



FICHE MÉTHODOLOGIQUE

Série de Fiches Méthodologiques en Recherche et en Rédaction Scientifique

Fiche Méthodologique n°8:

Comment évaluer la concordance entre deux mesures qualitatives par le test Kappa ?

How to assess the agreement between two categorical measures using Kappa test?

كيف تقييم التوافق بين تدبيرين نوعيين من قبل اختبار كابا

Houyem Khiari¹, Hajer Hannachi¹, Rym Mallekh¹, Faouzi Mehdi², Mohamed Hsairi¹, Réseau Maghrébin PRP2S*

* Réseau Maghrébin: Pédagogie- Recherche- Publication en Sciences de Santé (PRP2S)

1 : Service d'épidémiologie – Institut de cancérologie Salah Azaiz Tunis

2 : Direction Générale de la Santé Militaire – Tunis

Cette série...

Le Réseau Maghrébin PRP2S et la Rédaction de la revue « La Tunisie Médicale » ont l'honneur de vous présenter, régulièrement à partir du numéro de janvier 2020, une série des fiches techniques en épidémiologie et en bio statistique. Ces fiches méthodologiques décrivent, d'une manière standardisée, les modes d'usage des concepts, des outils et des méthodes utilisés lors des différentes phases de la rédaction médicale scientifique depuis la phase de la recherche documentaire jusqu'à la phase de la communication médicale scientifique.

Cette série est rédigée par des experts de méthodologie de recherche dans les universités du Grand Maghreb et les facultés sœurs au Nord de la Méditerranée. Chaque fiche répond à trois questions essentielles (Quoi ? Pourquoi ? Comment) du concept étudié, en se basant sur un article publié dans la revue Tunis Med.

Le coordinateur de la série « Fiches Méthodologiques ».

Professeur Ahmed Ben Abdelaziz (Président du Réseau Maghrébin PRP2S)
ahmedbenabdelaziz.prp2s@gmail.com

Série des Fiches méthodologiques

Sommaire

Fiche n°1 (janvier 2020):

Comment calculer la taille d'un échantillon pour une étude observationnelle

Serhier Z, et al. (Faculté de Médecine et de Pharmacie de Casablanca, Maroc)

Fiche n°2 (février 2020):

La recherche qualitative: méthodes, outils, analyse

Soulimane A. (Faculté de Médecine, Université Djillali Liabes, Sidi Bel Abbès, Algérie)

Fiche n°3 (mars 2020)

Et Allah ... créa la variabilité

Barhoumi T, et al (Réseau Maghrébin PRP2S)

Fiche n°4 (mai 2020)

Réussir votre recherche bibliographique sur PubMed

Ben Abdelaziz A, et al (Réseau Maghrébin PRP2S)

Fiche n°5 (juin 2020)

Réussir la rédaction de votre « Protocole de Recherche » en sciences de la santé

Ben Abdelaziz A, et al (Réseau Maghrébin PRP2S)

Fiche n°6 (juillet 2020)

Analyse multi variée par régression logistique

Ben Salem K, et al (Réseau Maghrébin PRP2S)

Fiche n°7 (octobre 2020)

Tests non paramétriques pour comparer deux ou plusieurs moyennes sur des échantillons indépendants

Bezzaoucha A, et al (Réseau Maghrébin PRP2S)

Fiche n°8 (novembre 2020)

Comment évaluer la concordance entre deux mesures qualitatives par le test Kappa ?

Khiari H, et al (Réseau Maghrébin PRP2S)

Correspondance

Ahmed Ben Abdelaziz

Laboratoire de Recherche LR19SP01 « Mesure et Appui de la Performance des Etablissements de Santé ». Université de Sousse (Tunisie)

ahmedbenabdelaziz.prp2s@gmail.com

ETUDE DE CAS

Dans l'extrait ci-dessous de l'article [1] intitulé «Etude de l'application des équations de référence spirométriques Caucasiennes pour l'interprétation des données spirométriques d'une population Tunisienne adulte»,

les auteurs ont utilisé le test Kappa afin d'étudier la concordance entre les profils spirométriques, d'un échantillon d'adultes tunisiens, selon deux équations de référence: une Caucasienne (ERS/ECSC1983) et l'autre Tunisienne (Tunisian 1995). Les résultats ont montré un faible niveau du coefficient de concordance Kappa.

«spirometry results (normal, LAOVD, SAOVD, TRVD, MVD) were interpreted for each subject. Spirometry results obtained using the two reference equations were compared after construction of contingency tables. Agreement between classification of spirometry results using the two reference equations were calculated using the Kappa estimate »

«Table 3 exposes the comparison of spirometry interpretation by ERS/ECSC1983 and Tunisian1995 reference equations. »

Table 3. Comparison of spirometry interpretation by Caucasian and Tunisian1995 reference equations.

		Tunisian1995				
		Normal (n=411)	LAOVD (n=80)	SAOVD (n=439)	TRVD (n=227)	MVD (n=35)
ERS/ECSC1983	Normal (n=1015)	411 (100.00)	17 (21.25)	414 (94.31)	170 (74.89)	3 (8.57)
	LAOVD (n=38)	0 (0.00)	26 (32.50)	0 (0.00)	2 (0.88)	10 (28.57)
	SAOVD (n=104)	0 (0.00)	37 (46.25)	25 (5.69)	29 (12.78)	13 (37.14)
	TRVD (n=28)	0 (0.00)	0 (0.00)	0 (0.00)	26 (11.45)	2 (5.71)
	MVD (n=7)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	7 (20.00)

LAOVD: Large Airway Obstructive Ventilatory Defect. SAOVD: Small Airway Obstructive Ventilatory Defect. TRVD: Tendency Through Restrictive Ventilatory Defect. MVD: Mixed Ventilatory Defect.
Data are number (%) of subjects.
Percentage is calculated as following: number of subjects/total of subjects for each spirometry interpretation.

«Table 4 exposes the Kappa estimate of agreement, and percentage of misclassified results, for spirometry interpretation between Tunisian 1995 and ERS/ECSC1983 reference equations. The last demonstrated poor agreement (in females, males and total sample)».

Table 4. Kappa estimate of agreement, and percentage of misclassified results, for spirometry interpretation between Tunisian1995 and Caucasian reference equations.

Males	Females	Total sample
Kappa estimate of agreement (95% confidence limits)		
0.144 (0.090 to 0.137) ^a	0.296 (0.173 to 0.410) ^a	0.124 (0.100 to 0.147) ^a
Percentage of misclassified results		
61.12	30.76	58.47

^ap < 0.05 (Kappa estimate).

QUIZ

- Pourquoi les auteurs de cet article ont utilisé le test Kappa ?
 - Pour tester la liaison entre les deux équations de références spirométrique (Caucasienne et Tunisienne).
 - Pour quantifier l'accord entre les résultats des profils spirométriques des deux équations de références.
 - Pour étudier la validité de l'équation de référence Tunisienne.
- Le Test Kappa dans cet exemple a permis de comparer des variables de nature :
 - Qualitative nominale
 - Quantitative
 - Qualitative ordinale
- Quelle est la valeur de concordance Kappa des résultats spirométriques entre les deux équations de référence pour la totalité de l'échantillon étudié ?
 - 0.144
 - 58.47%
 - 0.124

INTRODUCTION

Dans le domaine médical, les résultats des examens physiques, les interprétations radiographiques ou d'autres tests de diagnostic reposent souvent sur un certain degré d'interprétation subjective par les observateurs. C'est ainsi qu'il est indispensable d'évaluer la validité et la reproductibilité (ou fiabilité) des examens et des mesures en général [2, 3]. Des études qui mesurent l'accord entre deux ou davantage d'observateurs devraient inclure une statistique qui tienne compte du fait que les observateurs sont d'accord ou en désaccord simplement par hasard. Il faut tout d'abord distinguer cette étude de la reproductibilité de celle traitant de la liaison entre deux mesures et qui permet uniquement de mesurer l'association entre deux paramètres. En effet, l'existence d'une liaison statistique, même forte, n'implique pas nécessairement une bonne concordance. L'évaluation de la reproductibilité ou de la concordance suppose de comparer les résultats de deux séries de mesures et non de tester une liaison statistique entre elles. En médecine, la concordance des mesures, obtenues par deux (ou plusieurs) observateurs utilisant la même méthode (reproductibilité inter-observateurs) ou par deux méthodes différentes analysant le même échantillon (échantillon apparié) constitue un problème courant [4].

Plusieurs méthodes d'évaluation de la concordance peuvent être utilisées, le choix de la méthode dépend de la nature de la mesure effectuée [5, 6]. **Le test non paramétrique Kappa de Cohen ou coefficient Kappa** [7] permet de quantifier, en prenant en compte la concordance aléatoire, l'accord entre observateurs ou techniques, lorsque les jugements sont qualitatifs. Une valeur de Kappa égale à 1 indique un accord parfait, tandis qu'une valeur égale à 0 indique un accord équivalent à celui obtenu par le simple fait du hasard. Dans cette fiche méthodologique, nous allons expliquer l'intérêt du test Kappa, en quoi consiste ce test et comment peut-on l'utiliser et l'interpréter.

TEST KAPPA: POURQUOI ?

En lisant la littérature médicale sur le diagnostic et l'interprétation des tests de diagnostic ou de dépistage, notre attention est généralement portée sur la sensibilité, la spécificité, et les valeurs prédictives. Ces paramètres concernent la validité du test. Mais si les cliniciens qui

interprètent les résultats du test ne s'entendent pas sur les conclusions; ces tests seront de peu d'utilité. En effet, de nombreuses situations en médecine nécessitent le recours à plusieurs personnes ou techniques pour étudier les mêmes données. La question de concordance (accord, similitude..) entre les résultats se pose immédiatement. Les mesures de concordance sont à utiliser lorsque l'on souhaite évaluer à quel point deux ou plusieurs estimateurs (exemple: observateurs donnant leur avis sur une même série d'observations, appareils mesurant un paramètre sur une même série d'observations) sont concordants dans le résultat de leur analyse. L'analyse de la reproductibilité ne permet pas de savoir si un estimateur (observateur, méthode) est meilleur que l'autre (et que ses résultats sont justes).

Prenons le cas de deux praticiens examinant un cliché radiologique du même patient et proposent des interprétations différentes. En l'absence d'une référence, cette multiplication des avis n'apporte pas la sécurité et la continuité des soins pour le médecin traitant et le patient. Une solution consiste à estimer le taux d'accord entre les jugements des praticiens par le coefficient Kappa et d'étudier leurs désaccords pour y remédier.

Un autre exemple, on s'intéresse au diagnostic de la pneumonie par l'examen clinique: les résultats montrent qu'il y a une concordance entre la présence de crépitation à l'examen physique et d'une zone de condensation à la radiographie du thorax dans 79% des cas, avec un coefficient Kappa égal à 0.51; de même, il existe une concordance entre la présence d'expectorations purulentes à l'interrogatoire et d'une zone de condensation à la radiographie dans 85% des cas avec un coefficient Kappa égal à 0.01. Comment interpréter alors ces niveaux d'accord en tenant compte des valeurs du coefficient Kappa. Ce type de test revient à faire la distinction entre les véritables concordances entre les estimateurs et celles qui seraient le résultat du hasard. Le but est de déceler et de quantifier les désaccords pour les interpréter ou les corriger.

LE TEST KAPPA: QUOI ?

Le test Kappa est un test non paramétrique qui a été proposé par J Cohen en 1960, et qui permet d'estimer, en prenant en compte la concordance aléatoire, l'accord

entre des jugements catégoriels appliqués aux mêmes objets, fournis par deux ou plusieurs estimateurs (observateurs, techniques, tests..) dans le but de déceler et de quantifier les désaccords pour les corriger. Il est important aussi de souligner que l'accord observé entre un ou plusieurs jugements qualitatifs, résulte de la somme d'une composante d'accord «véritable» (réel) et d'une composante «aléatoire» (hasard).

Par exemple, imaginons une étude dans laquelle deux résidents en médecine de famille évaluent l'utilité d'une série de 100 conférences de formation. Le résident 1 et le résident 2 conviennent que les conférences sont utiles dans 15% des cas et non utiles dans 70% des cas (exemple 1). Toutefois, si les deux résidents assignent au hasard leur évaluation, ils seraient parfois d'accords, uniquement par l'effet du hasard. Le coefficient Kappa nous donne une évaluation numérique du niveau de concordance en soustrayant la part due au hasard. Pour contrôler le hasard, le coefficient Kappa (K) propose de chiffrer l'intensité ou la qualité de l'accord réel. C'est un indice qui permet de «retirer» la portion de hasard ou de subjectivité de l'accord entre les jugements.

TEST KAPPA: COMMENT ?

Afin d'expliquer l'utilisation du test Kappa, on va aborder le cas le plus simple: estimation de la concordance de deux mesures répétées par deux juges en utilisant le logiciel Excel.

1. Concordance entre deux séries de mesures binaire :

1.1. Exemple 1:

Prenons l'exemple 1 (sus cité): on commence par construire le tableau des effectifs croisés ou **effectifs observés** entre les deux variables appariées (tableau 1).

Tableau 1. Niveau d'accord entre deux résidents de médecine de famille sur l'utilité des conférences de formation

		Résident 1		
		Oui	Non	Total
Résident 2	Oui	15	5	20
	Non	10	70	80
	Total	25	75	100

- **Mesure de l'accord observé P_o :** On commence par calculer le pourcentage de concordance observé (P_o) des résultats convergents ou concordants. Il est égal à $(15+70)/100 = 0.85$.
- **Mesure d'accord véritable ($P_o - P_e$) :** On peut se demander s'il existe une partie de cette concordance qui est due au hasard.

On imagine alors une situation d'indépendance totale entre les deux résidents. On calcule alors **les effectifs théoriques**: pour chaque cellule en divisant le produit des effectifs marginaux (total ligne x total colonne) par l'effectif total (N). On obtient : $(20*25)/100=5$; $(80*75)/100=60$.

- **La proportion d'accord observée (P_o)** est le pourcentage de toutes les conférences pour lesquelles les évaluations de deux résidents sont d'accords. Dans notre exemple, il s'agit de $15 + 70/100 = 0.85$.
- **La proportion d'accord aléatoire liée au hasard (P_e) est égale à :** $P_e = (5+60)/100 = 0.65$

		Résident 1		
		oui	non	total
Résident 2	oui	15	5	20
	non	10	70	80
	total	25	75	100

La concordance observée ou proportion d'accord observée (P_o)
est : $(15+70)/100 = 0.85$

		Résident 1		
		oui	non	total
Résident 2	oui	5	15	20
	non	20	60	80
	total	25	75	100

La concordance aléatoire (P_e) ou proportion d'accord liée au hasard (P_e)
est : $(5+60)/100 = 0.65$

- **La mesure d'accord véritable (po-pe)** est la différence entre la **proportion d'accord observée (Po)** ou accord réellement présent (accord «observé») et la **proportion d'accord liée au hasard (Pe)** (accord «aléatoire»).
- **La valeur de Kappa** correspond à: $K = (Po - Pe) / (1 - Pe) = (0.85 - 0.65) / (1 - 0.65) = 0.57$

1.2. Cas général:

Les effectifs observés de l'examen de N patients par deux observateurs pour diagnostiquer un signe clinique (notés a, b, c, d) sont présentés par le tableau 2.

Tableau 2. Tableau de contingence

Observateur 2	Observateur 1		Total
	Signe +	Signe -	
Signe +	a (a')	b (b')	n1=a+b
Signe -	c (c')	d (d')	n2=c +d
Total	t1=a + c	t2=b + d	N

La présentation des résultats sous la forme d'un tableau de contingence montre que a et d correspondent au nombre de fois où les deux observateurs sont en accord (résultats concordants) alors que b et c correspondent au nombre de fois où les deux observateurs sont en désaccord (résultats discordants). S'il y a absence de désaccord entre observateurs, b et c seraient égales à 0 et le pourcentage d'accord observé (Po) serait égal à 1 ou 100% alors que s'il y a absence d'accord, a et d seraient égales à 0 et le pourcentage d'accord observé (Po) serait égal à 0.

- Dans une première étape on calcule **les effectifs attendus**: Les effectifs attendus se calculent pour chaque cellule en divisant le produit des effectifs marginaux (total ligne × total colonne) par l'effectif total (N). Par exemple $a' = [(a + b)(a + c)]/N = n1 * t1 / N$
- A partir de ces effectifs on déduit **la concordance observée et aléatoire**:

La proportion d'accord observée (Po) ou concordance observée (Co) égale à : $Po = (a+d)/N$.
 La proportion d'accord aléatoire (Pe) ou concordance aléatoire (Ca) égale à : $Pe = (a' + d')/N$

- **La concordance exprimée par le coefficient Kappa (K)** est égale à :

$$K = \frac{Po - Pe}{1 - Pe}$$

Avec

Po : La proportion d'accord observée ou concordance observée
 Pe : La proportion d'accord aléatoire ou concordance aléatoire

Remarque1 : Dans le cas particulier où deux catégories de réponses sont proposées, la formule de Kappa peut s'écrire :

$$K = \frac{2(ad - bc)}{n1.t2 + n2.t1}$$

Remarque 2: la valeur du coefficient Kappa est indépendante de la taille de l'échantillon étudié. Par exemple, si on multiplie par 10 chacun des effectifs des cases du tableau de contingence, on obtient alors le même coefficient Kappa que pour le premier échantillon N, mais la signification statistique de la valeur du coefficient Kappa sous l'hypothèse nulle ne sera plus grande.

- **Interprétation du coefficient Kappa :**

Le coefficient Kappa s'utilise comme un indicateur descriptif de la concordance. Il varie entre 0 et 1. La concordance est d'autant plus importante que le coefficient est proche de 1. Le coefficient Kappa est égal à 0 quand la concordance observée est égale à la concordance aléatoire. Rarement, un coefficient Kappa négatif peut être obtenu quand la concordance observée est moins bonne que la concordance aléatoire. Il s'interprète alors comme une absence de concordance. Le tableau 3 résume les valeurs de référence pour interpréter le coefficient Kappa selon Landis et Koch [8].

Tableau 3. Valeurs de référence utilisées pour interpréter le coefficient Kappa selon Landis et Koch

Accord	Kappa
Excellent	0,81
Bon	0.61- 0,80
Modéré	0.21- 0,60
Mauvais	0.00- 0,20
Très mauvais	<0,00

Les limites de ce classement sont donc arbitraires et peuvent varier selon l'étude réalisée; par exemple en psychiatrie où la part d'incertitude est grande, un accord «modéré» dans l'échelle proposée ci-dessus pourrait être considéré comme très satisfaisant, tel n'étant pas le cas devant un résultat d'analyse biologique. L'interprétation du coefficient Kappa doit se faire en fonction du contexte clinique et de l'utilisation de l'outil de mesure. En général, on accompagne le calcul du coefficient K de son degré de signification p afin de pouvoir interpréter le résultat avec certitude. Pour tester l'hypothèse nulle H0 «indépendance des jugements» (K = 0) contre l'hypothèse alternative H1

(K ≠ 0) on a recours au test du Kappa.

2. Concordance entre deux séries de mesure à plus de deux catégories :

La concordance entre deux mesures catégorielles à plus de deux classes est également évaluée par le coefficient Kappa. Le principe du calcul est le même que dans l'exemple précédent.

2.1. Exemple 2 :

Prenant l'exemple du «cas étudié» :

Table 3. Comparison of spirometry interpretation by Caucasian and Tunisian₁₉₉₅ reference equations.

		Tunisian ₁₉₉₅				
		Normal (n=411)	LAOVD (n=80)	SAOVD (n=439)	TRVD (n=227)	MVD (n=35)
ERS/ECSC ₁₉₈₃	Normal (=1015)	411 (100.00)	17 (21.25)	414 (94.31)	170 (74.89)	3 (8.57)
	LAOVD (n=38)	0 (0.00)	26 (32.50)	0 (0.00)	2 (0.88)	10 (28.57)
	SAOVD (n=104)	0 (0.00)	37 (46.25)	25 (5.69)	29 (12.78)	13 (37.14)
	TRVD (n=28)	0 (0.00)	0 (0.00)	0 (0.00)	26 (11.45)	2 (5.71)
	MVD (n=7)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	7 (20.00)

LAOVD: Large Airway Obstructive Ventilatory Defect. SAOVD: Small Airway Obstructive Ventilatory Defect. TRVD: Tendency Through Restrictive Ventilatory Defect. MVD: Mixed Ventilatory Defect.
Data are number (%) of subjects.
Percentage is calculated as following: number of subjects/total of subjects for each spirometry interpretation.

La concordance observée ou Proportion d'accord observée est :

$$Po = (411+26+25+25+26+7)/1195=41,6\%$$

La concordance aléatoire ou Proportion d'accord aléatoire est :

$$Pc = ((1015*411)/1195 + (38*80)/1195) + (104*439)/1195 + (31*230)/1195 + (7*25)/1195 = 31.1\%$$

$$Kappa = (Po - Pc) / (1 - Pc) = (41.6 - 31.1) / (100 - 31.1) = 0.12$$

Le coefficient Kappa est de 0.12 et qui correspond selon Landis et Koch à un mauvais accord.

Kappa estimate of agreement, and percentage of misclassified results, for spirometry interpretation between Tunisian₁₉₉₅ and Caucasian reference equations.

Males	Females	Total sample
Kappa estimate of agreement (95% confidence limits)		
0.144 (0.090 to 0.137) ^a	0.296 (0.173 to 0.410) ^a	0.124 (0.100 to 0.147) ^a
Percentage of misclassified results		
61.12	30.76	58.47

^ap < 0.05 (Kappa estimate).

2.2. Cas général :

modalités	1	2	...	i	Total A
1	O_{11}				n_1
2		O_{22}			n_2
...
i				O_{ii}	n_i
Total B	t_1	t_2		t_i	N

$$Co = \frac{O_{11} + O_{22} + \dots + O_{ii}}{N} \quad Ca = \frac{t_1 n_1 + t_2 n_2 + \dots + t_i n_i}{N^2}$$

$$K = \frac{Co - Ca}{1 - Ca}$$

- ❖ **Co: la concordance observée** : c'est la proportion des individus classés dans les **cases diagonales** de concordance du tableau de contingence, soit la somme de ces effectifs diagonaux, divisée par la taille de l'échantillon (n).
- ❖ **Ca: la concordance aléatoire** qui est égale à la somme des produits des effectifs marginaux, divisée par le carré de la taille de l'échantillon.

Cette situation n'est autre que celle vérifiant l'hypothèse nulle dans le cadre d'un test du chi-deux.

2.3 Calculs complémentaires :

- **L'intervalle de confiance du test Kappa (exemple: voir fiche Excel*)**

- **Le Test de Kappa :**

La valeur vraie du coefficient Kappa dans la population est une variable aléatoire qui suit approximativement une loi de Gauss de moyenne K et de variance $\text{Var}(K)$.

Hypothèse nulle (H_0): $Kappa = 0$. L'accord est le même que l'accord aléatoire.

Hypothèse alternative (H_a): $Kappa \neq 0$. L'accord est différent d'un accord aléatoire.

3. Le Kappa pondéré:

Parfois, on s'intéresse à l'accord entre les principales catégories dans lesquelles il existe une grande différence. Par exemple, supposons que nous avons cinq catégories de l'utilité des conférences de midi: «très utile», «utile», «neutre», «peu utile» et «très peu utile». Dans ce cas, on ne s'inquiète pas si un résident classe une conférence comme «très utile» tandis qu'un autre la classe comme «utile», mais plutôt si un résident classe comme «très utile» une conférence jugée comme «très peu utile» par l'autre. Prenons l'exemple clinique de l'interprétation d'une mammographie par deux radiologues. Le cas où l'un conclut à une radio normale et l'autre à une lésion bénigne n'est pas très inquiétant alors qu'on s'inquiète vraiment lorsque l'un la classe comme normale et l'autre comme lésion cancéreuse.

Le Kappa pondéré peut être utile si l'on souhaite pondérer des discordances «plus graves» que d'autres (par exemple en cas de notation sur une échelle, une discordance d'un point sera moins délétère qu'une discordance de plusieurs points). Dans l'exemple clinique précédent, un désaccord entre normal et bénin peut être interprété comme un accord partiel, cependant un désaccord entre normal et cancer serait plutôt une absence d'accord.

Pour calculer le **Kappa pondéré**, Cohen propose de donner à chacune des cases du tableau de contingence, un **pois** fixé a priori qui reflète l'importance que l'on attribue

au désaccord. Cependant, la détermination des poids pour un Kappa pondéré reste une question subjective [9].

CONCLUSION:

Le test Kappa est la mesure la plus souvent mentionnée dans la littérature médicale pour résumer l'accord entre deux observateurs concernant des mesures qualitatives. Le Kappa nous renseigne sur la fraction de concordance véritable (épurée des accords par le simple effet du hasard). Ce test est particulièrement utile dans le domaine de l'imagerie.

L'essentiel à retenir

Le test Kappa mesure la concordance entre deux mesures qualitatives; il est très utile notamment en imagerie pour mesurer la concordance entre deux radiologues concernant l'interprétation de clichés radiologiques. Comme il est possible d'avoir un certain accord entre deux observateurs, rien que par l'effet du hasard, le test Kappa offre la possibilité d'avoir une mesure épurée de cet effet du hasard.

C'est ainsi que le test Kappa (k) est égal à :

$$K = \frac{P_o - P_e}{1 - P_e}$$

avec P_o représentant la proportion d'accord observée (ou concordance observée) et P_e la proportion d'accord aléatoire ou concordance aléatoire.

La concordance est d'autant plus importante que le coefficient est proche de 1. Cette concordance est jugée excellente si kappa est supérieur à 0.81, bonne si kappa est compris entre 0.61 et 0.80, modérée si Kappa varie entre 0.21 et 0.60, mauvaise si kappa est entre 0.00 et 0.20 et très mauvaise si Kappa est négatif.

Réponses aux questions de quizz

1.b

2.a

3.c

* Pour avoir une copie du fichier Excel de calcul de l'Intervalle de Confiance, autour de l'Indice de Kappa, prière d'écrire au coordinateur de la série de cette collection des Fiches Méthodologiques, le Professeur Ahmed Ben Abdelaziz, à l'adresse suivante: ahmedbenabdelaziz.prp2@gmail.com

RÉFÉRENCES

1. El Attar M N, Hadj Mabrouk K, Ben Abdelaziz A, et al. Etude de l'application des équations de référence spirométriques Caucasiennes pour l'interprétation des données spirométriques d'une population Tunisienne adulte. *Tunis Med* 2014; 92(8-9):574-580.
2. Desquilbet L. Guide pratique de validation statistique de méthodes de mesure: répétabilité, reproductibilité et concordance. HAL. 2019. Disponible sur <https://hal.archives-ouvertes.fr/hal-02103716v1>.
3. Arnal F. Répétabilité & reproductibilité d'une méthode. Disponible sur <http://flarnal.e-monsite.com/medias/files/repet-1.pdf>. 2017.
4. Fuhrman C, Chouaid C. Concordance between two variables: numerical approaches (qualitative observations - the kappa coefficient; quantitative measures). *Rev Mal Respir* 2004 ;21(1):123-5.
5. Elie C, Colombet I. Méthodes d'estimation de la reproductibilité. *Sang Thrombose Vaisseaux* 2011; 23(3):138-45.
6. Fermanian J. Measurement of agreement between 2 judges. Qualitative cases. *Rev Epidémiol Sante Publique*. 1984;32(2):140-7.
7. Cohen J. A coefficient of agreement for nominal scales. *Educational Psychological Measurement* 1960; 20(1):37-46.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977: 159-17.
9. Viera AJ, Garrett JM. Understanding inter observer agreement: the Kappa statistic. *Fam Med* 2005; 37(5):360-3.