

Excel vo vyučovaní štatistiky

Mária Jurečková, Iveta Molnárová

ABSTRACT: The aim of this paper is to illustrate possibilities how to use EXCEL in the elementary course of Statistics.

Na konferencii Matematická štatistika a numerická matematika na vysokých školách v Kálnici zaznelo, že ak má mať predmet štatistika zmysel, mal by mať aspoň 40 hodín. Žiaľ, v poslednom období je na vysokých školách technického smeru tendencia znižovať počet hodín vyčlenených pre matematiku, a preto je prirodzené, že tento trend negatívne postihol i štatistiku.

Na Vojenskej akadémii v Liptovskom Mikuláši sa predmet štatistika vyučuje v 4. semestri v rozsahu 20-28 hodín, v závislosti od študijného odboru. Nasleduje po téme pravdepodobnosť, v ktorej sa preberajú základné pojmy a vety z teórie pravdepodobnosti, náhodná premenná a niektoré dôležité rozdelenia náhodnej premennej v rozsahu 22-26 hodín.

Po zriadení počítačových učební na jednotlivých fakultách Vojenskej akadémie, vznikli priaznivé podmienky pre využitie počítačov vo vyučovaní štatistiky. Vytvorili sa tak možnosti pre zefektívnenie práce na cvičeniach, čo istým spôsobom kompenzuje zníženie počtu hodín na tento predmet. I napriek tomu, že existujú špeciálne softwarové systémy na štatistickú analýzu, rozhodli sme sa pre vyučovanie základného kurzu štatistiky použiť Excel. Je niekoľko dôvodov pre toto rozhodnutie:

- Excel je súčasťou programového balíka kancelárskych aplikácií MS Office, takže študentom bude tento program dostupný i po skončení štúdia

- procedúry ponúkané Excelom svojim rozsahom vyhovujú obsahu základného kurzu štatistiky
- študenti poznajú základy Excelu z predmetu výpočtová technika

Predmet je rozdelený na prednášky a cvičenia. V rámci prednášok zdôrazňujeme vysvetľovanie základných pojmov, poukazujeme na geometrickú interpretáciu a aplikácie niektorých štatistických metód. Chceme aby študenti empirickým údajom vedeli priradiť teoretický model, vedeli interpretovať výstupy z počítača, snažíme sa spojiť predmet s praktickým životom. V prednáškach sa sústreďujeme na štyri hlavné témy:

- popisná štatistika
- bodový a intervalový odhad parametrov základného súboru
- testovanie hypotéz
- lineárna regresia

Popisná štatistika. Na výpočet parametrov štatistického súboru vyberieme z ponuky *Analytických nástrojov* procedúru *Popisná štatistika*. Pomocou tohto nástroja vytvoríme tabuľku, ktorá udáva hodnoty základných parametrov, s ktorými sa študenti oboznámili na prednáškach. Na triedenie súboru do tried použijeme procedúru *Histogram*, ktorá je vhodná i na tvorbu grafov.

Bodový a intervalový odhad parametrov základného súboru. Na cvičeniach z tejto témy je vhodné opäť použiť procedúru *Popisná štatistika*. Výstupná tabuľka, okrem iných parametrov, udáva bodový odhad m strednej hodnoty a bodový odhad s štandardnej odchýlky základného súboru. Pri zadaní hladiny spoľahlivosti α je v okienku hladiny spoľah-

livosti už vypočítaná hodnota $t_{1-\alpha/2} \frac{s}{\sqrt{n}}$, pomocou ktorej ľahko určíme interval spoľahlivosti $(m - t_{1-\alpha/2}s, m + t_{1-\alpha/2}s)$ pre strednú hodnotu μ základného súboru pre zvolenú hladinu α . Pre výpočet intervalu spoľahlivosti pre rozptyl základného súboru môžeme v Exceli použiť štatistické funkcie, konkrétne funkciu CHIINV, pomocou ktorej získame potrebné tabuľkové hodnoty spomínanej funkcie.

Nasledujúca Tabuľka 1. je ukážkou výstupu procedúry *Popisná štatistika*.

Tabuľka 1 Popisná štatistika

bodové hodnotenie	
Stř. hodnota	9,394736842
Chyba stř. hodnoty	0,939374985
Medián	10
Modus	10
Směr. odchylka	4,094640628
Rozptyl výběru	16,76608187
Špičatost	0,049699405
Šikmost	0,076945312
Rozdíl max-min	16,5
Minimum	1,5
Maximum	18
Součet	178,5
Počet	19
Hladina spolehlivosti (95,0%)	1,973555136

Testovanie hypotéz. V téme, ktorá sa zaoberá testovaním hypotéz, cielene volíme tie typy testov, ktoré sú v ponuke Excelu. Pri voľbe *Analytických nástrojov* nájdeme pre testovanie zhody stredných hodnôt základných súborov nasledujúce testy:

- dvojvýberový *z-test*
- dvojvýberový *t-test* s rovnosťou rozptylov základných súborov
- dvojvýberový *t-test* s nerovnosťou rozptylov základných súborov
- párový *t-test* na zhodu stredných hodnôt základných súborov

Pri rozhodovaní o výbere vhodného testu študenti z prednášok vedia, že je potrebné zohľadniť nasledujúce kritériá:

- rozsah výberových súborov
- porovnateľnosť rozptylov základných súborov
- vzájomná závislosť, resp. nezávislosť základných súborov
- skúmaný znak musí mať v oboch základných súborov normálne rozdelenie

Výstupy uvedených testov udávajú hodnoty parametrov, ktoré charakterizujú súbory a údaje, potrebné pre vyhodnotenie daného testu. Tieto sú pre jednostranné testy označené symbolom (1) a pre obojstranné testy symbolom (2). Vypočítanú testovaciu štatistiku nájdeme vo výstupnej tabuľke označenú symbolom *z*, prípadne *t-test*, podľa toho, či na výpočet bolo použité Gaussovo resp. Studentovo rozdelenie. Podobne i vypočítané kritické hodnoty sa uvádzajú pod názvom *z krit.* resp. *t krit.* Vo výstupnej tabuľke

je symbolom $P(Z \leq z)$, resp. $P(T \leq t)$, označená p hodnota, ktorá udáva najnižšiu hladinu významnosti na zamietnutie nulovej hypotézy. Na základe tejto hodnoty študenti hravo interpretujú výsledok testu.

Ako ukážku uvedieme výstupnú tabuľku dvojvýberového z -testu na strednú hodnotu a dvojvýberového t -testu s rovnosťou rozptylov.

Tabuľka 2 Dvojvýberový z -test na strednú hodnotu

Dvouvýběrový z-test na střední hodnotu		
	131-134	136-139
Stř. hodnota	12,1875	16,86486
Znamý rozptyl	35	44
Pozorování	32	37
Hyp. rozdíl stř. hodnot	0	
z	-3,0956659	
$P(Z \leq z)$ (1)	0,00098193	
z krit (1)	1,644853	
$P(Z \leq z)$ (2)	0,00196385	
z krit (2)	1,95996108	

Tabuľka 3 Dvojvýberový t -test na strednú hodnotu s rovnosťou rozptylov

Dvouvýběrový t-test s rovností rozptylů		
	131	136
Stř. hodnota	11	17,81818
Rozptyl	39,44737	58,71364
Pozorování	20	11
Společný rozptyl	46,09091	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	29	
t stat	-2,67542	
$P(T \leq t)$ (1)	0,006071	
t krit (1)	1,699127	
$P(T \leq t)$ (2)	0,012142	
t krit (2)	2,045231	

Z ponuky *Analytických nástrojov* je pre nás zaujímavý i F-test na testovanie zhody rozptylov dvoch základných súborov, ktorý je potrebné použiť v prípade voľby dvojvýberového t -testu pre zhodu stredných hodnôt základných súborov.

Do základného kurzu je vhodné zaradiť aspoň jeden test na overenie predpokladu o normálnom rozdelení znaku v základnom súbore. V závislosti od časovej rezervy v prednáškach, volíme χ^2 - test dobrej zhody, prípadne i test Shapira-Wilka. V ponuke Excelu síce nenájdeme spomínané testy, ale študenti zvyčajne nemajú problém s vytváraním potrebných funkcií a operácií v Exceli.

Lineárna regresia. Po prednáškach o vyšetrovaní vzájomného lineárneho vzťahu dvoch znakov skúmaných na súbore je vhodné riešiť konkrétne príklady na cvičení v Exceli. Excel má v ponuke rôzne procedúry na riešenie týchto úloh. Korelačný koeficient pre dva znaky je možné určiť pomocou funkcie *CORREL*, prípadne výberom *Analytické nástroje/ Korelácia* vypočítať maticu korelačných koeficientov všetkých kombinácií viacerých sledovaných znakov. V ponuke *Analýza údajov* je procedúra *Regresia*. Výstupné tabuľky tejto procedúry obsahujú nielen koeficienty regresnej priamky, ale i ďalšie informácie, ktoré síce presahujú obsah základného kurzu, ale na niektoré z nich, ako sú napríklad intervaly spoľahlivosti pre koeficienty regresnej priamky na zvolenej hladine významnosti, je vhodné študentov upozorniť.

Na základe skúseností, ktoré sme získali pri vyučovaní štatistiky s využitím Excelu, môžeme konštatovať, že tento predmet sa stal atraktívnejším pre študentov. Počítačové spracovanie údajov nielen spríjemňuje prácu, ale umožňuje i efektívnejšie využiť cvičenia. Ťažisko cvičení už nie je v numerickom spracovávaní údajov, ale presúva sa na výber vhodných štatistických metód a na správnu interpretáciu výstupov z počítača.

Literatúra:

1. Jurečková, M., Jurečka, S.: Uplatnenie štatistických metód v laboratórnych cvičeniach, Zborník konferencie Matematická štatistika a numerická matematika a ich aplikácie, Kočovce, 14-19.6.1999, s. 119-122
2. Jurečková, M., Jurečka, S.: Excel v štatistickom testovaní hypotéz, Zborník vojenskej akadémie 3 (2001), s. 27-35
3. Šťastný, Z.: Matematické a statistické výpočty v Microsoft EXCELu, Computer press, 1999

RNDr. Mária Jurečková, CSc.

Katedra matematiky VA
031 19 Liptovský Mikuláš
e-mail: jureckova@valm.sk

RNDr. Iveta Molnárová
Katedra matematiky VA
031 19 Liptovský Mikuláš
e-mail: imolnar@valm.sk