Cours 3* : *Les tests paramétriques

Introduction

Dans le chapitre précédent, nous avons vu qu'un test paramétrique nécessite que les données soient issues d'une distribution paramétrée; aussi les hypothèses nulle et alternatives du test portent sur un paramètre statistique (moyenne, variance ou proportion). Ces tests nécessitent des conditions de validité, notamment en ce qui concerne la taille de l'échantillon, la normalité de la distribution et l'égalité des variances.

Dans ce présent chapitre, nous allons traiter un autre type des tests qui sont moins exigeants par rapport à ces conditions, ces tests sont dits non paramétriques et ne font aucune hypothèse sur la distribution sous-jacente des données.

Lorsque les données sont quantitatives, les test non paramétriques transforment les données en rangs et mesurent l'accord entre les rangs observés et ce que devrait être ces rangs sous une hypothèse nulle.

Lorsque les données sont qualitatives, seuls les tests non paramétriques sont utilisables.

Il existe une grande diversité de tests non paramétriques.

1. Variables qantitatives

Un échantillon	deux échantillons	plus de deux échantillons
-Tester les valeurs douteuses dans un échantillon: **Test de Dixon** -Tester si la valeur médiane de l'échantillon s'écarte d'une valeur théorique:**Test de Wilcoxon** -Tester si la distribution dans l'échantillons suit une loi donnée:**Test d'ajustement de Kolmogorov-Smirnov** cas particulier: si la distribution de l'échant suit une loi normale: **Test de normalité de Kolmogorov-Smirnov**	-Comparer la distribution de 2 échants non appariés: **1.Test de Kolmogorov Smirnov** **2.Test de Wilcoxon Mann-Whitney** -Comparer la distribution de échants appariés:**Test de Test de Wilcoxon apparié**	-Comparer la distribution de plusieurs échants Non appariés(indépts): **Test de Kruskal-Wallis** -Comparer la distribution de plusieurs échants appariés **Test de Friedman**

2. Variables qualitatives

Un échantillon	deux échantillons	plus de deux échantillons
-Tester si la distribution de l'échant suit une loi binomiale:**Test binomial**	-Comparer la distribution de 2 échantillons non appariés **Test de Khi2** -Comparer la distribution de 2 échantillons appariés: **Test de McNemar**	-Comparer la distribution de plusieurs échants non appariés:**Test de Khi2** -Comparer la distribution de plusieurs échants appariés: **Test de Cochran**

Contraintes des tests paramétriques

- Normalité des distributions dans les populations.
- Homogénéité des variances.
- Peu applicables aux effectifs réduits-approximativement $n < 30$.

Caractéristiques des tests non paramétriques

- Pas de contraintes sur la distribution du caractère dans la population.
- Seuls tests permettant de comparer des échantillons issus de population ayant des distributions différentes.
- Seuls tests traitant des données qualitatives exprimées soit en variable nominale ou ordinal.
- Dans la plupart des cas, tests basés sur les rangs.

1. Test de Khi-Deux de Pearson

Le test de Khi-Deux est le plus célèbre des tests non paramétriques, il est à l'origine un test d'ajustement (ou adéquation) d'une loi totalement inconnue à une loi donnée; mais il peut être aussi utilisé pour vérifier l'indépendance ou l'homogénéité de deux variables aléatoires. C'est un test global qui porte sur l'ensemble des effectifs observés après l'expérience et calculés à partir de l'hypothèse testée.

Test d'indépendance de Khi-Deux

Le test d'indépendance de Khi-Deux aussi appelé **test de Khi-Deux de Pearson** consiste à mesurer l'écart entre la distribution des effectifs (ou fréquences) observés d'un échantillon et celle des effectifs théoriques, et tester ensuite si cet écart est suffisamment faible pour être imputables aux fluctuations de l'échantillonnage.

Tableau de contingence

C'est un tableau qui représente la répartition des effectifs d'un échantillon en fonction de la valeur d'une observation, il prend la forme suivante :

	X_1	X_2	Total
Échant 1	a	b	N_1
Échant 2	c	d	N_2
Total	C_1	C_2	N

Calcul des effectifs théorique

Les effectifs théoriques notés $a^{'}, b^{'}, c^{'}$ et $d^{'}$ sont calculés à partir des effectifs observés par les formules suivantes:

$$a^{'} = \frac{C_1 * N_1}{N};\ b^{'} = \frac{C_2 * N_1}{N};\ c^{'} = \frac{C_1 * N_2}{N};\ d^{'} = \frac{C_2 * N_2}{N}; \tag{1}$$

D'où nous obtenons le tableau des effectifs théoriques suivant :

	X_1	X_2	Total
Échant 1	$a^{'}$	$b^{'}$	N_1
Échant 2	$c^{'}$	$d^{'}$	N_2
Total	C_1	C_2	N

Statistique du test

La statistique du test de Khi-Deux calculée est donnée par

$$\mathcal{X}_c^2 = \sum \frac{(observés - théoriques)^2}{théoriques}$$

La règle de décision

L'intervalle d'acceptation de H_0 est $[0, \mathcal{X}^2_{\alpha,\nu}[$, où $\mathcal{X}^2_{\alpha,\nu}$ est la valeur critique (tabulé) de la loi de Khi-2 avec

$$\nu = (nombredelignes - 1) * (nombredecolonnes - 1)$$

.

On accepte H_0 si $\mathcal{X}_c^2 \in [0, \mathcal{X}^2_{\alpha,\nu}[$ i.e si $\mathcal{X}_c^2 < \mathcal{X}^2_{\alpha,\nu}$.

Exemple

On a étudié sur deux échantillons provenant de deux populations différentes la répartition des quatre groupes sanguins : O, A, B et AB. Les résultats obtenus sont le tableau de contingence suivant :

	O	A	B	AB	Total
échant 1	121	120	79	33	353
échant 2	118	95	121	30	364
Total	=239	=215	200	63	717

On veut tester l'hypothèse :

H_0 : Les quatre groupes sanguins sont répartis de la même manière sur les deux populations.

Contre l'hypothèse

H_1 : La répartition des groupes sanguins est différente dans les deux population. **Tableau de contingence des effectifs théoriques**

	O	A	B	AB	Total
échant 1	117.66	105.9	98.47	31.01	353
échant 2	121.33	109.15	101.53	31.98	364
Total	239	215	200	63	717

Ces effectifs théoriques sont calculés en utilisant les formules (1) précédents.

Statistique du test

$$\mathcal{X}_c^2 = \sum \frac{(observés - théoriques)^2}{théoriques}$$

$$\mathcal{X}_c^2 = \mathcal{X}_{c1}^2 + \mathcal{X}_{c2}^2$$

Les calculs sont dans le tableau suivant :

$\sum \frac{(observés - théoriques)^2}{théoriques}$					
	O	A	B	AB	Total
échant 1	$\frac{(3.34)^2}{117.66} = 0.095$	$\frac{(14.1)^2}{105.9} = 1.88$	$\frac{(-19.47)^2}{98.47} = 3.85$	$\frac{(1.99)^2}{31.01} = 0.13$	$\mathcal{X}_{c1}^2 = 5.075$
échant 2	$\frac{(-3.33)^2}{121.33} = 0.091$	$\frac{(-14.15)^2}{109.15} = 1.83$	$\frac{(19.47)^2}{101.53} = 3.73$	$\frac{(-1.98)^2}{31.98} = 0.12$	$\mathcal{X}_{c2}^2 = 5.771$

D'où la statistique du tests est $\mathcal{X}_c^2 = \mathcal{X}_{c1}^2 + \mathcal{X}_{c2}^2 = 5.075 + 5.771 = 10.846$.

D'autre part $\mathcal{X}_T^2 = \mathcal{X}_{\alpha,\nu}^2 = \mathcal{X}_{5\%,(2-1)(4-1)}^2 = \mathcal{X}_{5\%,3}^2 = 7.81$.

On a $\mathcal{X}_c^2 > \mathcal{X}_T^2$ donc on rejette l'hypothèse nulle H_0.

Conclusion

La répartition des quatre groupes sanguins n'est pas la même dans les deux populations.

2. Test des rangs signés de Wilcoxon

C'est un test pour comparer deux échantillons appariés (dépendants) par rapport à une variable quantitative ou qualitative ordinale.

L'appariement

Deux échantillons qui sont appariés c-à-d : qu'ils ne sont pas indépendants l'un de l'autre. Il peut s'agir par exemple d'une variable qui a été mesurée.

- ▷ Deux fois par le même opérateur dans des conditions de traitement différentes.
- ▷ Par deux opérateurs différents dans les même conditions de traitement.
- ▷ À deux instants t_1 et t_2 pour étudier l'effet d'un traitement au cours du temps.

Il peut s'agit aussi des observations obtenus chez les même individus.

- ▷ Étude avant/après (douleurs avant et après la prise d'un antalgique).

◦ *Exemple 1* : cas témoins (observations obtenues chez des individus différents présentant des caractères similaires).

◦ *Exemple 2* : cas témoins (appariement sur les facteurs de risque).

L'intérêt de l'appariement est qu'on peut contrôler la variabilité inter-individuelle et donc augmenter la puissance statistique dans la comparaison.

Principe du test de Wilcoxon

Nous disposons de 2 échantillons dépendants de même taille $n = n_1 = n_2$, et nous définissons le couple (X_1, X_2) pour chaque sujet.

Nous formulons une nouvelle variable : $D = X_1 - X_2$, variable aléatoire de moyenne μ.

Hypothèses à tester

Nous voulons tester les hypothèses suivants :

H_0 : "Les deux variables X_1 et X_2 ont la même distribution"

c-à-d fonctions de répartition identiques : $F_1(x) = F_2(x)$

H_1 : "X_1 et X_2 n'ont pas la même distribution de probabilité"

$F_1(x) \neq F_2(x)$

Nous procédons le test des rangs signé de Wilcoxon de la manière suivante :

1. On calcule les écarts entre les valeurs de chaque couple de mesures : $d_i = x_{i_1} - x_{i_2}$.
2. On élimine les écarts nuls.
3. On prend les valeurs absolues écarts $|d_i|$, mais on retient le signe de l'écart.
4. On classe les $|d_i|$ de façon croissante.
5. On calcule R^+ la somme des rangs positifs : $R^+ = \sum_{i:d_i>0} r_i$
   et R^- la somme des rangs négatifs : $R^- = \sum_{i:d_i<0} r_i$.
   Sachant que la somme totale des rangs des écarts est égale à $\frac{n(n+1)}{2}$.
   On peut écrire : $R^- = \frac{n(n+1)}{2} - R^+$.

Statistique du test

La statistique du test des rangs signés de Wilcoxon notée W est donnée par :

$$W = \min(R^+, R^-)$$

$$E[W] = \frac{n(n+1)}{4}$$

$$V[W] = \frac{n(n+1)(2n+1)}{24}$$

Règle de décision

Fixant un seuil de signification α, on distingue les deux cas suivants :

- Si $n < 15$: petit échantillon

on rejette l'hypothèse nulle H_0 si :

$$W \geq w_{crit}$$

où : w_{crit} est le seuil critique lu dans la table spécifique de Wilcoxon.

- Si $n \geq 15$: grand échantillon

Dans ce cas, on peut considérer une approximation de la loi normale et par conséquence on utilise la table des valeurs critiques z_α de la loi normale.

$$Z_{obs}(+) = \frac{(R^+ - m)}{\delta}$$
$$Z_{obs}(-) = \frac{(R^- - m)}{\delta}$$

avec : $m = \frac{n(n+1)}{4}$ $\quad et \quad$ $\delta^2 = \frac{n(n+1)(2n+1)}{24}$

Et on compare $Z = \max(Z_{obs}(+), Z_{obs}(-))$ avec la valeur critique z_α de la loi normale.

- Si $Z \geq z_\alpha$: on rejette H_0 au niveau α.
- Si $Z < z_\alpha$: on accepte H_0 au niveau α.

Exemple

On cherche à savoir si un entrainement régulier modifie la tension artérielle des personnes.

On a recueilli la tension systolique de n=8 personnes dont on a appliqué un programme d'entraînement spécifique pendant 6 mois. Après l'entaînement, on a mesuré de nouveau la tension chez ces personnes.

Les résultats obtenus sont les suivants:

N	avant	après	d_i	$\|d_i\|$	rang	rang(+)	rang(-)
1	130	120	10	10	5	5	-
2	170	163	7	7	4	4	-
3	125	120	5	5	2	2	-
4	170	135	35	35	7	7	-
5	130	143	-13	13	6	-	6
6	130	136	-6	6	3	-	3
7	145	144	1	1	1	1	-
8	160	120	40	40	8	8	-
						$\sum$=27	$\sum$=9

Peut-on conclure qu'il y a une amélioration dans la tension artérielle chez ces personnes après les entraînement ?

Hypothèses à tester :

H_0 : "La tension est la même avant et après les entaînement"

H_1 : "La tension après est moins qu'avant"

La statistique du test :

On a $W = \min(R^+, R^-) = \min(27, 9) = 9$

la règle de décision :

D'après la table de Wilcoxon; $w_{crit} = 4$

On remarque que $W > w_{crit}$ donc on rejette H_0 au seuil $\alpha = 5\%$

Conclusion la tension n'est pas la même avant et après l'entraînement : l'entraînement a un effet sur la tension.

3. Test de Kruskal-Walis

C'est un test pour comparer la distribution de $k > 2$ échantillons non appariés c-à-d vérifie si plusieurs échantillons appartiennent à la même population. Il s'agit de l'homologue non paramétrique de l'ANOVA à un facteur, mais avec le sérieux avantages de ne pas tenir compte de la loi de la distribution de la variable étudiée ni l'égalité des variances entre les échantillons. Ce test est une extension généralisée du test de Wilcoxon-Mann-Whitney et par conséquence il fonctionne de la même façon en remplaçant les valeurs de la variable étudiée par leurs rangs.

On admet les notations suivants :

- k : le nombre total d'échantillons.
- N : le nombre total d'observations.
- n_i : le nombre d'observations dans l'échantillon i.
- r_i : la somme des rangs dans l'échantillon i.

La statistique du test

Soit $\bar{r}$ la moyenne globale des rangs et $\bar{r}_i$ la moyenne des rangs dans l'échantillon i.

La statistique de Kruskal-Walis est défini comme suit :

$$K = \frac{12}{N(N+1)} \sum_i n_i (\bar{r}_i - \bar{r})^2 ...(*)$$

Qui est bien l'expression d'une variabilité inter-classe c-à-d la dispersion des moyennes conditionnelles autours de la moyenne globale.

La formule $(*)$ est équivalente à la formule :

$$K = \frac{12}{N(N+1)} \sum_i \frac{r_i^2}{n_i} - 3(N+1)$$

La région critique

1) Pour des effectifs faible, on utilise la table de Kuskal-Walis pour déterminer la valeur critique k_{crit} à ne pas dépasser au seuil α.

$\rightarrow$ On rejette l'hypothèse nulle H_0 qui suppose que les échantillons proviennent des populations identiques si $K > k_{crit}$

2) Pour des effectifs suffisamment grands ($n_i > 5$), la statistique K suit approximativement une loi de khi-deux à (k-1) degré de liberté (χ^2_{k-1}) lorsque l'hypothèse H_0 est vrai.

$\rightarrow$ Dans ce cas, On rejette l'hypothèse nulle H_0 si $K > k_{crit} = \chi^2_{k-1_1-\alpha}$

Exemple

Soit 3 forages d'eau dont on a mesuré la concentration en Magnésium dans l'eau de manière quotidienne pendant 5 jours (mg/l).

forage 1	forage 2	forage 3
15	15	19
20	16	20
20	21	21
22	23	23
25	25	25

Peut-on observer une différence de concentration en Magnésium dans l'eau entre les trois forages?

Solution

Les hypothèses du test

H_0 : "Il n'existe pas une différence de concentration en 3 forages".

H_1 : "Il existe une différence de concentration en 3 forages".

N=15 , $n_i = 5$, k=3

	forage 1	r_1	forage 2	r_2	forage 3	r_3
	15	1.5	15	1.5	19	4
	20	6	16	3	20	6
	20	6	21	8.5	21	8.5
	22	10	23	11.5	23	11.5
	25	14	25	14	25	15
total	/	37.5	/	38.5	/	44
$\bar{r}_i$	/	7.5	/	7.7	/	8.8

On a : $\bar{r}$= 8

et

$$K = \frac{12}{N(N+1)} \sum_i n_i(\bar{r}_i - \bar{r})^2$$

$$K = 0.245$$

$K < k_{crit} = 5.78$, donc on accepte l'hypothèse nulle au seuil α=5.

par conséquent, il n'y a de différence de concentration entre les trois forages.