Využitie programu MS Excel vo výučbe štatistiky

Stanislav Lukáč

ABSTRACT: The software Microsoft Excel provides a lot of basic commands and functions for statistical data processing and graphic interpretation of relations among data in charts. These resources can be conveniently exploited in a statistical course at the secondary schools. The article is focused on solving of basic types of statistical problems with exploitation of MS Excel.

1. ÚVOD

Základné pojmy popisnej štatistiky a základné charakteristiky polohy sa zavádzajú v školskej matematike už v 9. ročníku na základnej škole. Okrem výpočtov štatistických charakteristík sa žiaci učia vyberať informácie z číselných tabuliek a prehľadne ich zoskupovať do frekvenčných tabuliek, ktoré možno vhodne doplniť názornými grafmi. Na tieto poznatky nadväzuje základný kurz štatistiky na strednej škole, ktorý zavádza presné definície už známych štatistických veličín a rozvíja vzdelávací obsah o základné charakteristiky rozptýlenosti a popis štatistických závislostí dvojrozmerných súborov.

Teoretické vedomosti využívajú študenti pri riešení úloh, ktoré v mnohých prípadoch obsahujú údaje z reálneho života usporiadané vo forme tabuliek. Riešenie týchto úloh vyžaduje často časovo náročnejšie výpočty, na realizáciu ktorých sa väčšinou využívajú kalkulačky. Aj keď väčšina kalkulačiek disponuje štatistickými funkciami, výkonnejším prostriedkom pre spracovanie štatistických súborov je program MS Excel. Poskytuje možnosť zadania vstupných tabuliek na hárky pracovného zošita a jeho hlavnými výhodami sú komplexné nástroje pre štatistické aj grafické spracovanie vstupných tabuliek a ich prepojenie so získanými výsledkami aj grafmi, čo umožňuje okamžitú aktualizáciu výsledkov po ľubovoľnej zmene vstupných údajov. Študenti môžu vypočítať základné štatistické charakteristiky na základe postupov predpísaných definíciami, alebo môžu využiť príkazy programu a implementované štatistické funkcie. Uvedené prednosti programu MS Excel možno vhodne využiť pre motiváciu študentov a pre zvýšenie efektívnosti výučby. V úvodných etapách výučby zameraných na osvojenie a pochopenie definícií ponúka tento program alternatívnu možnosť pre kontrolu žiackych výsledkov a pre odstraňovanie chýb vyplývajúcich z nepozornosti.

2. ZÁKLADNÉ CHARAKTERISTIKY POPISNEJ ŠTATISTIKY

Výpočet základných štatistických charakteristík sa v programe MS Excel realizuje najčastejšie pomocou zabudovaných štatistických funkcií. Vybrané možnosti tohto programu pre jednoduché štatistické výpočty predstavíme v nasledujúcom príklade.

Príklad č. 1: Skupine 15 študentov bol zadaný test obsahujúci 5 úloh, pričom za správne vyriešenie každej úlohy bolo možné získať najviac 4 body. Doplňte do tabuľky, v ktorej sú zapísané hodnotenia študentov za jednotlivé úlohy, výpočet celkovej úspešnosti študentov a vypočítajte základné štatistické charakteristiky polohy a rozptýlenosti získaného súboru.

Na zaznamenanie bodových hodnotení a realizáciu výpočtov vytvoríme tabuľku podľa nižšie uvedenej predlohy. Najjednoduchšie získame celkové súčty pomocou automatického súčtu aplikovaného na oblasť buniek G2:G16. Tabuľka je usporiadaná vzostupne podľa úspešnosti študentov vyjadrenej v percentách. Základné štatistické charakteristiky: aritmetický priemer, modus a medián sme vypočítali pomocou štatistických funkcií: AVERAGE, MODE a MEDIAN. Argumentom týchto funkcií je oblasť buniek s úspešnosťami (H2:H16). Na výpočet smerodajnej odchýlky sme využili pomocný stĺpec Odch. Do bunky I2 bol zapísaný vzorec podľa obrázka, ktorý bol následne skopírovaný do oblasti I3:I16. Smerodajnú odchýlku sme vypočítali v bunke I20 pomocou vzorca: =SQRT(1/15*SUM(I2:I16)). Na kontrolu výpočtu sme zapísali do bunky H20 vzorec: =STDEVP(H2:H16), ktorý umožňuje vypočítať smerodajnú odchýlku súboru. Hodnoty základných štatistických priamo charakteristík možno získať aj jednoduchším spôsobom pomocou analytického nástroja Descriptive Statistics, ktorý je prístupný v položke Nástroje v hlavnej ponuke programu. Po vyvolaní príkazu zadáme adresu oblasti obsahujúcej vstupný súbor a adresu oblasti pre zápis výsledkov. Aplikovanie tohto nástroja na súbor úspešností zobrazuje pravá tabuľka so záhlavím Úspešnosť v %. Uvedená hodnota smerodajnej odchýlky je odlišná oproti predchádzajúcemu výpočtu, lebo je určená ako odhad smerodajnej odchýlky základného súboru na základe realizovaného výberu (vo vzorci je v menovateli výraz N - 1).

	12 = = (H2-\$H\$17)^2											
	А	В	С	D	Е	F	G	Н	·····	J		····Ľ····
1	Meno	Ul1	UI2	UI3	UI4	UI5	Spolu	Úsp (%)	Odch		Úspešnosť v 9	%
2	A.C.	1	1	2	2	0	6	30	1708,44			
3	S.L.	2	0	3	1	2	8	40	981,78		Mean	71,33
4	R.M.	0	3	2	1	3	9	45	693,44		Standard Error	5,531
5	J.S.	2	0	2	1	4	9	45	693,44		Median	80
6	P.F.	4	1	3	2	4	14	70	1,78		Mode	70
7	I.B.	3	2	4	3	2	14	70	1,78		Standard Deviation	21,42
8	S.K.	4	2	1	4	3	14	70	1,78		Sample Variance	458,8
9	A.P.	3	4	4	3	2	16	80	75,11		Kurtosis	-0,74
10	E.Č.	3	3	4	3	3	16	80	75,11		Skewness	-0,77
11	M.B.	3	4	3	4	3	17	85	186,78		Range	65
12	A.Ď.	3	4	4	2	4	17	85	186,78		Minimum	30
13	M.S.	4	3	3	4	4	18	90	348,44		Maximum	95
14	B.T.	4	4	3	3	4	18	90	348,44		Sum	1070
15	J.K.	4	4	3	4	4	19	95	560,11		Count	15
16	R.G.	4	3	4	4	4	19	95	560,11			
17	Priemer	2,93	2,53	3,00	2,73	3,07	14,27	71,33				
18	Modus	4	4	3	4	4	14	70				
19	Medián	3	3	3	3	3	16	80				
20	0 Smerodajná odchýlka							20,69	20,69			

Na grafickú reprezentáciu údajov z tabuliek sa využívajú grafy. Program MS Excel ponúka širokú paletu rôznych typov grafov, na zostrojovanie ktorých slúži interaktívny sprievodca grafmi. V ¾ avej èasti nasledujúceho obrázka je zostrojený jednoduchý kruhový diagram zobrazujúci podiel úspešností v jednotlivých úlohách na celkovom hodnotení vybraného

študenta. Pre zostrojenie histogramu absolútnych početností úspešností bolo potrebné vytvoriť frekvenčnú tabuľku. Horné hranice tried početností sú definované v stĺpci Triedy. Absolútne početnosti celkových úspešností patriacich do jednotlivých tried boli určené pomocou štatistickej funkcie FREQUENCY. Táto funkcia sa vkladá do tabuliek ako pole vzorcov a preto musí byť aplikovaná na oblasť buniek (P2:P6). Na základe získanej tabuľky je zostrojený histogram v pravej časti obrázka.



Pri väčšom rozsahu súborov môžu byť vstupné údaje zadávané vo forme tabuliek obsahujúcich súhrnné údaje vyjadrujúce početnosť jednotlivých hodnôt alebo početnosť hodnôt v zadaných intervaloch. V týchto prípadoch je potrebné pri výpočte aritmetického priemeru zohľadniť absolútne početnosti jednotlivých hodnôt.

Príklad č. 2: V zadanej tabuľke obsahujúcej dva stĺpce sú uvedené súhrnné údaje o dĺžke trvania cesty do školy u žiakov 1. ročníka. Vypočítajte priemernú dobu dochádzky do školy žiakov 1. ročníka.

iliaid.	A	В	С
1	Čas	Počet	Výpočet
2	3	11	=SUM(A1:A2)/2*B2
3	6	22	99
4	9	30	225
5	12	15	157,5
6	15	10	135
7	20	6	105
8	spolu	94	721,5
9	priemer	7,68	min.

V 1. stĺpci tabuľky je uvedená horná hranica intervalu vyjadrujúceho dĺžku trvania cesty do školy v minútach ((0 - 3>, (3 - 6>, atď.). Pre výpočet aritmetického priemeru vypočítame stred ktorý jednotlivých intervalov, vynásobíme počtom žiakov. Uvedené výpočty zapíšeme do pomocného stĺpca Výpočet. Vzorec v bunke C2 je zapísaný tak, aby sme ho bez zmien mohli skopírovať do oblasti C3:C7. V bunkách B8 a C8 určíme stĺpcové súčty a priemernú dobu

dochádzky vypočítame ako podiel čísel v bunkách C8 a B8.

Úlohy:

1. Na základe informácií poskytnutých Štatistickým úradom Slovenskej republiky za prvý štvrťrok 2003 sme vybrali priemerné nominálne mzdy a počty zamestnancov v niektorých odvetviach hospodárstva za mesiac január. Vypočítajte priemernú nominálnu mzdu pracovníka v rámci uvedených odvetví.

	Priemysel	Výroba elektr.	Stavebníctvo	Veľkoobchod	Doprava	Pošty	Telekom.
Mzda	13923	18312	11104	13019	13344	10796	24394
Počet	490233	46216	120646	109983	93180	18779	11754

2. V nižšie uvedenej tabuľke sú na základe údajov poskytnutých Štatistickým úradom Slovenskej republiky uvedené vybrané údaje o obyvateľoch Slovenska. Vypočítajte

	1995	1996	1997	1998	1999	2000
Počet sobášov	27489	27484	27955	27494	27340	25903
Počet živo narodených detí	61427	60123	59111	57582	56223	55151
Počet živo nar. detí na 1000 obyv.	11,4	11,2	11,0	10,7	10,4	10,21

priemerný počet sobášov pripadajúcich na 1000 obyvateľov Slovenska v jednotlivých rokoch aj za celé sledované obdobie.

3. GENEROVANIE NÁHODNÝCH ČÍSEL

Program MS Excel poskytuje príkazy pre generovanie náhodných čísel z danej množiny na základe zadaného rozdelenia pravdepodobnosti výberu čísel. Štandardným prostriedkom pre náhodný výber čísla z intervalu <0,1) na základe rovnomerného rozdelenia pravdepodobnosti je matematická funkcia RAND. Hodnota tejto funkcie sa automaticky mení po každom prepočítaní vzorcov v pracovnom zošite. Komplexnejšie služby ponúka Generátor náhodných čísel nachádzajúci sa medzi nástrojmi na analýzu údajov v položke Nástroje z hlavnej ponuky. V dialógovom okne generátora možno nastaviť požadovaný počet náhodných čísel a ďalšie vlastnosti v závislosti od vybraného typu rozdelenia pravdepodobnosti. Do zvolenej výstupnej oblasti tabuľky sa vložia vygenerované čísla ako konštanty. Využitie funkcie RAND ukážeme v nasledujúcom príklade.

Príklad č. 3: Už Galileo Galilei zdôvodnil, prečo padá pri hode tromi hracími kockami častejšie súčet 10 ako súčet 9. Modelujte 500 hodov troch hracích kociek a určte početnosti jednotlivých súčtov.

	Α	B	С	D	E	F	G	Н
1	K1	K2	K3		Zmrazenie			
2					ZK1	ZK2	ZK3	Súčet
3	5	3	6		5	3	6	
:	:	:	:		:	:	:	
502	1	6	2		1	6	2	

Vytvoríme tabuľku podľa predlohy. Pre náhodný výber čísla z množiny {1,2,3,4,5,6} zapíšeme do bunky A3 vzorec: =INT(6*RAND())+1. Uvedený vzorec skopírujeme do oblasti A3:C502. Označíme oblasť A3:C502 a skopírujeme ju do schránky. Pomocou príkazu Prilepiť špeciálne v položke Úpravy z hlavnej

ponuky skopírujeme údaje zo schránky ako hodnoty do oblasti E3:G502. Týmto postupom sme "zmrazili" hodnoty funkcie RAND v skopírovanej oblasti. Ak by sme chceli vyhodnocova• ï alších 500 hodov, zopakovali by sme uvedenú postupnos• akcií. Do bunky H3 zapíšeme vzorec pre súèet bodov: =SUM(E3:G3) a skopírujeme ho do oblasti H4:H502. Potom vzostupne usporiadame tabu¾ku v oblasti E2:H502 pod¾a hodnôt v ståpci Súèet pomocou príkazu Zoradi• v položke Údaje z hlavnej ponuky.

Na usporiadanú tabu³/4ku aplikujeme Medzisúèty nachádzajúci príkaz sa v položke Údaje z hlavnej ponuky. Pre každú zmenu v ståpci Súèet použijeme funkciu Poèet hodnôt na urèenie poèetnosti jednotlivých súètov. V 3/4 avej èasti pracovnej plochy sa po potvrdení príkazu objavia ikony pre zobrazovanie úrovní rôznych preh3/4adu údajov. Zatlaèíme ikonu pre úroveò 2 na zobrazenie poèetností súètov. Vytvorenú preh¾adovú tabu¾ku využijeme pre



zostrojenie histogramu absolútnych početností. Prehľadová tabuľka spolu s histogramom je zobrazená na obrázku.

Do buniek tabuľky možno vkladať aj dátumy. V programe MS Excel je dátum vnútorne reprezentovaný ako prirodzené číslo, ktoré predstavuje poradie daného dňa určované vzhľadom na deň 1.1.1900. Pri zápise dátumov do buniek sa automaticky nastaví štandardný dátumový formát určujúci spôsob zobrazenia údajov v bunkách. Po zmene formátu bunky s dátumom na všeobecný alebo číselný možno v bunke zobraziť odpovedajúce poradové číslo. Takýmto spôsobom možno v bunkách náhodne generovať aj dátumy.

Príklad č. 4: V škole vyučuje 30 učiteľov vo veku od 26 do 60 rokov. Vygenerujte náhodne dátumy narodenia učiteľov a určte absolútne početnosti učiteľov podľa zvolených vekových skupín.

Na identifikáciu jednotlivých položiek dátumu slúžia dátumové funkcie: YEAR, MONTH a DAY. Ak do bunky A1 vložíme dnešný dátum pomocou funkcie TODAY a do bunky A2 dátum z minulosti, možno do bunky A3 zapísať na približné určenie počtu celých rokov, ktoré uplynuli medzi zadanými dátumami vzorec: =YEAR(A1)-YEAR(A2). Na presné určenie počtu celých rokov rešpektujúce celé zadané dátumy využijeme vzorec: =IF(MONTH(A2)> MONTH(A1);A3-1;IF(AND(MONTH(A2)=MONTH(A1);DAY(A2)>DAY(A1));A3-1;A3)). Na rozlíšenie jednotlivých prípadov vo výpočte boli využité logické funkcie IF a AND.

Vzhľadom na zložitý zápis presného vzorca budeme v tabuľke využívať jednoduchší vyššie uvedený vzorec. Podľa zadania úlohy sú dátumy narodenia učiteľov z obdobia od 1.1.1943 do 31.12.1977. Týmto dátumom odpovedajú poradové čísla 15707 a 28490. Na generovanie náhodného prirodzeného čísla z tohto intervalu možno využiť funkciu RANDBETWEEN alebo vzorec: =15707+INT(RAND()*12784). Vytvoríme tabuľku podľa predlohy. Do bunky

	Α	В	С	D
1	Dnes	=TODAY()		
2	Priezvisko	Náh. číslo	Dátum nar.	Vek
3	Barátová	20497	12.2.1956	47
:	:	:	:	:
32	Velesová	23893	31.5.1965	38

zapíšeme vyššie uvedený vzorec B3 a skopírujeme ho do oblasti B4:B32. Označíme oblasť s vygenerovanými číslami a skopírujeme ju do schránky. Potom pomocou príkazu Prilepiť špeciálne v položke Úpravy z hlavnej ponuky vložíme zo schránky hodnoty vzorcov do tretieho stĺpca tabuľky. Označíme

skopírované čísla v oblasti C3:C32 a zmeníme formát buniek na dátumový. Do bunky D3 zapíšeme vzorec: =YEAR(\$B\$1)-YEAR(C3), ktorý skopírujeme do celej tabuľky. Na záver rozdelíme zadaný interval pre vek učiteľov na päťročné intervaly a pomocou funkcie FREQUENCY určíme početnosť učiteľov v jednotlivých vekových skupinách.

Úlohy:

1. Podľa tabuľky z predchádzajúceho príkladu určte, ako často budú v jednotlivých mesiacoch oslavovať učitelia narodeniny svojho kolegu.

2. D'Alembert sa mylne domnieval, že pravdepodobnos• udalosti, že pri hode dvoma mincami padne aspoò raz znak, je 2/3. Simulujte v programe MS Excel 300 hodov dvoma mincami a urète relatívnu poèetnos• tejto udalosti.

4. ŠTATISTICKÁ ZÁVISLOSŤ ZNAKOV

Pre skúmanie vzťahu medzi dvoma znakmi štatistického súboru ponúka program MS Excel numerické aj grafické prostriedky, ktoré umožňujú charakterizovať závislosti medzi údajmi v tabuľkách a následne realizovať prognózy ďalšieho vývoja skúmaných znakov. Na kvantitatívne vyhodnocovanie závislostí slúžia najmä štatistické funkcie (napr. COVAR, CORREL a pod.). Pre grafickú interpretáciu vzťahu medzi dvoma veličinami sa využíva graf typu XY (závislosť), ktorý obsahuje dve číselné osi na nanášanie číselných údajových radov.

Príklad č. 5: V tabuľke sú uložené údaje o výrobe a hospodárení istej americkej automobilky v miliónoch USD. Preskúmajte vzťah medzi nákladmi na reklamu a ziskom automobilky v sledovanom období. Za predpokladu lineárnej závislosti medzi týmito údajmi odhadnite zisk automobilky v roku 1987. ak náklady reklamu boli na 40 miliónov USD.

	Α	В	С	D
1	Rok	Nákl. na výrobu	Nákl. na reklamu	Zisk
2	1981	12,8	2,5	28
3	1982	42	8,8	170
4	1983	89	12	700
5	1984	143	19	2384
6	1985	184	27	1885
7	1986	212	34	2197

Aj keď v skutočnosti závisí zisk od mnohých premenných, sústredíme sa len na



vyhodnocovanie parciálnej závislosti medzi nákladmi na reklamu a ziskom. Na charakteristiku vzťahu medzi znakmi slúži koeficient korelácie, ktorý podľa zadanej tabuľky vypočítame pomocou =CORREL(C2:C7;D2:D7). vzorca: Vypočítaná hodnota 0,87 poukazuje na dosť tesnú väzbu medzi skúmanými znakmi. Údaje z tretieho a štvrtého stĺpca tabuľky zobrazíme v grafe XY (závislosť) ako izolované body. V tomto zjednodušenom modeli za predpokladu lineárnej závislosti medzi skúmanými znakmi preložíme zobrazenými bodmi lineárnu trendovú čiaru pomocou kontextovej ponuky, ktorú vyvoláme kliknutím pravého tlačidla

myši na niektorý zo zobrazených bodov. Na základe zobrazenej priamky možno odhadnúť zisk automobilky v roku 1987 pre zadané náklady na reklamu. Odhad možno realizovať aj výpočtom pomocou vzorca: =FORECAST(40;D2:D7;C2:C7).

Úloha: Charakterizujte numericky aj graficky závislosť nákladov na reklamu od nákladov na výrobu podľa zadania z predchádzajúcej tabuľky.

5. ZÁVER

Pri preberaní tematického celku o spracovaní štatistických súborov v 3. ročníku gymnázia by už mali študenti ovládať základy práce s tabuľkovým kalkulátorom MS Excel z hodín informatiky. Z tohto dôvodu by na hodinách matematiky nebolo potrebné venovať veľa času vysvetľovaniu koncepcie počítačového spracovania údajov v tabuľkách a príkazov tohto programu, ale študenti by sa mohli sústrediť na využívanie štatistických metód založených na numerických aj grafických prostriedkoch. Využívaním matematického obsahu pri spracovaní údajov v tabuľkách by rozvíjali aj svoje informatické zručnosti, čo vedie k užitočnému a potrebnému upevňovaniu medzipredmetových vzťahov vo vzdelávaní.

Literatúra a použité zdroje:

- 1. Šedivý, J. a kol.: Matematika pre 3. ročník gymnázia. SPN Bratislava, 1986.
- 2. Šťastný, Z.: Matematické a statistické výpočty v MS Excelu. Computer Press, Brno, 1999.
- 3. Lukáč, S.: Multimédiá a počítačom podporované učenie sa v matematike. Prírodovedecká fakulta UPJŠ Košice, 2001.
- 4. www.statistics.sk

Adresa autora: S. Lukáč, Prírodovedecká fakulta UPJŠ Košice, slukac@kosice.upjs.sk