

# TEST KAPPA

*Kappa de Cohen* **K**

El Missaoui Imane - Bakary Nene - Sabri Ouahiba - Djohi Ezechiel

# Histoire :

- **Fiabilité inter-juges** : indice statistique qui mesure la cohérence des évaluations effectuées par deux ou plusieurs juges ou évaluateurs indépendants sur un même ensemble de données. Elle est importante car elle représente la mesure dans laquelle les données recueillies dans l'étude sont des représentations correctes des variables mesurées.
- Traditionnellement mesuré en **% d'accord**
- En **1960**, Jacob Cohen a critiqué l'utilisation du pourcentage d'accord en raison de son incapacité à tenir compte de l'accord aléatoire. Il a introduit le kappa de Cohen, développé pour tenir compte de la possibilité que les évaluateurs devinent en fait sur au moins certaines variables en raison de l'incertitude.



Le kappa est un **coefficient de corrélation** de symbol **κ** fréquemment utilisé pour tester la **fiabilité inter-juges**.

# Caractéristiques du $\kappa$ :

- Permet de contrôler ce facteur de **concordance aléatoire**.
- Variable qualitative nominale
- Le kappa peut **varier de -1 à +1**
- Le kappa est une **valeur standardisée** et est donc interprétée de la même manière dans plusieurs études.
- 2 évaluateurs

# Comment est calculé le k ?

Juge 1

Modalités	1	2		i	total A
1	O11				n1
2		O22			n2
...	...	...		...	...
i				Oii	ni
total B	t1	t2		ti	N

Juge 2

Concordance réelle

$$kappa(\kappa) = \frac{P_o - P_e}{1 - P_e}$$

Concordance parfaite

Po : représente l'accord réel observé  
 Pe : représente l'accord aléatoire

$$Po = \frac{o11 + o22 + .... + o_{ii}}{N}$$

$$Pe = \frac{t1n1 + t2n2 + ... + tini}{N^2}$$

# Exemple de calcul du Kappa :

		Rater 1		Row marginals
		Normal	Abnormal	
Rater 2	Normal	147	3	150
	Abnormal	10	62	72
Column marginals		157	65	222

$$Po = \frac{147 + 62}{222} = 0.94$$

$$Pe = \frac{(150 * 157) + (72 * 65)}{222^2} = 0.57$$

→ le pourcentage d'accord est de 0,94 alors que le Kappa est de 0,85

Une réduction considérable du niveau de congruence.  
Accord aléatoire important → valeur résultante du kappa est faible.

$$\text{kappa} = \frac{0.94 - 0.57}{1 - 0.57} = 0.85$$

# Comment on l'interprète ?

Coefficient Kappa	Estimation du degré de concordance
0,8 à 1	Excellent
0,6 à 0,8	Bon
0,4 à 0,6	Moyen
0,2 à 0,4	Faible
0 à 0,2	Négligeable
< 0	Mauvais

# Calcul sur R & Python

En **Python** :

**sklearn.metrics** : ce package fournit la fonction `cohen_kappa_score()` qui permet de calculer le coefficient kappa.

En **R** :

**psych** : ce package fournit la fonction `cohen.kappa()` qui permet également de calculer le coefficient kappa.

# Intervalle de confiance et comparaison de deux Kappa :

Variance du kappa :

$$Var(kappa) = \frac{Po(1 - Po)}{n(1 - Pe)^2}$$

Ecart type du kappa en prenant la racine carrée de la variance :

$$Sk = \sqrt{Var(kappa)}$$

Calculez l'intervalle de confiance à 95 % :

$$IC \text{ à } 95 \% = kappa \pm (1,96 \times Sk)$$

Comparaison de deux Kappa Test de Wald :

$$Z = \frac{(Kappa \text{ estimé} - Kappa \text{ fixe})^2}{Var(kappa \text{ estimé})}$$



# Intervalle de confiance et comparaison de deux Kappa : (Application)

		Reviewer 1	
		Inclus	Exclus
Reviewer 2	Inclus	80	20
	Exclus	10	90

Calcul de l'intervalle de confiance

$$\text{kappa} = 0.7$$

$$S_k = 0.05049$$

$$\text{IC à 95 \%} = \text{kappa} \pm (1.96 \times S_k)$$

$$\text{IC} = [0.6011; 0.7989]$$

Comparaison de deux Kappa

**H0** : kappa estimé = kappa fixe

**H1** : kappa estimé  $\neq$  kappa fixe

Kappa fixe = 0.80


Valeur de Z de Wald

$$Z = 3.92 > 3.84$$

Conclusion: rejet de H0

Tests de mesure de la fiabilité inter-juge	contextes d'utilisation	Type de variables	Limites
Kappa de Cohen	Évaluation de concordance entre deux évaluateurs	variable qualitative nominale	<ul style="list-style-type: none"> <li>non adapté pour plus de 2 juges</li> <li>sensible à la prévalence et au déséquilibre des effectifs marginaux</li> </ul>
Kappa de Cohen pondéré	extension du kappa de cohen prenant en compte les catégories nominales ordinales ou pondérées	variable nominale ordinale et pondérée	peut être sensible à la façon dont les poids sont choisis
kappa Fleiss	extension du kappa de Cohen entre trois ou plusieurs évaluateurs	variable qualitative nominale	difficile à interpréter si le nombre de catégories nominales est élevé
Kendalls w	lorsque on souhaite évaluer la concordance entre plus de deux évaluateurs.	variable qualitative ordinale	difficile à interpréter si les classements sont complexes ou si le nombre d'évaluateurs est élevé
Intra-class correlation	mesure la concordance entre les évaluateurs en termes de mesures continues	variable métrique	difficile à interpréter si l'échelle de mesure n'est pas bien définie

# Conclusion:

- Mesure de la fiabilité inter-juge
- Valeur seuil de significativité dépend du contexte d'utilisation
- Différents types de tests selon le contexte d'étude
-  N'indique pas la véracité des résultats

**Merci de votre attention :)**

Limites	Alternatives
Sensibilité aux erreurs systématiques entre les juges du à un déséquilibre des effectifs marginaux	Kappa de Fleiss
Sensibilité à la prévalence du signe cherché	Solution de Feinstein et Cicchetti: calcul des proportions observées d'accords positifs et négatifs  Kappa max (Hui et Walter)

**Tableau VI - Etude avec des effectifs marginaux non biaisés**

Résultat du test B	Réponses	Résultat du test A		Total
		+	-	
	+	40	20	60
	-	20	20	40
	Total	60	40	100

**Tableau VII - Etude avec des effectifs marginaux biaisés**

Résultat du test B	Réponses	Résultat du test A		Total
		+	-	
	+	40	35	75
	-	5	20	25
	Total	45	55	100

Le tableau [VI](#) donne un Kappa = **0,17** ( $P_o = 0,6$  et  $P_e = 0,52$ ) et le tableau [VII](#) un Kappa = **0,24** ( $P_o = 0,6$  et  $P_e = 0,475$ ).

# Calcul sur R & Python

En **Python** :

**sklearn.metrics** : ce package fournit la fonction **cohen\_kappa\_score()** qui permet de calculer le coefficient kappa.

```
from sklearn.metrics import cohen_kappa_score

# Les deux évaluateurs ont noté les mêmes objets selon l'échelle de 1 à 5
eval1 = [1, 2, 3, 4, 5]
eval2 = [1, 2, 2, 4, 5]

# Calcul du coefficient kappa
kappa = cohen_kappa_score(eval1, eval2)

print("Coefficient kappa :", kappa)
```

Coefficient kappa : 0.75

En **R** :

**psych** : ce package fournit la fonction **cohen.kappa()** qui permet également de calculer le coefficient kappa.

```
library(psych)

# Calcul du coefficient kappa
kappa <- cohen.kappa(data.frame(eval1, eval2))

# Affichage du coefficient kappa
print(paste("Coefficient kappa :", kappa))
```

# Exemples de cas d'utilisation du kappa :

## Etudes sur les escarres

Des études sur les escarres lorsque les variables comprennent des éléments tels que la quantité de rougeur, d'œdème et d'érosion dans la zone affectée. Alors que les collecteurs de données peuvent utiliser des outils de mesure pour la taille, la couleur est assez subjective, tout comme l'œdème

## Les frottis de Papanicolaou

En laboratoire, on a constaté que les personnes qui lisent les frottis de Papanicolaou (Pap) pour le cancer du col de l'utérus interprètent différemment les cellules sur les lames.

## Traumatismes crâniens

Dans les recherches sur les traumatismes crâniens, les collecteurs de données estiment la taille des pupilles du patient et la mesure dans laquelle les pupilles réagissent à la lumière en se contractant.