1 Triedenie podľa kvalitatívnych znakov

POJMY

- ✓ Kvalitatívny štatistický znak
- ✓ Dichotomický (alternatívny) znak
- Multinomický (polynomický) znak
- ✓ Štatistické triedenie
- ✓ Triediaci znak
- ✓ Jednostupňové triedenie
- ✓ Viacstupňové triedenie
- ✓ Asociačné triedenie
- ✓ Trieda
- ✓ Triedny symbol

- ✓ Triedna početnosť
- ✓ Podvojné triedenie
- ✓ Množné (multinomické) triedenie
- ✓ Asociačná tabuľka
- ✓ Kontingenčná tabuľka
- ✓ Triedna početnosť
- ✓ Absolútne početnosti
- ✓ Relatívne početnosti
- ✓ Kumulatívne početnosti absolútne
- ✓ Kumulatívne početnosti relatívne

RIEŠENÝ PRÍKLAD

Úloha: Získať z dostupných údajov čo najviac informácií o analyzovaných poľnohospodárskych podnikoch.

Zadanie1:

Roztrieď te poľnohospodárske podniky podľa výrobných podmienok a podľa právnej formy podnikania.

Riešenie:

Pri riešení vychádzame z dvoch kvalitatívnych znakov a to: podmienky a forma, ktoré môžeme vidieť aj tabuľke 1 prezentujúcej časť databázy, z ktorej vychádzame pri analýzach. Uvažujeme, že analyzované poľnohospodárske podniky pochádzajú z výrobných oblastí, ktoré členíme na horšiu a lepšiu výrobnú oblasť (kritériom pre dané členenie boli ukazovatele ako úrodnosť pôdy, % zornenia, atď.). Štatistický znak podmienky predstavuje teda alternatívny kvalitatívny znak. Druhým skúmaným znakom je právna forma podnikania. Aj napriek tomu, že zákon povoľuje rôzne formy podnikania (napr. družstvo, a.s., s.r.o., samostatne hospodáriaci roľník) v našom príklade budeme uvažovať len s dvomi kategóriami. Jednu kategóriu budú predstavovať družstvá a ostatné formy podnikania zlúčime do samostatnej kategórie, ktorá je v príklade označená ako ostatné. Aj znak forma nadobúda len dve obmeny, t.j. jedná sa opäť o alternatívny kvalitatívny znak.

Číslo podniku	Кгај	Okres	Podmienky	Forma	RPEPP2004	RPEPP2005
3610	Nitriansky	Komárno	lepšie	družstvo	121	113
4004	Žilinský	Námestovo	horšie	ostatné	52	50
3611	Nitriansky	Komárno	lepšie	ostatné	83	80
4005	Žilinský	Tvrdošín	horšie	ostatné	62	47
3612	Nitriansky	Komárno	lepšie	družstvo	40	37
3613	Nitriansky	Komárno	lepšie	družstvo	123	120
4006	Žilinský	Námestovo	horšie	družstvo	39	30
3614	Nitriansky	Komárno	lepšie	ostatné	295	272
3615	Nitriansky	Komárno	lepšie	ostatné	71	61
4007	Žilinský	Dolný Kubín	horšie	ostatné	41	37
3616	Nitriansky	Komárno	lepšie	družstvo	55	53
3617	Nitriansky	Komárno	lepšie	ostatné	204	185
4008	Žilinský	Dolný Kubín	horšie	ostatné	64	58
3618	Nitriansky	Komárno	lepšie	družstvo	368	364
4009	Žilinský	Námestovo	horšie	družstvo	23	20
3619	Nitriansky	Komárno	lepšie	ostatné	179	159
3621	Nitriansky	Komárno	lepšie	družstvo	44	43
4010	Žilinský	Námestovo	horšie	ostatné	57	60
3622	Nitriansky	Komárno	lepšie	družstvo	45	38
3624	Nitriansky	Komárno	lepšie	družstvo	35	30
4011	Žilinský	Dolný Kubín	horšie	ostatné	27	25
3625	Nitriansky	Komárno	lepšie	družstvo	4	7
4012	Žilinský	Dolný Kubín	horšie	ostatné	58	57

Tabuľka 1: Časť databázy poľnohospodárskych podnikov

Vysvetlivky: RPEPP 2004: ročný priemerný evidenčný počet pracovníkov za rok 2004 RPEPP 2005: ročný priemerný evidenčný počet pracovníkov za rok 2005 Našou úlohou je roztriediť štatistický súbor podľa dvoch alternatívnych kvalitatívnych znakov. Výstupom bude asociačná tabuľka, ktorá má rozmer 2x2.

Postup riešenia v Exceli:

Triedenie kvalitatívnych znakov sa v Exceli realizuje prostredníctvom voľby PivotTable and PivotChart Report, ktorú nájdeme v ponuke Data

Slide 1.1:

PivotTable and PivotChart \	Wizard - Step 1 of 3	<u>?×</u>
	Where is the data that you want to analyze? • Microsoft Office Excel list or database • External data source • Multiple consolidation ranges • Another PivotTable report or PivotChart report • Mutat kind of report do you want to create? • PivotTable • PivotTable • PivotTable	
	Cancel < Back <u>N</u> ext > <u>F</u> in	ish

Po výbere sa dostávame do procesu tvorby kontingenčnej tabuľky, ktorý pozostáva z troch krokov. V prvom kroku (Slide 1.1) definujeme zdroj údajov, ktoré budeme triediť a formu prezentácie výsledkov. V našom prípade budeme čerpať údaje z excelovskej databázy, t.j. akceptujeme predznačený výber (**Microsoft Office Excel list or database**). Tvoríme asociačnú tabuľku, t.j. potvrdíme voľbu Pivot Table. Do ďalšieho kroku sa dostaneme prostredníctvom tlačítka **Next**.

- V druhom kroku (Slide 1.2) sa definuje oblasť štatistického súboru. Ak triedime údaje v Exceli, musí byť štatistický súbor definovaný ako databáza, t.j. v prvom riadku sa musia nachádzať popisy jednotlivých polí, tzn. názvy štatistických znakov. Informácie o podniku sa musia byť definované v jednom riadku. Pokiaľ sa skôr ako vyberieme voľbu Pivot Table postavíme kurzorom do vnútra štatistického súboru na ľubovoľnú bunku, dokáže Excel vstupnú oblasť ohraničiť sám. Ak by oblasť nebola správne určená, v tomto kroku tvorby ju môžeme správne nadefinovať. Cez tlačítko Next prechádzame do posledného kroku tvorby tabuľky.
 - PivotTable and PivotChart Wizard Step 2 of 3

 Where is the data that you want to use?

 Range:
 \$A\$11\$G\$130

 Browse...

 Cancel
 < Back</td>
 Next >

 Einish

Slide 1.2:

Ak vytvárame ďalšie (nie prvú) kontingenčné tabuľky, objaví sa nám nasledovné informačné okno (slide 1.3).

Slide 1.3:

Microsoft	Excel
i)	Your new report will use less memory if you base it on your existing report PivotTable1, which was created from the same source data. Do you want your new report to be based on the same data as your existing report? • If you click Yes, you will save memory and your workbook file will be smaller. • If you click No, the two reports will be separate.
	<u>Yes</u> <u>N</u> o

Voľba **Yes** nám hovorí, že pri tvorbe novej tabuľky sa bude vychádzať už z existujúcej kontingenčnej tabuľky, t.j. výsledný xls súbor bude mať menší počet KB. V tomto prípade je potrebné v ďalšom okne určiť, z ktorej už vytvorenej tabuľky sa má vychádzať. Voľby **No** znamená, že budú vytvorené dva oddelené výstupy.

V treťom, poslednom kroku (slide 1.4) sa definuje ako má výstupná tabuľka vyzerať ako aj oblasť výstupu.



PivotTable and PivotChart Wizard - Step 3 of 3	? X
Where do you want to put the PivotTable report?	
Layout Options Cancel < Back Next >	inish

Vzhľad tabuľky sa definuje prostredníctvom tlačítka **Layout...** . Po jeho výbere sa dostávame do nasledovného okna (slide 1.5), v ktorom je v strede načrtnutú kontingenčná tabuľka a v pravej časti na nachádza zoznam štatistických znakov.

Slide 1.5:



Kontingenčná tabuľka sa skladá zo štyroch rôznych polí (**ROW**-riadok, **COLUMN**-stĺpec, **DATA**-údaje a **PAGE**-strana), do ktorých sa presúvajú štatistické znaky z pravej strany. Na ich základe sa potom realizuje triedenie.

V našom príklade triedime údaje podľa podmienok a formy, t.j. presunieme jeden znak (podmienky) do riadku a druhý znak (forma) do stĺpca. Máme teda zadefinované, na základe ktorých znakov sa má uskutočniť triedenie. Rovnako nevyhnuté je naplniť aj pole **DATA** (údaje), pretože na základe tohto štatistického znaku sa uskutoční vlastné triedenie. V prípade triedenia, t.j. zisťovania počtu štatistických jednotiek môžeme do poľa **DATA** premiestniť ľubovoľnú premennú. Musí byť však nastavená správna funkcia (t.j. funkcia **COUNT**) na zistenie počtu hodnôt v jednotlivých triedach¹.

V našom príklade ak presunieme do poľa DATA štatistický znak Číslo podniku, objaví sa v danom poli SUM of Číslo podniku, t.j. nebude zisťovaný počet, ale súčet hodnôt (funkcia SUM). Je preto potrebné, aby funkcia SUM bola zmenená, čo dosiahneme dvojklikom na položke v poli DATA. Dostávame nasledovné okno, v ktorom zmeníme funkciu SUM na funkciu COUNT a následne potvrdíme cez OK.

Slide 1.6:

¹ Pozn.: Ak robíme jednostupňové triedenie, jedno z polí **ROW**, resp. **COLUMN** môže ostať prázdne, pole **DATA** však musí byť vždy naplnené aspoň jedným štatistickým znakom.

PivotTable Fie	×	
Source field:	Číslo podniku	ок
Na <u>m</u> e: <mark>Sum o</mark>	f Číslo podniku	Cancel
<u>S</u> ummarize by: Sum		<u>H</u> ide
Count Average		<u>N</u> umber
Min Product		Options >>
Count Nums	•	

Správne nadefinovaná kontingenčná tabuľka vyzerá nasledovne:

Slide 1.7:



Poznámka: Pole **PAGE** je možné využiť, ak chceme triediť súbor podľa viacerých znakov. Pole **PAGE** (strana) nám umožní akoby stránkovať rôzne kontingenčné tabuľky podľa obmien znaku v poli **PAGE**.

Potvrdením OK sa dostávame späť do posledného okna tvorby tabuľky, v ktorom ešte určíme miesto výstupu. Štandardne je nastavený výstup do nového listu (New worksheet). V našom prípade nastavíme bunku v liste, v ktorom sa nachádza náš súbor. Stačí vyznačiť jednu bunku, ktorá predstavuje pravú hornú bunku výstupnej oblasti.

Slide 1.8:

PivotTable and PivotChart Wizard - Step 3 of 3	<u>?</u> ×
Where do you want to put the PivotTable report?	
C New worksheet	
Existing worksheet	
Click Finish to create your PivotTable report.	
Layout Options Cancel < Back Next > Ein	ish

Proces tvorby kontingenčnej tabuľky ukončíme prostredníctvom tlačítka Finish, po stlačení ktorého prebehne triedenie a výsledná kontingenčná (asociačná) tabuľka bude vložená do výstupnej oblasti, tak ako sme si ju nadefinovali v treťom kroku tvorby tabuľky.

Interpretácia výsledkov:

Výsledná asociačná tabuľka podľa zadefinovaných kritérií vyzerá nasledovne:

× 7 7 .	1 1
Vigtun	1 1 .
vystup	1.1.

Count of Číslo podniku	Forma 🔻		
Podmienky 🗾 👻	družstvo	ostatné	Grand Total
horšie	18	31	49
lepšie	44	36	80
Grand Total	62	67	129

Aké informácie môžeme získať z údajov v tabuľ ke?

V prvom rade je potrebné si uvedomiť, že výsledkom triedenia sú počty hodnôt v jednotlivých triedach, ktoré predstavujú absolútne početnosti. V pravom dolnom rohu sa nachádza početnosť nultého stupňa, čiže hodnota 129 nám hovorí, že v štatistickom súbore sme analyzovali 129 poľnohospodárskych podnikov. V poslednom riadku a poslednom stĺpci sa nachádzajú početnosti 1. stupňa predstavujúce sumy riadkov a sumy stĺpcov. Z tabuľky vidíme, že zo 129 poľnohospodárskych podnikov je 49 podnikov v horších výrobných podmienkach a 80 v lepších výrobných podmienkach. V súbore máme 62 družstiev a 67 podnikov majú inú formu hospodárenie. Vo vnútri v tabuľke sa nachádzajú početnosti 2. stupňa, t.j. pri triedení sme do úvahy brali obidva kvalitatívne znaky. Z údajov vidíme, že zo

62 družstiev je 18 v horších a 44 v lepších výrobných oblastiach. Z ostatných 67 poľnohospodárskych podnikov sa nachádza 31 v horších a 36 lepších výrobných podmienkach. Z asociačnej tabuľky vyplýva, že v našom súbore poľnohospodárskych podnikov majú takmer polovičné zastúpenie družstvá a väčšina podnikov hospodári v lepších výrobných oblastiach.

Zmena absolútneho vyjadrenia na relatívne vyjadrenie

Často krát je vhodnejšie miesto absolútneho vyjadrenia použiť relatívne vyjadrenie, t.j. vyjadrenie v %, pretože si vieme lepšie a rýchlejšie predstaviť, akú časť zaberá daná trieda z celku. Pri percentuálnom vyjadrení je potrebné si uvedomiť % z čoho chceme vyjadriť. Je rozdiel, či chceme vypočítať % z celkového počtu poľnohospodárskych podnikov alebo % z družstiev.

Postup v Exceli:

Vychádzame z existujúcej asociačnej tabuľky, ktorá je zobrazená v absolútnom vyjadrení. Zmenu na relatívne vyjadrenie urobíme nasledovne:

Nastavíme sa kurzorom do vnútra asociačnej tabuľky na ľubovoľnú číselnú hodnotu. Môžeme využiť panel nástrojov, ktorý sa automaticky objaví po vytvorení tabuľky. V takom prípade klikneme na ikonu Field Settings (druhá ikona od konca).



Druhý spôsob je prostredníctvom menu cez DATA/PivotTable and PivotChart Report. Po výbere z ponuky sa priamo presunieme do tretieho kroku tvorby kontingenčnej tabuľky, v ktorom opäť cez tlačítko Layout sa dostaneme na miesto, v ktorom sme definovali asociačnú tabuľku. Podobne ako pri zmene funkcie 2x klikneme na položku v poli DATA (COUNT od Číslo podniku) a dostávame sa do okna PivotTable Field. V spodnej časti okna zmeníme zobrazenie z Normal ²na percetuálne vyjadrenie z celkového počtu podnikov, t.j. z ponuky vyberieme voľbu % of total.

² Pozn.: Zobrazenie Normal znamená, že v kontingenčnej tabuľke sa budú počítať absolútne početnosti.

PivotTable Field	×
Source field: Číslo podniku	ОК
Name: Count of Cisio podniku	Cancel
Sum 🔺	Hide
Count Average Max	<u>N</u> umber
Min Product	Options >>
Show data as:	
% of total	-
Base field: Base item:	
Číslo podniku Kraj Okres Podmienky Forma RPFPP2004	*

Prepočítaná asociačná tabuľka bude vyzerať nasledovne:

Count of Číslo podniku	Forma 🔻		
Podmienky 🗾 🔻	družstvo	ostatné	Grand Total
horšie	13,95%	24,03%	37,98%
lepšie	34,11%	27,91%	62,02%
Grand Total	48,06%	51,94%	100,00%

Pri zobrazení % of total sa 100% (základ pre nás predstavujú všetky (129) poľnohospodárske podniky) nachádza v pravom dolnom roku kontingenčnej tabuľky. Hodnoty, ktoré sme v predchádzajúcej asociačnej tabuľke interpretovali sú teraz prepočítané na %, t.j. napr. v súbore máme 48,06% družstiev, z ktorých 13,95% hospodári v horšej a 34,11% v lepšej výrobnej oblasti. Podobné závery by sme mohli urobiť aj v prípade ostatných podnikov. V ich prípade vidíme, že zastúpenie v horších a lepších podmienkach je rovnomernejšie ako pri družstvách.

V situácii, keď nás nezaujíma celkový počet podnikov, ale len určitá časť, je vhodnejšie použiť zobrazenie % of row, resp. % of column.

Napr. zaujíma nás akú časť tvoria družstvá z počtu podnikov, ktoré hospodária v horších podmienkach. Keďže podmienky sa nachádzajú v poli **ROW**, je potrebné nastaviť zobrazenie % of row a dostaneme nasledovnú tabuľku:

PivotTable Field	×
Source field: Číslo podniku	ОК
Name: Count of Číslo podniku	Cancel
Summarize by:	Hide
Count Average	Number
Max Min Product	Options >>
Show data as:	
1% of row	
Base field: Base item: Cislo podniku Kraj Okres Podmienky Forma RPEPP2004	×

Count of Číslo podniku	Forma 🔻		
Podmienky 📃 👻	družstvo	ostatné	Grand Total
horšie	36,73%	63,27%	100,00%
lepšie	55,00%	45,00%	100,00%
Grand Total	48,06%	51,94%	100,00%

100 % sa v tabuľke nachádza v poslednom stĺpci, pretože na tomto mieste sa v prvej asociačnej tabuľke nachádzal celkový počet podnikov v horšej a lepšej výrobnej oblasti. Ako vyplýva z tabuľky, družstvá z celkového počtu podnikov v horšej výrobnej oblasti predstavujú 36,73%. Zvyšok, 63,27% tvoria ostatné podniky. V horších podmienkach sa nachádza teda viac podnikov z inou formou hospodárenia ako družstvá, na rozdiel od lepších podmienkach, v ktorých aj keď nepatrne prevláda zastúpenie družstiev.

Posledná situácia je taká, keď základom sú družstvá, resp. ostatné formy podnikov, z ktorých chceme vyčísliť % podnikov v lepších a horších podmienkach. Keďže sa forma nachádza v poli **COLUMN**, nastavíme spôsob zobrazenie **% of column**.

V tabuľke je možné vidieť, že v zastúpení v ostatných podnikov v horších a lepších podmienkach sú menšie rozdiely ako pri družstvách, v prípade ktorých sa z celkového počtu družstiev 29,03% nachádza v horších podmienkach a prevláda zastúpenie v lepších podmienkach (70,97%).

PivotTable Field	×
Source field: Číslo podniku	ОК
Name: Count of Číslo podniku	Cancel
Summarize by:	Hide
Count Average	<u>N</u> umber
Max Min Product	Options >>
Count Nums	
Show d <u>a</u> ta as:	
Base field: Base item:	
Kraj Okres Podmienky	
Forma RPEPP2004	v

Count of Číslo podniku	Forma 💌		
Podmienky 🗾 👻	družstvo	ostatné	Grand Total
horšie	29,03%	46,27%	37,98%
lepšie	70,97%	53,73%	62,02%
Grand Total	100,00%	100,00%	100,00%

Ukážka využitia poľa PAGE:

Do poľa **PAGE** sme presunuli *Formu* a do poľa **COLUMN** sme vložili nový znak *Kraj*. Kontingenčná tabuľka vyzerá nasledovne (nastavené je zobrazenie pre družstvá):

Forma	družstvo	▼				
Count of Číslo podniku	Kraj	•				
Podmienky 🗾 🔻	Košický		Nitriansky	Prešovský	Žilinský	Grand Total
horšie				7	11	18
lepšie		15	29			44
Grand Total		15	29	7	11	62

Keďže máme zobrazenú tabuľku za družstvá, vidíme, že v pravom dolnom rohu nemáme hodnotu 129, ale 62, čo predstavuje celkový počet družstiev v súbore. Z tohto počtu vidíme rozdelenie podľa podmienok a kraja. Družstvá v Prešovskom a Žilinskom kraji sa nachádzajú v horších podmienkach a družstvá v Košickom a Nitrianskom kraji hospodária v lepších podmienkach. Najväčšie zastúpenie (29) je družstiev z Nitrianskeho kraja. Podobným spôsobom by sme si mohli urobiť výpis aj za ostatné, poprípade všetky formy hospodárenie.³

³ Ak nastavíme formu na All (všetky formy hospodárenia), údaje budú triedené len podľa znakov, ktoré sa nachádzajú v poli ROW a COLUMN.