

# LINEÁRNÍ REGRESE

**regrese** - vystižení charakteru závislosti mezi dvěma nebo více náhodnými veličinami (v případě více než dvou náhodných veličin jde o vícenásobnou regresi)

**metoda nejmenších čtverců odchylek** - metoda pro nalezení koeficientů zvolené regresní funkce

**korelace** - míra (intenzita) závislosti

**koeficient korelace** - hodnoty se pohybují v intervalu  $[-1,1]$ , přičemž hodnoty blízké  $|1|$  znamenají, že existuje silná závislost mezi proměnnými (odmocnina z koeficientu determinace)

**index korelace** - hodnoty z intervalu  $[0,1]$ ; slouží k posouzení vhodnosti regresní funkce

**koeficient determinace** - udává kvalitu regresního modelu, poskytuje informaci o tom, kolik procent rozptylu vysvětlované proměnné je vysvětleno modelem (jak velká část hodnot je vysvětlena regresní přímkou)

Postup (Excel):

1) vytvoříme graf ( x ... nezávislá proměnná, y ... závislá proměnná)

Vložení -> Bodový graf

2) proložíme graf regresním modelem

Nástroje grafu -> Rozložení -> Spojnice trendu (Další možnosti ...)

- vybereme Typ trendu

- zaškrtneme Zobrazit rovnici v grafu, Zobrazit hodnotu spolehlivosti R

Příklad: Vyšetřujeme souvislost znečištění ovzduší oxidy síry a výskyt alergií u dětí do 15 let. V 10 obcích jsme zjistili průměrnou koncentraci (v  $\mu\text{g}/\text{m}^3$ ) a procento dětí s alergií:

koncentrace:

0,10 0,12 0,21 0,15 0,05 0,10 0,10 0,12 0,15 0,11

alergie:

15 20 28 11 7 18 20 13 25 10

Souvisí výskyt alergií se znečištěním?

TESTY

## Příklad testu:

- 1) Kombinace  $k$ -té třídy z  $n$  prvků jsou
  - a) uspořádané skupiny po  $k$  prvcích
  - b) skupiny po  $k$  prvcích vybraných z  $n$  prvků bez ohledu na uspořádání
  - c) uspořádané skupiny po  $n$  prvcích
  
- 2) Jsou dány dva jevy  $A$ ,  $B$ , přičemž  $A$  je podmnožinou  $B$ . Pak platí:
  - a)  $A \cdot B = A$
  - b)  $A \cdot B = B$
  - c)  $A \cdot B = A + B$
  
- 3) Číselná charakteristika směrodatná odchylka
  - a) charakterizuje rozptýlenost hodnot NV kolem její střední hodnoty
  - b) vyjadřuje, zda je rozložení náhodné veličiny symetrické
  
- 4) Parametry hypergeometrického rozdělení jsou:
  - a) počet pokusů, pravděpodobnost úspěšnosti v každém pokusu, počet prvků základního souboru
  - b) počet pokusů, počet prvků základního souboru, počet prvků základního souboru s požadovanou vlastností
  - c) počet pokusů, střední hodnota, rozptyl

5) Necht' NV představuje vzdálenost mezi dvěma poruchami ve struktuře krystalu, pokud známe průměrnou vzdálenost. Jaké má tato NV rozložení pravděpodobnosti?

- a) Poissonovo
- b) exponenciální
- c) normální

6) Počet všech prvků statistického souboru se nazývá

- a) základní soubor
- b) rozsah souboru
- c) statistická jednotka

7) Pomocí centrálního empirického momentu 2. řádu je vyjádřena momentová charakteristika

- a) empirický rozptyl
- b) empirický koeficient šikmosti
- c) empirický koeficient špičatosti

8) Blíží-li se hodnota indexu korelace k nule, pak

- a) se mezi danými veličinami jedná vždy o slabou lineární závislost
- b) se mezi danými veličinami jedná vždy o silnou lineární závislost
- c) vypočtená regresní funkce nevystihuje statistickou závislost mezi veličinami

9) Stupeň spolehlivosti vyjadřuje pravděpodobnost toho, že skutečná hodnota odhadovaného parametru

- a) leží uvnitř intervalu spolehlivosti
- b) neleží uvnitř intervalu spolehlivosti
- c) překročí  $\alpha$ -kvantil očekávaného rozdělení

10) Významnost rozdílu dvou rozptylů testujeme pomocí

- a)  $F$ -testu
- b)  $t$ -testu

---

Na procvičení:

Závěrečný kontrolní test I.