Autres modèles statistiques : 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 (n < 30)**

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

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

Le test $t_{calculé}$ est donné par la relation suivante :

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

1.1) Test t pour échantillon unique (test de conformité)

But : Les tests de conformité sont destinés à vérifier si un échantillon peut être considéré comme extrait d'une population donnée ou représentatif de cette population, vis-à-vis d'un paramètre comme la moyenne.

Exemple d'application : Le taux de cholestérol dans la population est connu et vaut 4,3 mg/l. Les résultats de 10 dosages sont présentés dans le tableau suivant.

Sujet	Taux de cholestérol
1	4,5
2	4,8
3	4,3
4	5
5	5,2

6	3,9
7	3,8
8	5,1
9	5,2
10	4,5

La moyenne de l'échantillon conforme la valeur de la population ?

Le test $t_{calculé}$ pour un échantillon (test de conformité) est donné par la relation suivante :

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Solution

L'erreur type moyen ou Standard Erreur Moyen (SEM) = Sx / Racine de n

Moyenne m = 4,63

Ecart type ET = 0,52

SEM ou Erreur type = 0,52/3,16 = 0,16

$t_{calculé}$ = 4,63 – 4,3/0,16 = 2,06

En fonction de deux critères suivants :

1) Seuil de signification = 0,05 ou 5%
2) d.d.l. = n-1 = 10-1 = 9

On a $t_{théorique}$ = 2,26

$t_{calculé}$ (2,06) < $t_{théorique}$ (2,26)

Il existe une différence non significative entre la moyenne de l'échantillon et la moyenne de la population ; implique l'acceptation de H0 et le rejet de H1.

Conclusion scientifique

Oui l'échantillon conforme la population ; l'échantillon possède une valeur normale de cholestérol.

1.2) Test t pour deux échantillons indépendants

Il s'agit de deux séries de mesure pour lesquelles il n'y a aucune correspondance entre les éléments de la première série et ceux de la deuxième ; les deux séries de mesures sont obtenues avec des sujets différents. Dans ce cas le but de l'application du test t est de voir si les deux moyennes calculées sur les deux échantillons diffèrent significativement.

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 :

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

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

= 173159,5

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

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

Sx = 117,19

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

= 0,52

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

= (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.

1.3) Test t pour deux échantillons dépendants ou appariés ou liés

Il s'agit de deux séries de mesures pour lesquelles il y a une correspondance stricte, terme à terme, entre les éléments de l'une et les éléments de l'autre.

Exemple d'application : La quantité de bactéries par cm^3 de lait provenant de 8 vaches différentes est estimée juste après la traite et 24 h plus tard.

Vache	Juste après la traite En 1000	24h après la traite En 1000	différences
1	12	14	2
2	13	20	7
3	21	31	10
4	17	28	11
5	15	26	11
6	22	30	8
7	11	16	5
8	21	29	8
Somme			62 /8
Moyenne	16,5	24,25	7,75

Existe-t-il un accroissement significatif du nombre de bactéries par cm^3 de lait au cours du temps ?

Solution

$T_{calculé}$ = D/SEM

D : LA Moyenne des Différence = ∑différences/n

Erreur type moyen (SEM) = Sx/racine de n

$S^2x = \sum(xi1-m1)^2 + \sum(xi2-m2)^2 / n-1$

$\sum(xi1-m1)^2 = (12-16,5)^2 + (13-16,5)^2 + (21-16,5)^2 + (17-16,5)^2 + (15-16,5)^2 + (22-16,5)^2 + (11-16,5)^2 + (21-16,5)^2$

$\sum(xi2-m2)^2 = (14-24,25)^2 + (20-24,25)^2 + (31-24,25)^2 + (28-24,25)^2 + (26-24,25)^2 + (30-24,25)^2 + (16-24,25)^2 + (29-24,25)^2$

$S^2x = 9,64$ implique $Sx = 3,10$
Erreur type (SEM) = $3,10 / \sqrt{8} = 3,10 / 2,82$ implique SEM = 1,09

$t_{calculé}$ = 7,75 / 1,09 = 7,05

POUR un seuil de 5% et une valeur de ddl = n-1 = 8-1 = 7

La valeur théorique de t = 2,36

On a **$t_{calculé}$** > **$t_{théorique}$** ce qui implique l'existence d'une différence significative entre les deux moyennes testées.

Conclusion
Oui il existe un accroissement significatif du nombre de bactéries par cm^3 de lait au cours du temps.

Analyse de la variance à un critère (facteur) (ANOVA)

But : C'est un test permet de chercher et de comparer la différence entre plusieurs échantillons (moyennes) quantitatifs.

Principe : Leur principe repose sur la comparaison d'un facteur calculé ($F_{calculé}$) (observé) par rapport à un autre facteur théorique (critique) en fonction de degrés de liberté (v_1 et v_2) et au seuil de signification (α).

Le facteur calculé est donné par la formule suivante :

$F_{calculé}$ = CMF/CMR

CMF = SCF/v_1

$SCF = T^2{}_1/n_1 + T^2{}_2/n_2 + T^2{}_3/n_3 + \ldots\ldots T^2{}_n/n - T^2/N.$

$T = T_1+T_2+T_3+\ldots\ldots T_n$

N = n1 + n2 + n3 + n.

v_1 = nombre de niveaux – 1 ou nombre d'échantillons – 1 = 3-1 = 2

SCT = SCF + SCR

$SCT = \sum x^2 + \sum y^2 + \sum z^2 + \ldots + \sum k^2 - T^2 /N$

CMR = SCR/v_2

v_2 = nombre de toutes les valeurs – nombre de niveaux.

La valeur de facteur théorique est trouvée dans le tableau théorique en fonction de deux critères ;

- Le seuil de signification (α) ;
- Les degrés de liberté v_1 et v_2.

La comparaison donne généralement deux cas :

- $F_{calculé} \geq F_{critique}$ $\Longrightarrow$ Il existe une différence significative entre les moyennes ;
- $F_{calculé} < F_{critique}$ $\Longrightarrow$ Il existe une différence non significative entre les moyennes.

Exemple

Un essai comparatif de la teneur en azote dans trois variétés de blé, les résultats sont représentés dans le tableau suivant :

V1	V2	V3
$12^2 = 144$	$16^2 = 256$	$20^2 = 400$
$14^2 = 196$	$18^2 = 324$	$21^2 = 441$
$13^2 = 169$	$17^2 = 289$	$21^2 = 441$
$T^2_1 = (12+14+13)^2 /3 = 507$	867	1281,33
T = 152 N= n1+n2+n3 = 9		
T^2 /N	2567,11	
SCT = **2660 – 2567,11 = 92,89**		
M1 = 13	M2 = 17	M3 = 20,66

Est-ce que les différences entre les teneurs en azote au sein des variétés sont significatives ?

Solution

Groupes = échantillons = modalités = niveaux

$F_{calculé}$ = CMF/CMR

CMF = SCF/v_1

V1 = 3-1 = 2

SCF = $T^2_1/n_1 + T^2_2/n_2 + T^2_3/n_3 + \ldots\ldots T^2_n/n$ - $\mathbf{T^2/N}$.

SCF = (507 + 867 + 1281,33) - 2567,11

SCF = 88,22 Implique CMF = 88,22 / 2 = **44,11**

$T = T_1+T_2+T_3+.....T_n$

N = n1 + n2 + n3 + n.

v_1 = nombre de niveaux – 1 ou nombre d'échantillons – 1 = 3-1 = 2

SCT = SCF + SCR

$SCT = \sum x^2_i + \sum y^2 + \sum z^2 + \ldots + \sum k^2 - \mathbf{T^2 / N}$

SCR = SCT – SCF = 92,89 – 88,22 = 4,67

CMR = SCR/v_2

v_2 = nombre de toutes les valeurs – nombre de niveaux. N-3 = 9-3 = 6

CMR = SCR / V2 = 4,67/6 = 0,77

SCT = SCF + SCR implique SCR = SCT-SCF = 92,89 – 88,22 = 4,67

Pour $\alpha = 0{,}05$ et $v_1 = 2$ et $v_2 = 6$ on a $F_{critique}$ = **5,14**

C'est-à-dire $F_{calculé}$ (56,71) > $F_{critique}$ (5,14) implique l'existence d'une différence significative entre les variétés.

	Somme des Carrés SC	d.d.l.	Carré Moyen CM	$F_{calculé}$
Inter-groupes (F)	88,222	v1 = 2	44,111	**56,714**
Intragroupes (R)	**4,667**	v2 6	0,778	
Total (T)	92,889			

Conclusion

Les teneurs en azote sont significativement différentes au sein de trois variétés testées ; la variété 3 possède la teneur la plus élevée.

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:

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

Exemple

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 $\chi^2_{\text{calculé}} = \Sigma (O_i - C_i)^2 / C_i$ = $(17-21)^2/21$ + ……..+$(15-19)^2/19$ = **3,2**.

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

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

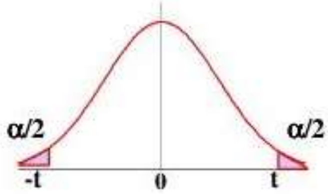
On a $X^2_{\text{cal.}}$ (3.2) < $X^2_{\text{thé.}}$ (3,84) implique il existe une différence non significative ou pas de relation entre les deux variables type de traitement et guérison.

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 de Fisher (Test Anova à un facteur) α = **0,05**

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	>25
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.08	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
>120	3.84	2.99	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.00

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