

Súbory údajov využívané vo výučbe štatistiky na EU v Bratislave

*Jozef Chajdiak**

Vo výučbe štatistiky sa využívajú súbory údajov v dvoch špecifických formách. Prvú predstavuje bezprostredný proces výučby formou prednášok a cvičení. Druhou formou je spracovanie samostatných zápočtových prác. V oboch formách ústrednou zásadou je pracovať s reálnymi súbormi údajov. V súčasnej dobe, keď prítomnosť počítačov a špecializovaného štatistického softvéru potláča aspekt výpočtovej náročnosti riešených úloh do úzadia, riešenie štatistických úloh na reálnych údajoch umožňuje precvičiť zvládnutie príslušných postupov ako na ľubovoľných iných súboroch údajov. Použitie reálnych údajov však dáva navyše študentovi vedomosti o konkrétnych vecných číselných hodnotách riešených úloh.

Špecifikácia súborov údajov v prvej forme presahuje rámec tohto príspevku. V rámci samostatnej práce v súčasnej dobe študenti analyzujú tri súbory údajov.

Prvý súbor predstavuje po 200 vygenerovaných čísiel so zadanými rozdeleniami. Úlohou vytvorenia a spracovania tohto súboru je zvládnuť prácu s počítačom a formálne zvládnuť rozličné štatistické procedúry.

Druhým súborom je zadaný časový rad. Na jeho získanie študenti musia pracovať s Historickou ročenkou a nasledujúcimi ročenkami. Aspekt práce s ročenkami je snáď dôležitejší ako následné riešenie zodpovedajúcich úloh analýzy a prognóz.

Tretí súbor predstavujú štvrt'ročné údaje z výkazu Priem P 3-04 za 1900 podnikov. Súbor je k dispozícii v pamäti počítača a úlohou je analýza týchto údajov a interpretácia vypočítaných výsledkov.

Pri veľmi heterogénnej hardvérovej základni s pomerne zastaralým základom je stále otvorenou otázkou analýza súboru údajov s niekoľko desiatok tisíc či stotisíc pozorovaní ako väčšieho stredného súboru a s miliónmi pozorovaní ako veľkého súboru.

Ukážky používaných zadanií pre jednotlivé úlohy nasledujú ďalej.

ROČENKA

Z údajov v Historickej štatistickej ročenke ČSSR, Štatistických ročeniek ČSSR, ČSFR a SR zistíte ročné údaje za Slovenskú republiku pre zadaný ukazovateľ od roku 1920 (prípadne 1945) do roku 1997.

Úlohy:

1. Vývoj ukazovateľa zobrazte graficky a špecifikujte prípadné etapy jeho vývoja.
2. Urobte jednoduchý štatistický rozbor časového radu ukazovateľa
 - min, max, priemer, std; vytriediť; histogram;
 - absolútny prírastok, tempo prírastku, priemerný absolútny prírastok (z regresnej priamky).
3. Vypočítajte a analyzujte autokorelačnú funkciu do 24. rádu
4. Urobte prognózu vývoja ukazovateľa do roku 2005.

* Katedra štatistiky FHI, Ekonomická univerzita, Dolnozemska 1, 85325 Bratislava,
e-mail: chajdiak@euba.sk

Príklad GENER

Úlohou je vygenerovať po 200 náhodných čísel z nasledovných rozdelení:

- normálne rozdelenie $N(\text{číslo}, \text{číslo})$
- rovnomerné rozdelenie $\langle 0, \text{číslo} \rangle$
- exponenciálne rozdelenie ($\lambda = \text{číslo}$)
- binomické rozdelenie ($n = 4, p = 0.2 + 0.001 * \text{číslo}$)

Číslo pre parametre rozdelení zadá cvičiaci každému študentovi na začiatku semestra.

Postupnosti náhodných čísel pre jednotlivé rozdelenia uložte do piatich premenných a riešte nasledovné úlohy:

1. Urobte základný štatistický rozbor pre jednotlivé premenné (minimum, maximum, priemer, úhrn, smerodajná odchylka; vytriediť; histogram).
2. Pre jednotlivé premenné zostrojte boxplot grafy.
3. Pre všetky možné dvojice premenných vypočítajte koeficienty korelácie a usporiadajte ich do korelačnej matice.
4. Pre dvojicu s najvyšším stupňom vzájomnej závislosti vypočítajte regresnú priamku. Výsledky zobrazte aj graficky.
5. Otestujte, či vygenerované súbory náhodných čísel majú zadané parametre.
6. Urobte analýzu rozptylu premennej s normálnym rozdelením a ako faktor použite premennú s binomickým rozdelením.

Príklad PODNIKY

Číselné hodnoty sa nachádzajú v súbore PODNIKY v adresári D:\SAS\UDAJE. Ide o súbor 1895 podnikov SR v jednotlivých štvrt'rokoch 1995 (písmená A, B, C a D na začiatku mena premennej špecifikujú 1., 2., 3. a 4. štvrt'rok). Od cvičiaceho na začiatku semestra dostane študent konkrétnu špecifikáciu vzťahu $p=Q/q$ (napríklad $w=VT/T$, kde w (p) je produktivita práce, VT (Q) výroba tovaru a T (q) počet pracovníkov). Symbolu 0 zodpovedá prvý polrok 1995 a 1 zodpovedá druhý polrok 1995.

Úlohy:

1. Špecifikujte názov a obsah premennej p a vypočítajte hodnoty p_0, p_1 a ďalej $Q=Q_1-Q_0, q=q_1-q_0, p=p_1-p_0$ a $Q_0=p_0*q_1$. Hodnoty všetkých premenných vypočítajte pre 1. a 2. polrok v porovnateľnej základni.
2. Indexná analýza:
 - a) Indexom charakterizovať vývoj úhrnu zloženej extenzitnej veličiny Q ; charakterizovať ako sa na tomto vývoji podieľa: - len zmena q ; - len zmena p ($I_H=I_q*I_p$)
 - b) Indexom charakterizovať vývoj priemeru intenzitnej veličiny p ; ako sa na tomto vývoji podieľa: len zmena q ; len zmena p ($i_pz=i_s*i_{sz}$)
 - c) Určiť celkovú absolútnu zmenu Q ; akou čiastkou sa na tejto zmene podieľa len zmena: priemeru intenzitnej veličiny p ; len zmenu úhrnu jednoduchej extenzitnej veličiny q
3. Štatistická analýza:
 - a) Urobte základný štatistický rozbor (minimum, maximum, priemer, smerodajná odchylka; vytriediť; histogram) z ukazovateľov Q a p (len za 1. štvrt'rok).
 - b) Vypočítajte korelačnú maticu medzi všetkými dvojicami ukazovateľov (z $Q_0, Q_1, q_0, q_1, p_0, p_1, Q, q$ a p) a analyzujte ju.
 - c) Závislosť Q_1 od q_1 modelujte regresnou priamkou. Výsledky zobrazte graficky.
 - d) Otestujte významnosť zmien $H_0: Q_0=Q_1, q_0=q_1$ a $p_0=p_1$.
 - e) Urobte analýzu rozptylu premených Q, q a p podľa faktora príslušnosti do formy, okeč.
 - f) Závislosť Q_1 od p_0, q_0, p a q modelujte regresnou funkciou.