **H_1:**$m \neq \mu$
- n> 30 et la distribution supposant qu'elle suit la loi normale
- application du test Z $\mathbf{Z} = \frac{/m-\mu/}{\frac{s}{\sqrt{n}}}$
- **Z_0**$=120 - 110.6)/(11.3/\sqrt{100}) = 8.31$

• 8.31>1,96 →**On rejette H0** →

La moyenne de La population et de l'échantillon sont significativement différentes.

III.1.2.2.Comparaison de deux moyennes observées

- **Conditions d'application**

Effectif de chaque échantillon ≥ 30

- **Hypothèses**

– H_0 : $\mu_1=\mu_2$

– H_1 : $\mu_1 \neq \mu_2$

$\mu_1$et μ_2 : moyennes inconnues des deux populations d'où sont tirés les échantillons.

- **Calcul** **(Eq5)**

$$Z = \frac{/m_1 - m_2/}{\sqrt{\frac{S^2}{n_1} + \frac{S_2^2}{n2}}}$$

– m_1 et m_2 : moyennes observées des 2 échantillons.

– s^2 1 et s^2 2 : variances des 2 échantillons.

– n_1 et n_2 : effectifs des 2 échantillons.

- **Interprétation**

$Zo < 1,96 \rightarrow$ Ho non rejetée $\rightarrow \mu_1$ n'est pas significativement différent de μ_2

$Zo \geq 1,96 \rightarrow$ Ho est rejetée $\rightarrow \mu_1$ diffère significativement de μ_2

Exemple :

On a mis en place un système de surveillance des diarrhées nosocomiales dans un service de pédiatrie. Le questionnaire relevait pour chaque cas plusieurs variables dont la durée d'hospitalisation.

En 2014 et 2015 respectivement 153 et 108 cas de diarrhée ont été recensés.

- Comparer la moyenne de la durée d'hospitalisation de tous les cas de diarrhée qui était de 21,5+/- 1,4 jours en 2014 et de 24,5+/- 2,1 jours en 2015.

Comparaison de deux moyennes :

- **H_0** : $m_1=m_2$
- **H_1** : $m_1 \neq m_2$
- Comparaison de deux moyennes, n> 30 et la distribution suit la loi normale

$$Z = \frac{/m_1 - m_2/}{\sqrt{\frac{S^2}{n_1} + \frac{S_2^2}{n2}}}$$

• **Sd** = $\sqrt{(17.3^2/153) + (21\ .8^2/108)}$

Zo = 1.2 à α= 5%

- 1.2< 1,96 : on accepte H0 $\rightarrow$ les deux moyennes ne sont pas significativement différentes

III.3. Comparaison de deux variances

Comment comparer les variances ?

- On teste l'hypothèse **H_0 : $\sigma^2_1 = \sigma^2_2$**

σ^2_1et **σ^2_2**étant les variances (vraies) respectives des variables.

III.1. test F de Snedecor.

- Pour comparer deux variances calculées sur des échantillons indépendants, on utilise **letest F de Snedecor**.

- On forme le rapport :

$\mathbf{F_0 = S^2_1 / S^2_2}$

En mettant la plus grande variance au numérateur.

- Ce rapport F_0 est comparé à la valeur $\mathbf{F\alpha}$ lue sur **la table de Fisher** à ddl1 =**k1**= (n_1 -1) et ddl2 = **k2**= (n_2 – 1) au **point 2,5%.**

- **Si $F_0 \geq F_\alpha \rightarrow H_0$ est rejetée, les variances sont significativement différentes.**
- **Si $F_0 < F_\alpha \rightarrow H_0$ n'est pas rejetée, on peut supposer l'égalité des variances.**