

TÉMA 3

BODOVÉ A INTERVALOVÉ ODHADY

CO BYSTE MĚLI PO PROSTUDOVÁNÍ TOHOTO TÉMATU UMĚT?

1. Podstata, význam, výhody a nevýhody bodových a intervalových odhadů
2. Bodové odhady (význam, interpretace, možnosti použití pro data určitých vlastností)
3. Intervalové odhady (význam, interpretace, možnosti použití pro data určitých vlastností)
4. Praktický výpočet intervalových odhadů v Excelu
5. Praktický výpočet intervalových odhadů ve Statistice

OSNOVA

1. Teorie bodových a intervalových odhadů
2. Výpočet intervalu spolehlivosti střední hodnoty v Excelu
3. Výpočet intervalu spolehlivosti střední hodnoty ve Statistice
4. Výpočet intervalu spolehlivosti směrodatné odchylky v Excelu
5. Výpočet intervalu spolehlivosti směrodatné odchylky ve Statistice
6. Interpretace výsledků
7. Příklady na procvičení

TEORIE BODOVÝCH A INTERVALOVÝCH ODHADŮ

Odkaz na literaturu:

http://user.mendelu.cz/drapela/Statisticke_metody/teorie%20text%20I.pdf

Strany 90 - 101

Odkaz na prezentaci:

http://user.mendelu.cz/drapela/Statisticke_metody/Prezentace/zakladni/odhady.ppt

VÝPOČET INTERVALU SPOLEHLIVOSTI STŘEDNÍ HODNOTY V PROGRAMU EXCEL

Interval spolehlivosti střední hodnoty je symetrický kolem výběrového průměru. Stačí tedy vypočítat polovinu intervalu spolehlivosti (IS) a tuto hodnotu odečíst od průměru (tím získáme dolní hranici) a přičíst k průměru (tím získáme horní hranici).

V Excelu máme (kromě využití definičního vzorce) tři možnosti, jak vypočítat polovinu IS:

- V doplňku Analýza dat – Popisná statistika – zatrhneme volbu Hladina spolehlivosti pro střední hodnotu. Výsledná hodnota je polovina IS (vypočítána na základě t-rozdělení);
- Použití funkce CONFIDENCE.T – zadáme α (obvykle 0,05), Sm_odch (bodový odhad směrodatné odchylky) – např. z výstupu popisné statistiky v Excelu a Velikost (počet prvků výběru) – vzhledem k tomu, že tato funkce je založena na t-rozdělení, vyjde polovina IS naprosto stejně jako u předchozí metody;



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

- Použití funkce CONFIDENCE.NORM – zadání je stejné jako u předchozí funkce, ale tato funkce je založena na normovaném normálním rozdělení. Polovina IS vypočítané touto funkcí vychází pro stejná data menší než u funkce CONFIDENCE.T (t-rozdělení má větší variabilitu než normální rozdělení). Rozdíl výsledků obou funkcí bude tím větší, čím menší je výběr. Pro malé výběry (cca do 30 hodnot) je nutné používat CONFIDENCE.T, pro větší výběry (především řádově vyšší desítky nebo stovky a víc hodnot) je to prakticky jedno, protože rozdíly jsou zcela zanedbatelné.

VÝPOČET INTERVALU SPOLEHLIVOSTI STŘEDNÍ HODNOTY V PROGRAMU STATISTICA

Menu Statistika – Popisná statistika – na kartě Detailní výsledky se zatrhne kolonka s názvem Meze spolehlivosti průměru a interval se nechá nastaven na 95 % a ve výsledné tabulce se objeví dva řádky navíc – Interval spolehlivosti -95% a Interval spolehlivosti 95%, to je dolní a horní hranice intervalu spolehlivosti střední hodnoty. Výpočet také počítá s t- rozdělením, takže výsledky jsou totožné s výsledkem podle Popisné statistiky nebo funkce CONFIDENCE.T.

VÝPOČET INTERVALU SPOLEHLIVOSTI SMĚRODATNÉ ODCHYLKY V PROGRAMU EXCEL

Výpočet intervalu spolehlivosti směrodatné odchylky se v EXCELU provádí podle vzorce.

$$\sqrt{\frac{(n-1) \cdot \hat{S}^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}} \leq \sigma \leq \sqrt{\frac{(n-1) \cdot \hat{S}^2}{\chi_{\frac{\alpha}{2}, n-1}^2}}$$

Všimněte si, že levá a pravá část vzorce nejsou totožné a to z toho důvodu, že interval spolehlivosti směrodatné odchylky není souměrný, ale řídí se Pearsonovým (χ^2) rozdělením, které je nesouměrné. Čím menší je velikost výběru, tím větší je nesouměrnost. Levá část vzorce udává hodnotu dolní hranice intervalu spolehlivosti a pravá část hodnotu horní hranice intervalu spolehlivosti směrodatné odchylky. n je velikost výběru, S^2 je bodový odhad rozptylu a $\chi_{\frac{\alpha}{2}}^2; \chi_{1-\frac{\alpha}{2}}^2$ jsou hodnoty chi-kvadrát (Pearsonova) rozdělení pro α stupeň

významnosti pro $n-1$ stupňů volnosti. Velikost výběru a bodový odhad rozptylu spočítáte pomocí popisné statistiky EXCELU a hodnoty rozdělní chi-kvadrát se dají buďto vyhledat ve statistických tabulkách nebo se dají v EXCELU vypočítat pomocí funkce CHISQ.INV. I v tomto případě si ponecháme hladinu významnosti α nastavenou na hodnotu 0,05. Pro výpočet funkce CHISQ.INV musíte znát ještě počet stupňů volnosti (viz teorie), což je „velikost výběru – 1“. Funkci CHISQ.INV najdete mezi statistickými funkcemi.

Při zadávání argumentů funkce CHISQ.INV pojem „pravděpodobnost“ označuje danou hladinu významnosti ($\alpha/2 = 0,025$ pro horní hranici a $1-\alpha/2=0,975$ pro dolní hranici) a pojem „volnost“ počet stupňů volnosti $n-1$. Takto vypočítané hodnoty funkce CHISQ.INV již



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

můžete přímo dosadit do jmenovatele vzorce a vypočítat interval spolehlivosti směrodatné odchylky.

VÝPOČET INTERVALU SPOLEHLIVOSTI SMĚRODATNÉ ODCHYLKY V PROGRAMU STATISTICA

Menu Statistika – Popisná statistika – na kartě Detailní výsledky se zatrhne kolonka s názvem Meze spolehlivosti směrodatné odchylky a interval se nechá nastaven na 95 % a ve výsledné tabulce se objeví dva řádky navíc se stejným názvem – Spolehlivost směrodatné odchylky, kde nižší hodnota je dolní hranice a vyšší hodnota je horní hranice intervalu spolehlivosti směrodatné odchylky.

VZOROVÉ PŘÍKLADY - INTERPRETACE VÝSLEDKŮ

Odkaz na data k analyzování pro příklad 1 a 2:

http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Excel/odhadyEDA1.XLS

Příklad 1:

Vypočítejte pomocí funkce CONFIDENCE.NORM a pomocí popisné statistiky intervalový odhad (interval spolehlivosti) střední hodnoty základního souboru v souboru „pitná voda“.

Popisná statistika

<i>Obsah Fe iontů (mg/l)</i>	
Stř. hodnota	0,194658
Chyba stř. hodnoty	0,003342
Medián	0,1995
Modus	0,207
Směr. odchylka	0,023152
Rozptyl výběru	0,000536
Špičatost	0,170859
Šikmost	0,182056
#REF!	0,099
Minimum	0,149
Maximum	0,248
Součet	9,3436
Počet	48
Hladina spolehlivosti (95,0%)	0,006723

Bodový odhad střední hodnoty
 Bodový odhad směrodatné odchylky
 Bodový odhad rozptylu
 Velikost výběru
 Polovina int. spolehl. střední hodnoty

Interpretace:



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Intervalový odhad střední hodnoty základního souboru pomocí popisné statistiky je s pravděpodobností 95% $\langle 0,1879; 0,2014 \rangle$.

Funkce CONFIDENCE.NORM

Polovina interval spolehlivosti střední hodnoty vypočítaná pomocí funkce CONFIDENCE.NORM je 0,0065 (je menší oproti polovině intervalu spolehlivosti z popisné statistiky, protože je počítána z normovaného normálního rozdělení; popisná statistika používá Studentovo t rozdělení a to dává větší hodnotu poloviny IS z důvodu větší variability t-rozdělení). Intervalový odhad střední hodnoty základního souboru pomocí funkce CONFIDENCE.NORM je tedy s pravděpodobností 95% $\langle 0,1881; 0,2012 \rangle$.

Příklad 2:

Vypočítejte pomocí vzorce uvedeného výše intervalový odhad (interval spolehlivosti) směrodatné odchylky základního souboru v souboru „pitná voda“.

Interpretace:

Velikost výběru a rozptyl viz příklad 1. CHISQ.INV funkce pro $\alpha/2$ je 67,821 a pro $1-\alpha/2$ je 29,956. Po dosažení všech těchto hodnot dostaneme tyto výsledný interval. Intervalový odhad směrodatné odchylky základního souboru je tedy s pravděpodobností 95% $\langle 0,0193; 0,0290 \rangle$.

Odkaz na data k analyzování pro příklad 3 a 4:

http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Statistica/odhadyEDA1.sta

Příklad 3:

Vypočítejte pomocí popisné statistiky intervalový odhad (interval spolehlivosti) střední hodnoty základního souboru v souboru „letokruhy“.

Proměnná	1 letokruhy
N platných	38
Průměr	3,19736842
Int. spolehl.	2,8986983
Int. spolehl.	3,49603854
Medián	3,15
Modus	Vícenás.
Četnost	3
Minimum	1,9
Maximum	4,8
Rozpětí	2,9
Rozptyl	0,825668563
Sm.odch.	0,908663064
Spolehlivost Sm.Odch.	0,740800099
Spolehlivost Sm.Odch.	1,17558035
Koef.prom.	28,4190917
Šikmost	0,145339012
Špičatost	-1,22507716

Bodový odhad střední hodnoty

Dolní hranice IS střední hodnoty

Horní hranice IS střední hodnoty

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Interpretace:

Intervalový odhad střední hodnoty základního souboru pomocí popisné statistiky je s pravděpodobností 95% $\langle 2,898; 3,496 \rangle$.

Příklad 4:

Vypočítejte pomocí popisné statistiky intervalový odhad (interval spolehlivosti) směrodatné odchylky základního souboru v souboru „letokruhy“.

Proměnná	1 letokruhy
N platných	38
Průměr	3,19736842
Int. spolehl.	2,8986983
Int. spolehl.	3,49603854
Medián	3,15
Modus	Vícenás.
Četnost	3
Minimum	1,9
Maximum	4,8
Rozpětí	2,9
Rozptyl	0,825668563
Sm.odch.	0,908663064
Spolehlivost Sm.Odch.	0,740800099
Spolehlivost Sm.Odch.	1,17558035
Koef.prom.	28,4190917
Šikmost	0,145339012
Špičatost	-1,22507716

Bodový odhad směrodatné odchylky

Dolní hranice IS směrodatné odchylky

Horní hranice IS směrodatné odchylky

Interpretace:

Intervalový odhad směrodatné odchylky základního souboru je s pravděpodobností 95% $\langle 0,7408; 1,1755 \rangle$.

PŘÍKLADY NA PROCVIČENÍ

Odkaz na procvičování statistických charakteristik v Excelu:

http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Excel/odhadyEDA2.xls

Odkaz na procvičování statistických charakteristik ve Statistice:

http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Statistica/odhadyEDA2.sta

Na procvičení jsou vhodné soubory E 202 dusičnany a E 213 pH.

Pro Vaši kontrolu jsou zde uvedeny intervalové odhady střední hodnoty vypočítané pomocí popisné statistiky i pomocí funkce CONFIDENCE.NORM i intervalové odhady směrodatné odchylky.

Soubor E 202 - dusičnany

Intervalový odhad střední hodnoty základního souboru je s pravděpodobností 95% v rozmezí $\langle 0,4954; 0,5045 \rangle$ (vypočítaný pomocí funkce CONFIDENCE.NORM) a v rozmezí $\langle 0,4953; 0,5046 \rangle$ (vypočítaný pomocí popisné statistiky). Intervalový odhad směrodatné odchylky základního souboru je s pravděpodobností 95% v rozmezí $\langle 0,013; 0,020 \rangle$.

Soubor E 213 - pH

Intervalový odhad střední hodnoty základního souboru je s pravděpodobností 95% v rozmezí $\langle 6,665; 7,028 \rangle$ (vypočítaný pomocí funkce CONFIDENCE.NORM) a v rozmezí $\langle 6,657; 7,035 \rangle$ (vypočítaný pomocí popisné statistiky). Intervalový odhad směrodatné odchylky základního souboru je s pravděpodobností 95% v rozmezí $\langle 0,410; 0,693 \rangle$.