Inferenční statistika - úvod

- z-skóry
- normální rozdělení
- pravděpodobnost
- rozdělení výběrových průměrů

Pravděpodobnost

- postupy induktivní statistiky vycházejí z teorie pravděpodobnosti
- **pravděpodobnost**, že nastane určitý výsledek, **definujeme** jako podíl

počet pokusů, kdy nastal jev A

deleno

celkový počet jevů

- pravděpodobnost bývá uváděna nejčastěji jako **podíl** (0,33), **zlomek** (1/3) nebo **procento** (33,3%)
- pravděpodobnost určitého jevu nebo třídy jevů můžeme odhadnout z rozdělení hodnot (četností)

Pravděpodobnost - příklady

- představme si, že máme krabici se 40 očíslovanými žetony s čísly 1 – 5
- v tabulce jsou uvedeny absolutní i relativní četnosti jednotlivých čísel žetonů

- hodnoty jsou v pořadí: číslo žetonu, absolutní četnost, relativní četnost

Tabulka:

- 1 4 0,10
- 2 8 0,20
- 3 16 0,40
- 4 10 0,25
- 5 2 0,05
- vaším úkolem je vytáhnout 1 žeton
- **jaká je pravděpodobnost, že vytáhnete žeton s číslem 3?**
- **p (3) = f/N = 16/40 =0,40**
  nebo 2/5 či 40%

- **Jaká je pravděpodobnost, že vytáhnete žeton s číslem vyšším než 2?**
  $p(X > 2) = ?$
  0,05 + 0,25 + 0,40 = **0,70**

tj. 70%

- **Jaká je pravděpodobnost, že vytáhnete žeton s číslem nižším než 5?**
  $p(X < 5) = ?$
  0,10 + 0,20 + 0,40 + 0,25 = **0,95**

tj. 95%

Jaká je pravděpodobnost, že vytáhnete žeton s číslem nižším než 4 a vyšším než 1?
$p(4 > X > 1) = ?$
0,20 + 0,40 = **0,60**

tj. 60%

Pravděpodobnost

- pravděpodobnost odpovídá hustotě **oblasti pod křivkou** pro daný interval

Z-skóry

- transformace hodnot proměnné
- umožňují najít a popsat **pozici každé hodnoty** v rámci rozdělení hodnot
- a také **srovnávání hodnot** pocházejících z měření **na rozdílných stupnicích**
- hrubé skóry jsou převedeny na **standardizovanou stupnici** (jednotkou je směrodatná odchylka)

Z-skóry - příklad

- např. skóry ze dvou testů – biologie a psychologie
- v testu z biologie byl průměr celého ročníku m=18 (sd=6)
- v testu z psychologie byl průměr celého ročníku m=500 (sd=100)
- student získal 26 bodů z biologie a 620 z psychologie. Ve kterém předmětu byl lepší?
- přímé porovnání není snadné – skóry z obou testů mají rozdílné průměry i směrodatné odchylky
- z skór =odchylka skóru od průměru vzhledem k velikosti směrodatné odchylky
- z = odch. od průměru/směr. odch.
- skór z biologie: (26-18)/6 = 1,33
- skór psychologie: (620-500)/100=1,2
- v biologii byl student lepší: 1,33 směrodatné odchylky nad průměrem

Z-skóry

- z-skór přesně udává pozici každé hodnoty vzhledem k ostatním hodnotám
- znaménko (+ nebo -) ukazuje, zda je hodnota nad nebo pod průměrem rozdělení
- hodnota z-skóru upřesňuje, kolik směrodatných odchylek byla hodnota od průměru vzdálena
- průměr rozdělení z-skórů je vždy 0

- směrodatná odchylka je 1

Z-skóry

vzorec pro výpočet z-skóru hodnoty X

- <u>u populace</u>: $z = (X_i - \mu) / \sigma$

- <u>u vzorku</u>: $z = (X_i - m) / s$

Z-skóry

- podobně můžeme i z-skór převést na hrubý skór, známe-li průměr a směrodatnou odchylku
- např. u stupnice IQ
- $\mu = 100$, $\sigma = 15$
- pro osobu se z=-3 (3 směrodatné odchylky pod průměrem) bude IQ ?

$$X = Z * \sigma + \mu$$
$$X = -3 * 15 + 100$$
$$X = 55$$

Rozdělení z-skórů

- **tvar** rozdělení z-skórů je **stejný** jako tvar původního rozdělení hrubých skórů
- průměr je 0, směrodatná odchylka 1
- transformace změní jen označení hodnot na ose X

Normální rozdělení

- normální rozdělení je <u>symetrické</u>, <u>unimodální</u>, zvonovitého tvaru
- označuje se i jako <u>Gaussova křivka</u>

- 34.13% skórů spadá mezi průměr a 1 směr. odchylku
- 13.59% hodnot spadá mezi 1. a 2. směr. odchylku
- 2.28% hodnot spadá mezi 2. a 3. směr. odchylku

Normální rozdělení

- tabulka normálního rozdělení (z rozdělení)
- důležitý nástroj, obvykle jako apendix v učebnicích statistiky (spolu s dalšími tabulkami)
- umožňuje zjistit hustotu oblasti pod křivkou (tj. pravděpodobnost) pro jednotlivé z-skóry

Normální rozdělení - příklady

- postup při zjišťování **pravděpodobnosti** z tabulky:
  - načrtnout si normální rozdělení, s hodnotou průměru a směr. odch.
  - zakreslit hledanou hodnotu (v přibližné vzdálenosti od průměru), vystínovat hledanou oblast
  - převést hodnotu X na z-skór
  - najít v tabulce pravděpodobnost

Normální rozdělení - příklady

- Kolik procent osob z populace má IQ 130 nebo vyšší? (μ = 100, σ =15)

otázku je možno formulovat i jako:

- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 130 nebo vyšší? (μ = 100, σ =15)
- z = (130 – 100)/15
- **z = 2**

- z tabulky z-rozdělení plyne, že p = 0.0228 tj. **2,3%**

Normální rozdělení - příklady

- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 85 nebo nižší?

z = (100-85)/15

z = -1

- p = 0.5 – 0.3413 = **0.1587**
- tj. **15,9%**

Normální rozdělení - příklady

- postup při zjišťování **z-skóru** z tabulky:
  - načrtnout si normální rozdělení
  - vystínovat oblast odpovídající zadané pravděpodobnosti
  - v tabulce vyhledat příslušný z-skór
  - vypočítat z něj hrubý skór

Normální rozdělení - příklady

- Jakou minimální hodnotu IQ musí člověk mít, aby patřil mezi 5% osob s nejvyššími hodnotami IQ?
- p = 0.05

Normální rozdělení

- z tabulky: **z = 1.645**
- X = (1.645)*(15) + 100 = **124.675**
- musí mít IQ 124 bodů

Normální rozdělení - příklady

- pomocí tabulky normálního rozdělení je možno nalézt také hodnotu percentilu
- <u>příklad</u>: kolik procent osob má nižší hodnoty IQ než člověk s IQ 130?
- z= 2
- z tabulky: pro z = 2

p = 0.4772 (+ 50% pod průměrem)

97.72% osob má nižší skór než IQ 130

Rozdělení výběrových průměrů

- cílem induktivní statistiky je odhadnout parametry populace z charakteristik vzorku (výběrového souboru)
- např. odhadem průměru populace bude průměr vzorku
- odhad je vždy zatížen určitou **výběrovou chybou**

Rozdělení výběrových průměrů

- předpokládejme, že z jedné populace vybereme 3 různé vzorky
- budou se nejspíš navzájem lišit ve tvaru rozdělení hodnot, průměru i variabilitě
- jak se rozhodneme, který z nich zvolit pro odhad průměru populace ??

Rozdělení výběrových průměrů

- pokud bychom spočítali průměry ze všech možných výběrů o určité velikosti *n*, budou tvořit tzv. **rozdělení výběrových průměrů** (sampling distribution)

Rozdělení výběrových průměrů

- **příklad**: populace hodnot 2, 4, 6, 8
- průměr $\mu = 5$
- předpokládejme, že průměr neznáme a pokoušíme se ho odhadnout ze vzorku n=2
- v tabulce jsou uvedeny všechny možné výběrové soubory (v pořadí výběr, první skór, druhý skór, průměr vzorku)

Tabulka:

1	2	2	2
2	2	4	3
3	2	6	4
4	2	8	5
5	4	2	3
6	4	4	4
7	4	6	5
8	4	8	6
9	6	2	4

10	6	4	5
11	6	6	6
12	6	8	7
13	8	2	5
14	8	4	6
15	8	6	7
16	8	8	8

- jaká je pravděpodobnost, že z této populace vybereme vzorek s průměrem vyšším než 7?
- v rozdělení výběrových průměrů je takový vzorek jen 1 ze 16 – tj. pravděpodobnost takového vzorku je 1/16 = 0.0625, tj. 6%
- jaká je pravděpodobnost, že náhodně vybraný vzorek 2 čísel z této populace bude mít průměr roven průměru populace, tj. 5?
- tato pravděpodobnost je 4/16, tj. 25%

Rozdělení výběrových průměrů

- většina populací i vzorků je mnohem větší
- ale existují určité základní vlastnosti rozdělení výběrových průměrů (RVP)
- **tvar** – RVP se při dostatečně velkém vzorku (30 a více) blíží **normálnímu rozdělení**
- **průměr** tohoto rozdělení (=průměr průměrů všech teoretických výběrů) je roven **průměru populace**
- označuje se také jako očekávaná hodnota průměru vzorku
- **variabilita** – směrodatná odchylka RVP se označuje jako výběrová nebo standardní chyba průměru (standard error)
- jde o směrodatnou odchylku výběrových průměrů od průměru populace
- ukazuje, jak spolehlivý je odhad populačního průměru z průměru vzorku – tj. jak velkou chybou je odhad zatížen
- velikost výběrové chyby je dána dvěma charakteristikami: variabilitou v populaci a velikostí výběru
- **variabilita znaku v populaci**: čím je vyšší, tím je vyšší i

variabilita výběrových průměrů
- velikost výběru – čím větší výběr *(n),* tím méně průměrů výběrů se odchyluje od průměru populace (= výběrová chyba je menší)
- vzorec pro výpočet směrodatné chyby:

$$\sigma_x = \sigma/\sqrt{n}$$

tj. směrodatná odchylka populace děleno odmocninou z velikosti vzorku

Rozdělení výběrových průměrů
- platí zjednodušení **tzv. centrálního limitního teorému** – pro každou populaci o průměru μ a směrodatné odchylce σ se bude rozdělení výběrových průměrů výběrů (pro rozsah výběru jdoucí do nekonečna) blížit normálnímu rozdělení s průměrem μ a směrodatnou odchylkou $\sigma_x = \sigma/\sqrt{n}$

Rozdělení výběrových průměrů
- **příklad**: když vybereme z populace náhodně vzorek 9 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 112?
- ptáme se vlastně: jaká je pravděpodobnost, že vzorek 9 osob z populace o průměru 100 bude mít průměr 112 nebo vyšší? Kolik % je takových vzorků ze všech možných vzorků o této velikosti?
- musíme zjistit charakteristiku rozdělení výběrových průměrů pro tuto velikost vzorku (N=9) u populace s $\mu = 100$, $\sigma = 15$
- průměr RVP = 100
- směrodatná odchylka = standardní chyba: **$\sigma_x = \sigma/\sqrt{n} = 15/3 = 5$**
- známe průměr a směrodatnou odchylku rozdělení, převedeme tedy skór 112 na z-skór
- $\mu = 100$, $\sigma_x = 5$

- **z** = (112-100)/ σ_x = 12/5 = **2.4**
- pak najdeme v tabulce z-rozdělení pravděpodobnost pro z=2.4
- z tabulky $P(Z \geq 2.4)$ = **0.4918**
- odečteme od 50% (celá jedna strana z-rozdělení) a vyjde nám pravděpodobnost:
- p = 0.5000 – 0.4918 = **0.0082**

Kontrolní otázky

- výpočet a především interpretace z-skórů
- normální rozdělení – charakteristiky
- rozdělení výběrových průměrů
- výpočet směrodatné chyby

Literatura

- Hendl: kapitoly 4 a 5