

Fórum:Testy2/Cohenovo kappa

Cohenovo kappa ^[1] měří shodu mezi dvěma hodnotiteli, kteří hodnotí stejnou skupinu n studentů. Shodu by bylo možné vyčíslit jednoduše v procentech. Cohenovo kappa je univerzálnější v tom smyslu, že bere v úvahu také pravděpodobnost náhodné shody.

Cohenovo kappa κ je dáno vztahem:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)},$$

kde $\Pr(a)$ je relativní shoda mezi hodnotiteli a $\Pr(e)$ je odhad pravděpodobnosti náhodné shody. Pokud se hodnotitelé shodli v hodnocení všech jedinců, pak $\kappa = 1$. Pokud je celkové procento shody rovno pravděpodobnosti očekávané shody při náhodném rozhodování, je $\kappa = 0$. Pokud je dokonce procento shody menší, je κ záporné. Interpretace hodnot κ je obvykle následující: Hodnotu $\kappa > 0,75$ považujeme za výbornou shodu, κ mezi 0,40 a 0,75 za dobrou shodu a $\kappa < 0,40$ považujeme za špatnou shodu.

■ Příklad

Představme si, že testujeme dvěma testy 5 studentů a obdržíme tyto výsledky:

Tab. 10.5 Výsledky studentů ve dvou testech

	skóre v prvním testu	skóre v druhém testu
1. student	18	48
2. student	45	75
3. student	33	63
4. student	48	78
5. student	51	81

Pearsonův korelační koeficient je roven 1, mezi dvěma výsledky je přímo lineární závislost (druhý je vždy právě o 30 bodů vyšší než první). Budeme-li však rozhodovat u složení zkoušky na základě 50 bodové hranice, první test úspěšně absoluuje 1 student z 5, zatímco v druhém to budou 4 studenti z 5. Testy by rozhodly o neúspěchu shodně jen u prvního a posledního studenta, tedy ve 40 % případů. Jaká je pravděpodobnost náhodné shody testů? První test rozhoduje o úspěchu v 20 % případů a druhý v 80 % případů, náhodné shodné rozhodnutí o úspěchu by tedy nastalo v 20 % · 80 % = 16 % případů, podobně náhodné shodné rozhodnutí o neúspěchu by nastalo v 80 % · 20 % = 16 % případů, celkem tedy náhodná shoda nastává v 32 %. Cohenovo kappa je proto

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} = \frac{0,40 - 0,32}{1 - 0,32} = 0,12,$$

což svědčí o velmi malé shodě mezi dvěma testy.

■ Příklad

Představme si situaci, kdy o udělení atestace rozhodují nezávisle dvě komise. Každý atestovaný je hodnocen dvakrát a každá komise rozhodne o udělení atestace buď kladně "Ano", nebo záporně "Ne". Výsledky jsou jako v tabulce 10.6, přičemž řádky odpovídají komisi A a sloupce komisi B:

Tab. 10.6

		B	B	
		Ano	Ne	Celkem
A	Ano	20	5	25
A	Ne	10	15	25
	Celkem	30	20	50

Z celkem 50 atestovaných tedy bylo 20 atestovaných hodnoceno kladně oběma komisemi a 15 atestovaných hodnoceno záporně. Procento shody je tedy $\Pr(a) = (20 + 15)/50 = 0,70$.

Abychom odhadli pravděpodobnost náhodné shody $\Pr(a)$, všimněme si nejdříve, že:

- komise A hodnotila kladně 25 atestovaných a záporně také 25 atestovaných. Kladně tedy hodnotila v 50 % případů.
- komise B hodnotila kladně 30 atestovaných a záporně také 20 atestovaných. Kladně tedy hodnotila v 60 % případů.

Proto pokud by komise rozhodovaly náhodně, pravděpodobnost, že obě komise řeknou "Ano" je $0,50 \cdot 0,60 = 0,30$ a pravděpodobnost, že obě řeknou "Ne" je $0,50 \cdot 0,40 = 0,20$. Celková pravděpodobnost náhodné shody je tedy $\Pr(a) = 0,30 + 0,20 = 0,50$.

Po doplnění do vzorce pro Cohenovo kappa dostáváme:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} = \frac{0,70 - 0,50}{1 - 0,50} = 0,40.$$

Další koeficienty shody

Cohenovo kappa lze rozšířit na případ více kategorií. V případě, kdy nás zajímá více hodnotitelů, je namístě použít Fleissovo kappa ^[2] ^[3].

Reference

1. COHEN, Jacob. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960, roč. -, vol. 20, no. 1, s. 37-46, ISSN (Print) 0013-1644, (Online) 1552-3888.
2. FLEISS, Joseph L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971, roč. -, vol. 76, no. 5, s. 378-382, ISSN (Print) 0033-2909, (Online) 1939-1455.
3. FLEISS, Joseph L. *Statistical methods for rates and proportions*. 2. vydání. New York : Wiley, 1981. ISBN 9780471064282.