Chapitre IV : Les statistiques inférentielles (Tests de comparaison)

But : Les statistiques inférentielles, consistant en des tests permettant de confirmer ou infirmer une hypothèse.

Variables quantitatives

1) **Test t ou test Student**

But : C'est un test permet de compares deux distributions quantitatives ou deux moyennes.

Principe : Leur principe est de comparer la valeur de $t_{calculé}$ (observé) avec une valeur théorique (critique = tabulaire) selon deux critères :

- Le seuil de signification (α) ; 0,05 ou 5%
- Le degré de liberté (d.d.l. = $n_1 + n_2 - 2$). 10+12 -2 = 20 Implique $t_{théorique}$ = 2,08

Hypothèses

- **Hypothèse nulle (H_0)**: $x_1 = x_2$.
- **Hypothèse alternative (H_1)**: $x_1 \neq x_2$.

La valeur de $t_{calculé}$ est donné par la relation suivante :

$t_{calculé} = \left| m_1 - m_2 \right| / Sx \sqrt{(1/n_1+1/n_2)}$, S : écart-type. $S= \sqrt{S^2_X}$. $S^2_X = \sum (x_{i1} - m_1) + \sum (x_{i2} - m_2) / n_1 + n_2 - 2$

La comparaison donne généralement deux cas :

- $t_{calculé} \geq t_{critique}$ $\Longrightarrow$ Il existe une différence significative ;
- $t_{calculé} < t_{critique}$ $\Longrightarrow$ Il existe une différence non significative.

Exemple d'application. Le domaine vital de l'ours noir (km^2) a été mesuré pour des individus mâles et femelles.

Mâles	Femelles
94	37
504	72
173	60
274	49
560	102
168	18
	50
	49
	20

Existe-t-il une différence significative dans la taille du domaine vital chez le mâle et la femelle de l'ours noir ?

Le test $t_{calculé}$ pour deux échantillons séparés ou indépendants est donné par la relation suivante :

$t_{calculé} = |m_1 - m_2| / Sx \sqrt{(1/n_1+1/n_2)}$, Sx : écart-type. $Sx = \sqrt{S^2_X}$.

$S^2_X = \sum (x_{i1} - m_1)^2 + \sum (x_{i2} - m_2)^2 / n_1 + n_2 - 2$

Solution

m1 = 295,5 km²
m2 = 50,77 km²

$\sum (x_{i1} - m_1)^2 = (94-295,5)^2 + (504-295,5)^2 + (173-295,5)^2 + (274-295,5)^2 + (560-295,5)^2 + (168-295,5)^2$

= 173159,5

$\sum (x_{i2} - m_2)^2 = (37-50,77)^2$ **+ (72-50,77)² + (60-50,77)² + (49-50,77)² + (102-50,77)² + (18-50,77)² + (50-50,77)² + (49-50,77)² + (20-50,77)² = 5377,54**

S^2_X = 173159,5 + 5377,54 / 13 = 13733,62

Sx = 117,19

$\sqrt{(1/n_1+1/n_2)}$ = (1/6 + 1/9) = 0,16 + 0,11 = √0,27

= 0,52

$t_{calculé} = |m_1 - m_2| / Sx \sqrt{(1/n_1+1/n_2)}$

= (295,5 – 50,77) / 117,19 (0,52) = **3,83**

$t_{calculé}$ = **3,83**

Pour un seuil de 5% et une valeur de ddl = (n1-1) + (n2-1) = n1+n2 -2 = 6+9-2 = 13

La valeur théorique de t = 2,16

$t_{calculé} > t_{théorique}$ implique Il existe une différence significative entre les deux moyennes testées.

3,83 2,16

Conclusion

La surface occupée par le mâle est presque six fois plus large que celle occupée par la femelle.

Variables qualitatives

Test Khi-deux

But : C'est un test permet de chercher et de comparer la différence entre deux variables qualitatives.

Principe : Leur principe repose sur la comparaison d'une valeur de $\chi^2_{calculé}$ par rapport à une autre valeur de $\chi^2_{théorique}$ en fonction de degrés de liberté (v). Le v = (nbr de colonnes -1) × (nbr de lignes -1).

Le test Khi-deux est noté par χ^2 et donné par formule suivante:

$\chi^2_{calculé} = \Sigma$ (Effectifs observés$_i$ – effectifs calculés$_i$)^2 / effectifs calculés$_i$ ou $\boldsymbol{\chi^2_{calculé} = \Sigma (O_i - C_i)^2 / C_i}$

Exemple d'application

L'efficacité de deux traitement A et B a été testée vis-à-vis deux lots de 40 animaux, l'un soumis à A et l'autre à B. les résultats sont les suivants :

Effectifs observés	**Succès**	**Echec**	Total
Traitement A	17	23	40
Traitement B	25	15	40
Total	42	38	80

Testez l'efficacité de deux traitements au seuil α = 5%.

Solution

Méthode de calcul des effectifs calculés

Effectif calculé de (17) = somme de ligne (40) × somme de la colonne (42) / la somme totale (80) = **21**

1/ Calculer les effectifs théoriques (appelés également attendus ou calculés)

Effectifs calculés ou théoriques	**Succès**	**Echec**	Total
Traitement A	21	19	40
Traitement B	21	19	40
Total	42	38	80

On a $\chi^2_{calculé} = \Sigma (O_i - C_i)^2 / C_i = (17-21)^2/21 + \ldots\ldots.+(15-19)^2/19 = \mathbf{3{,}2}$.

La valeur de $\chi^2_{calculé}$ doit être comparée à la valeur critique de Khi-deux (sur la table) au seuil α = 0,05 : si $\chi^2_{calculé}$ est supérieur ou égale au $\chi^2_{théorique}$, on considère la différence significative.

Pour α = 0,05 et ddl = (2-1) × (2-1) = 1 on a $\chi^2_{théorique}$ = **3,84**.

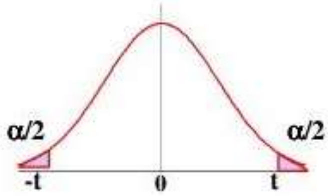
On a $X^2_{cal.}$ (3.2) < $X^2_{thé.}$ (3,84) implique il existe une différence non significative ou pas de relation entre les deux variables type de traitement et guérison.

Conclusion
Les deux traitements ont la même efficacité au seuil α = 5%.

Tables statistiques

Table de t (lois de Student)*

La table donne la probabilité α pour que *t* égale ou dépasse, en valeur absolue, une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



ddl / α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	1,000	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,765	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,190	1,553	2,132	2,776	3,747	4,604	8,610
5	0,132	0,727	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,718	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,706	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,703	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,700	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,697	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,695	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,694	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,690	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,740	2,110	2,567	2,898	3,956
18	0,127	0,688	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,688	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,687	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,686	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,686	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,685	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,685	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,684	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,684	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,684	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,683	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,683	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,683	1,055	1,310	1,697	2,042	2,457	2,750	3,646
+ ∞	0,126	0,674	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Exemple : avec d. d. l. = 10, pour *t* = 2,228, la probabilité est α = 0,05

*(d'après Fisher et Yates, Statistical tables for biological, agricultural, and medical research (Oliver and Boyd, Edinburgh) avec l'aimable autorisation des auteurs et des éditeurs)

Table Khi-deux

Loi du χ^2

$1-\alpha$	**0,001**	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**	**0,999**
α	**0,999**	**0,995**	**0,99**	**0,975**	**0,95**	**0,9**	**0,5**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**
ν = ddl													
1	0,00	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88	10,83
2	0,00	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60	13,82
3	0,02	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84	16,27
4	0,09	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86	18,47
5	0,21	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75	20,51
6	0,38	0,68	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55	22,46
7	0,60	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28	24,32
8	0,86	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95	26,12
9	1,15	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59	27,88
10	1,48	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19	29,59