Inferenční statistika

Populace vs. Vzorek
Distribuce výběrových průměrů
Interval spolehlivosti
Testování hypotéz

Populace vs. vzorek

	Populace	Vzorek
Průměr	μ	$\bar{x}$
Rozptyl (variance)	σ^2	s^2
Směrodatná(standardní) odchylka	σ	s

parametry

statistiky (odhady parametrů)

Skrze reprezentativní (=náhodný) vzorek odhadujeme parametry populace

Techniky odhadování

- Bodový odhad
  - jaký je nejlepší odhad charakteristiky populace?(např. $\bar{x}$ je bodovým odhadem μ)
- Intervalový odhad
  - jaký je interval který s vysokou pravděpodobností obsahuje charakteristiku populace?(např. μ leží s 99% pravděpodobností mezi 95 a 105) = interval spolehlivosti
- testování hypotéz
  - Za předpokladu že populační průměr 100, pravděpodobnost že vytáhnu vzorek s průměrem 110 a větším je dostatečně vysoká a proto neodmítám nulovou hypotézu že průměr populace je 100

Distribuce výběrových průměrů

= opakovaně vybírám vzorek a jeho průměry nanáším na novou distribuci

Vzniká nová distribuce s těmito charakteristikami:

A) průměr distribuce = průměr výběrových průměrů = populační průměr (zákon velkých čísel)

B) Odchylka = chybu průměru = $\sigma_{m(\bar{x})} = \sigma / \sqrt{n}$

C) čím vyšší počet vzorků, tím víc se distribuce blíže normální distribuci, bez ohledu na tvar populační distribuce (Central limit theorem)

ad C) distribuce se blíží normálnímu rozdělení když populační distribuce je normálně rozdělena nebo když velikost výběru je větší než 30

Příklad: distribuce výběrových průměrů

- Př. Výsledky IQ testu jsou aproximovány (blíží se) normálním rozložením o průměru $\mu = 100$ a $\sigma=16$. Třída 36 dětí dosáhne průměru 105 bodů, jak chytrá je tato třída?
- vypočítám chybu průměru = $\sigma_{m(\bar{x})} = \sigma \ / \sqrt{n} = 16 / \sqrt{36} = 2.67$
- A z-skór $Z = (\bar{x} - \mu) / \sigma_{m(\bar{x})} = (105 – 100) / 2.67 = 1.87$
- A příslušnou pravděpoodbost z tabulky pro $Z = 1.87$
- Výsledek: $P(\bar{x} > 105) = 0.03$ = třída je velmi chytrá a pravděpodobně patří do populace s $\mu > 100$

Příklad: testování hypotéz

- Př. Výsledky IQ testu jsou aproximovány (blíží se) normálním rozložením o průměru $\mu = 100$ a $\sigma=16$. Třída 36 dětí dosáhne průměru 105 bodů. Učitel si myslí že děti patří do populace $\mu > 100$
- H 0: $\mu <= 100$, H 1: $\mu > 100$
- vypočítám chybu průměru = $\sigma\ m(\bar{x}) = \sigma\ / \sqrt{n} = 16 / \sqrt{36} = 2.67$
- A z-skór $Z = (\bar{x} - \mu\ ) / \sigma\ m(\bar{x}) = (105 – 100) / 2.67 = 1.87$
- A příslušnou pravděpodobnost z tabulky pro $Z = 1.87$
- Výsledek: $P\ (\bar{x} > 105) = 0.03$,
- Interpretace: Za předpokladu že průměr populace je 100, tak pravděpodobnost že vytáhnu průměr o velikosti 105 nebo větší je 0.03 což je velmi nízká pravděpodobnost, proto nedůvěřuji H O a odmítám ji, klaním se k H 1.

Příklad: interval spolehlivosti

- Př. Výsledky IQ testu jsou aproximovány (blíží se) normálním rozložením o průměru $\mu = 100$ a $\sigma=16$. Třída 36 dětí dosáhne průměru 105 bodů. Vypočítejte 95 % interval spolehlivosti.
- Postup:
  - vypočítám chybu průměru
    $\sigma\ m(\bar{x}) = \sigma\ / \sqrt{n} = 16 / \sqrt{36} = 2.67$
  - Vypočítám horní hranici intervalu 95:
    $1.96\ \sigma_{(\bar{x})} + \bar{x} = (1.96 * 2.67) + 105 = 105 + 5.23 = 110.23$
  - Vypočítám spodní hranici intervalu 95:
    $\bar{x} - 1.96\ \sigma_{(\bar{x})} = 105 - (1.96 * 2.67) = 105 – 5.23 = 99.\ 77$

Interpretace: Když budu tahat vzorky nekonečně mnohokrát a pokaždé vytvořím 95 % interval, pak v 95% případů bude tento interval spolehlivosti obsahovat skutečný průměr

Alternativní interpretace: Existuje 95 % pravděpodobnost že skutečný průměr leží v intervalu 99.77 až 110.23

Analogicky vypočteme a interpretujeme 99 % interval spolehlivosti, namísto hodnoty 1.96 dosadíme hodnotu 2.58 (ověřte v tabulce z skoru)