

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

TÉMA 1

ZÁKLADNÍ STATISTICKÉ CHARAKTERISTIKY

CO BYSTE MĚLI PO PROSTUDOVÁNÍ TOHOTO TÉMATU UMĚT?

1. Podstata, význam, výhody a nevýhody momentových a kvantilových charakteristik
2. Charakteristiky polohy (význam, hlavní zástupci včetně jejich interpretace, možnosti použití pro data určitých vlastností)
3. Charakteristiky variability (význam, hlavní zástupci včetně jejich interpretace, možnosti použití pro data určitých vlastností)
4. Charakteristiky tvaru (význam, hlavní zástupci včetně jejich interpretace, možnosti použití pro data určitých vlastností)
5. Vzorové příklady včetně interpretace v Excelu
6. Praktický výpočet těchto charakteristik ve Statistice

OSNOVA

1. Teorie statistických charakteristik
2. Výpočet v Excelu
3. Výpočet ve Statistice
4. Interpretace výsledků
5. Příklady na procvičení

TEORIE STATISTICKÝCH CHARAKTERISTIK

Odkaz na literaturu:

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

Strany 24 - 48

Odkaz na prezentaci:

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

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

VZOROVÉ PŘÍKLADY - INTERPRETACE VÝSLEDKŮ

Odkaz na data k analyzování:

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

Příklad 1:

Vypočítejte základní statistické charakteristiky pro „tloušťky“ ze souboru „odhadyEDA1“ a tyto charakteristiky interpretujte.

Výčetní tloušťka v cm	
Stř. hodnota	44,715
Chyba stř. hodnoty	0,759634
Medián	44,4
Modus	45,9
Směr. odchylka	7,596336
Rozptyl výběru	57,70432
Špičatost	-0,48092
Šikmost	0,091143
#REF!	34,6
Minimum	27
Maximum	61,6
Součet	4471,5
Počet	100

← Název charakteristiky, kterou analyzuji.
 ← Aritmetický průměr
 ← Medián
 ← Modus
 ← Směrodatná odchylka
 ← Rozptyl
 ← Koeficient špičatosti (zahrocenosti)
 ← Koeficient šikmosti (nesouměrnosti)
 ← Variační rozpětí
 ← Minimální hodnota studovaného souboru
 ← Maximální hodnota studovaného souboru
 ← Počet měření – velikost výběru

Charakteristiky polohy

Při srovnání aritmetického průměru a mediánu vidíme, že se jejich hodnoty vzájemně příliš neliší a to znamená, že v základním souboru se nebudou vyskytovat žádné extrémní hodnoty.

Charakteristiky variability

Pokud se podíváme na charakteristiky variability, tak si můžeme potvrdit, že směrodatná odchylka je skutečně odmocninou z rozptylu. Pokud vydělíme směrodatnou odchylku aritmetickým průměrem a výsledek vynásobíme hodnotou 100, tak dostaneme relativní míru variability „Variační koeficient“. Tímto koeficientem můžeme srovnávat variabilitu souborů i s různými jednotkami. Zde je hodnota variačního koeficientu 16,98 %.

Charakteristiky tvaru

Hodnota koeficientu špičatosti je menší než 0, to znamená, že data jsou plochá.

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Hodnota koeficient šikmosti je větší než 0, takže data jsou levostranná (v tomto případě jen velmi mírně), ale jedná se o hodnotu tak blízko 0, že taková data můžeme ještě označit za souměrná.

Shrnutí:

Můžeme říci, že data pocházejí ze základního souboru, který má souměrné a ploché rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Příklad 2:

Vypočítejte základní statistické charakteristiky pro „studny“ ze souboru odhadyEDA1 a tyto charakteristiky interpretujte.

<i>Obsah dusičnanů (mg/l)</i>	
Stř. hodnota	93,05051
Chyba stř. hodnoty	8,192695
Medián	80
Modus	80
Směr. odchylka	81,51629
Rozptyl výběru	6644,906
Špičatost	20,29007
Šikmost	3,869086
#REF!	597
Minimum	3
Maximum	600
Součet	9212
Počet	99

Popis charakteristik viz Příklad 1.

Charakteristiky polohy

Při srovnání aritmetického průměru a mediánu vidíme, že se jejich hodnoty vzájemně významně liší a to znamená, že v základním souboru se budou vyskytovat extrémní hodnoty. V dalších analýzách by se počítalo s hodnotou mediánu, protože ten není zkreslen extrémními hodnotami na rozdíl od aritmetického průměru.

Charakteristiky variability

Hodnota směrodatné odchylky je vzhledem k průměru dat značně vysoká, z čehož vyplývá vysoká hodnota variačního koeficientu 87,60 %. Tento datový soubor má tedy mnohem větší

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

variabilitu než soubor tloušťky - $87,60 > 16,98$, což je způsobeno převážně sešikmeným rozdělením (viz Charakteristiky tvaru dále) a extrémními hodnotami.

Charakteristiky tvaru

Hodnota koeficientu špičatosti je výrazně větší než 0, to znamená, že data jsou výrazně špičatá.

Hodnota koeficient šikmosti je výrazně větší než 0, takže data jsou silně levostranná.

Shrnutí:

Můžeme říci, že data pocházejí ze základního souboru, který má levostranné a špičaté rozdělení, vysokou míru variability a vyskytují se v něm extrémní hodnoty.

Příklad 3:

Vypočítejte základní statistické charakteristiky pro „letokruhy“ ze souboru odhadyEDA1 a tyto charakteristiky interpretujte.

<i>Průměrná šířka letokruhu (mm)</i>	
Stř. hodnota	3,197368
Chyba stř. hodnoty	0,147405
Medián	3,15
Modus	2,1
Směr. odchylka	0,908663
Rozptyl výběru	0,825669
Špičatost	-1,22508
Šikmost	0,145339
#REF!	2,9
Minimum	1,9
Maximum	4,8
Součet	121,5
Počet	38

Popis charakteristik viz Příklad 1.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Charakteristiky polohy

Při srovnání aritmetického průměru a mediánu vidíme, že se jejich hodnoty vzájemně příliš neliší a to znamená, že v základním souboru se nebudou vyskytovat žádné extrémní hodnoty.

Charakteristiky variability

Zde je hodnota variačního koeficientu 28,42 %. Tento datový soubor má tedy větší variabilitu než soubor tloušťky - $28,42 > 16,98$; ale zároveň má menší variabilitu než soubor studny.

Charakteristiky tvaru

Hodnota koeficientu špičatosti je menší než 0, to znamená, že data jsou plochá.

Hodnota koeficient šikmosti je větší než 0, to znamená, že data jsou levostranná, ale pouze mírně.

Shrnutí:

Můžeme říci, že data pocházejí ze základního souboru, který má levostranné a ploché rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

VÝSTUP V PROGRAMU STATISTICA:

Příklad 4:

Odkaz na data k analyzování pro příklad 4:

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

Abyste byli schopni interpretovat výsledky jak z Excelu, tak i z programu Statistica, tak je zde přidán ještě příklad 4. Jsou analyzována stejná data jako u příkladu 1, takže interpretace je stejná a nebude zde tedy opakována. Výstup dat z popisné statistiky poskytuje ty samé výsledky jako výstup z popisné statistiky v Excelu, ale má jiné zobrazení, které je níže popsáno. Velkou výhodou Statisticy je to, že si můžete navolit, které výstupy chcete (na kartě Details, podrobněji viz „návod“

http://user.mendelu.cz/drapela/Statisticke_metody/Navody_k_pouziti_statistickyh_programu/Statistica/Stat-chrakteristiky.exe).

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Proměnná	1 tloušťky
N platných	100
Průměr	44,715
Medián	44,4
Modus	Vícenás.
Četnost	2
Minimum	27
Maximum	61,6
Rozpětí	34,6
Rozptyl	57,7043181
Sm.odch.	7,59633584
Koef.prom.	16,9883391
Šikmost	0,0911433121
Špičatost	-0,480916981

Název charakteristiky, kterou analyzují.

Počet měření - velikost výběru.

Aritmetický průměr

Medián

Druh modu

Početnost modu

Minimální hodnota studovaného souboru

Maximální hodnota studovaného souboru

Variační rozpětí

Rozptyl

Směrodatná odchylka

Variační koeficient

Koeficient šikmosti (nesouměrnosti)

Koeficient špičatosti (zahrocenosti)

Pokud je modus unimodální, tak je v kolonce modus napsána jeho vlastní hodnota. Pokud je modus polymodální, tak je v kolonce modus napsáno vícenásobný. V kolonce četnost je udáno, kolikrát se v souboru hodnota modu vyskytuje. V tomto případě se jedná o polymodální modus a hodnoty všech modů jsou v souboru obsaženy dvakrát.

PŘÍKLADY NA PROCVIČENÍ

Odkaz na procvičování statistických charakteristik v Excelu:

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

Pro Vaši kontrolu je zde uvedena výstupní tabulka všech příkladů a jejich interpretace.

E202		E203		E213		E218		E221	
Stř. hodnota	0,5	Stř. hodnota	1,2555	Stř. hodnota	6,8467	Stř. hodnota	26,483	Stř. hodnota	5,4285
Chyba stř. hodno	0,0023	Chyba stř. hodno	0,0023	Chyba stř. hodno	0,0926	Chyba stř. hodno	0,1525	Chyba stř. hodno	0,1562
Medián	0,5	Medián	1,25	Medián	6,95	Medián	26,6	Medián	4,8
Modus	0,51	Modus	1,25	Modus	7,2	Modus	26,8	Modus	4,2
Směr. odchylka	0,0163	Směr. odchylka	0,0165	Směr. odchylka	0,507	Směr. odchylka	0,8355	Směr. odchylka	1,8748
Rozptyl výběru	0,0003	Rozptyl výběru	0,0003	Rozptyl výběru	0,2571	Rozptyl výběru	0,698	Rozptyl výběru	3,5148
Špičatost	-0,2105	Špičatost	-1,2439	Špičatost	-0,5049	Špičatost	-1,0176	Špičatost	3,9401
Šikmost	-0,2656	Šikmost	0,0762	Šikmost	-0,4271	Šikmost	-0,2068	Šikmost	1,6742
#REF!	0,07	#REF!	0,05	#REF!	1,9	#REF!	2,8	#REF!	11,4
Minimum	0,46	Minimum	1,23	Minimum	5,8	Minimum	25	Minimum	3,2
Maximum	0,53	Maximum	1,28	Maximum	7,7	Maximum	27,8	Maximum	14,6
Součet	25	Součet	64,03	Součet	205,4	Součet	794,5	Součet	781,7
Počet	50	Počet	51	Počet	30	Počet	30	Počet	144

Soubor E 202

Data pocházejí ze základního souboru, který má mírně ploché a mírně pravostranné rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor E 203

Data pocházejí ze základního souboru, který má ploché a velmi mírně levostranné (dalo by se ohodnotit i ještě jako souměrné) rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor E 213

Data pocházejí ze základního souboru, který má ploché a pravostranné rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor E 218

Data pocházejí ze základního souboru, který má ploché a mírně pravostranné rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor E 221

Data pocházejí ze základního souboru, který má špičaté a levostranné rozdělení a vyskytují se v něm extrémní hodnoty.

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odkaz na procvičování statistických charakteristik ve Statistice:

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

Pro Vaši kontrolu je zde uvedena výstupní tabulka všech příkladů a jejich interpretace.

Proměnná	1 dusičnany	2 fosfor	3 pH	4 MCPA	5 CHSK
N platných	50	51	30	30	144
Průměr	0,5	1,2554902	6,84666667	26,4833333	5,42847222
Medián	0,5	1,25	6,95	26,6	4,8
Modus	,5100000	1,250000	7,200000	Vícenás.	4,200000
Četnost	13	14	4	3	15
Minimum	0,46	1,23	5,8	25	3,2
Maximum	0,53	1,28	7,7	27,8	14,6
Rozpětí	0,07	0,05	1,9	2,8	11,4
Rozptyl	0,00026530612	0,00027325490	0,25705747	0,697988501	3,51477807
Sm.odch.	0,016288220	0,016530423	0,50700835	0,83545706	1,87477414
Koef.prom.	3,25764407	1,31665094	7,40518537	3,15465227	34,5359447
Šikmost	-0,26564738	0,076162086	-0,42711744	-0,20684567	1,67424208
Špičatost	-0,21051554	-1,24386096	-0,50494353	-1,0176356	3,94011274

Soubor Dusičnany

Data pocházejí ze základního souboru, který má ploché a pravostranné rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor Fosfor

Data pocházejí ze základního souboru, který má ploché a velmi mírně levostranné (dalo by se ohodnotit i ještě jako souměrné) rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor pH

Data pocházejí ze základního souboru, který má ploché a pravostranné rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor MCPA

Data pocházejí ze základního souboru, který má ploché a pravostranné rozdělení a nevyskytují se v něm žádné extrémní hodnoty.

Soubor CHSK



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Data pocházejí ze základního souboru, který má špičaté a levostranné rozdělení a vyskytují se v něm extrémní hodnoty.