

CURSUL XIV

PROGRAME APLICATIVE PENTRU PRELUCRAREA ȘI ANALIZA DATELOR STATISTICE

De mai mult de 30 de ani Minitab Inc. a oferit software și servicii pentru managementul calității, educație și cercetare. Mii de companii din mai mult de 80 de țări folosesc produse Minitab pentru îmbunătățirea calității și proiecte care folosesc metodele Six Sigma.

În domeniul educațional este folosit de mai mult de 4000 de colegii și universități.

Minitab Inc. a dezvoltat, dezvoltă și vinde software statistic pentru cei interesați de soluții pentru instruirea în analiza datelor și pentru analiza în sine a datelor, ajutându-i să obțină rezultate precise, de încredere și care să se remarce prin ușurința folosirii.

Produsul de bază al Minitab Inc. este **MINITAB**, ajuns acum la versiunea 14.1.

Totul a plecat de la dezvoltarea sa în 1972 în scopul de a ajuta profesorii în instruirea legată de statistici de bază, ajungând la cifra prezentată mai sus de universități și colegii care îl folosesc în întreaga lume. Referința la MINITAB se face în aproape 450 de lucrări de specialitate. Așadar, MINITAB a început ca fiind software-ul cel mai utilizat de profesori în procesul de instruire în statistică.

MINITAB Statistical Software a devenit răspândit și în lumea afacerilor datorită instrumentelor pe care le oferă în controlul calității, DOE (Design of Experiments), analizelor de fidelitate și statistică, în general. În momentul de față, MINITAB este instrumentul ales în multe țări ale lumii pentru afaceri de toate dimensiunile, de la cele mai mici până la companiile listate în Fortune 500, inclusiv Ford Motor Company, 3M, Honeywell International, General Motors și Six Sigma.

Minitab Inc. este o companie privată cu 165 de angajați în State College, biroul din Pennsylvania, cu 20 Minitab Ltd și subsidiare cu sediul în Coventry din Anglia și Paris din Franța, ca și cu o rețea bogată de reprezentanțe și distribuitori în China, Spania, Germania, Ungaria, Mexic, Brazilia, Argentina, Africa de Sud, Australia/Noua Zeelandă și Coreea.

Piața academică reflectă rădăcinile MINITAB, pentru că inițial a fost dezvoltat de trei membri ai Facultății de Statistică a Pennsylvania State University. În 1976 Duxbury Press a publicat un îndrumar pentru studenți (MINITAB Student Handbook) și a demonstrat profesorilor cum se poate integra utilizarea software-ului în cursurile introductive pentru statistică.

Combinarea între înalta calitate și ușurința folosirii a făcut ulterior din MINITAB software-ul statistic preferat pentru instruirea academică, o poziție pe care MINITAB și-o menține și astăzi.

Succesul MINITAB s-a arătat mai pregnant după decizia din 1982 de a se desprinde de universitate pentru a se dezvolta într-o companie privată, fondată în 1983 în campusul universitar al Pennsylvania State University.

Puterea sa pe piața academică a fost valorificată la fel de bine pe piețele de afaceri. Pe măsură ce studenții au ocupat poziții în industrie și în domeniul guvernamental au adus cu ei și cunoașterea MINITAB. Aceasta a luat un aspect practic în fabricile de autoturisme, în 12 din cele mai mari 15 companii din industria chimică a Statelor Unite, în 8 dintre cei mai mari 10 fabricanți de computere, în 10 din cele mai importante 15 companii din industria electronică și în 8 dintre cele mai mari 10 companii forestiere.

În prezent, se speră că s-a atins scopul de a fi software-ul statistic cel mai răspândit pe piața statisticii industriale. Produsul oferă în plus față de un pachet statistic general următoarele module: Statistical Process Control, Design of Experiments (DOE), Reliability/Survival Analysis și Measurement Systems Analysis. Suportul tehnic excelent și trainingul de calitate sunt părți integrale ale unei oferte valoroase aduse clienților de Minitab Inc.

Minitab oferă profesioniștilor angrenați în îmbunătățirea calității instrumente software puternice, de încredere și ușor de folosit de care au nevoie pentru a analiza efectiv datele în timp ce conduc proiecte ample și complexe.

MINITAB Statistical Software este pachetul ideal pentru proiectele Six Sigma și alte proiecte de optimizare a calității. De la controlul statistic al procesului până la proiectarea experimentelor, el oferă metodele de care este nevoie pentru a implementa fiecare fază a proiectului de calitate, împreună cu facilități ca StatGuide și ReportPad care ajută în înțelegerea și comunicarea rezultatelor. Este un pachet statistic precis, de mare încredere și ușor de folosit.

MINITAB 14.1 oferă multe facilități, cum ar fi: un nou motor grafic, care furnizează rezultate interesante și noi facilități relativ la prelucrarea datelor; metode facile de creare, editare și actualizare a graficelor; abilitatea de a personaliza meniurile.

Astfel, în domeniul managementului datelor și a fișierelor identificăm următoarele trăsături: dimensiunea nelimitată a worksheet-urilor, importă/exportă: Excel, text, și alte formate, interogare ODBC, autofill, manipularea datelor: unește, sortează și schimbă tipul datelor, salvează datele în format HTML, 4000 coloane pe worksheet.

În domeniul reprezentării grafice identificăm: metode facile de creare, editare și actualizare a graficelor, editarea interactivă a atributelor (axe, scala, etc.) și recrearea graficelor, grafice în dimensiunea 3D, plasarea mai multor grafice pe aceeași pagină, setarea implicită a unor atribute conform preferințelor utilizatorului, salvarea graficelor în format TIFF, JPEG, PNG, BMP, GIF etc.

La realizarea de statistici de bază, permite elaborarea de statistici descriptive, intervale de încredere, corelația și covarianța, testul de normalitate, testează egalitatea dispersiilor.

Se poate utiliza regresia lineară, regresia polinomială, regresia logistică, diagrama reziduală pentru analiza de corelație și regresie.

Pentru analiza dispersională se poate folosi: metoda ANOVA, metoda MNOVA, modelul linear general, comparații multiple și analiza mediei.

Serii de timp și tendințe: diagrame de serii de timp, auto-corelația, partial auto-corelația și corelația încrucișată, analiza Box-Jenkins (ARIMA), analiza trendului, netezirea exponențială, metoda Winter.

Poate funcționa pe sistemele de operare: Windows 98, 2000, Me*, XP, sau Windows NT 4 (Service Pack 4.0 și recente), cu următoarele cerințe de sistem: minim 32MB RAM, procesor Pentium I - 133 Mhz sau mai puternic, CD-ROM pentru instalare, monitor VGA sau SVGA, minim 800 X 600 - recomandabil 1024 X 768; 35MB spațiu pe HDD.

SPSS Inc. este unul din liderii mondiali în furnizarea de software și soluții de predicție analitică. Tehnologia companiei face legătura dintre date și acțiunile strategice efective prin tragerea de concluzii fiabile despre condițiile actuale și evenimentele critice viitoare. Peste 250.000 de clienți din sectorul comercial, academic și public se bazează pe tehnologia SPSS pentru a contribui la creșterea veniturilor, reducerea costurilor, îmbunătățirea proceselor importante și detecția și prevenirea fraudelor. Peste 95 de procente dintre companiile din lista Fortune 1000 sunt clienți SPSS.

SPSS (Statistical Package for the Social Sciences)¹ este unul dintre cele mai puternice și utilizate programe statistice. Acest pachet integrat asigură acoperirea procedeelelor specifice din Statistica descriptivă, Statistica inferențială și Analiza datelor. Programul a devenit

¹ Jaba, E., Grama, A. *Analiza statistică cu SPSS sub Windows*, Ed. Polirom, Iași, 2004. pp. 18-18

deosebit de atractiv pentru utilizatori deoarece permite tratarea datelor statistice fără a impune cunoașterea formulelor de calcul, îmbinând posibilitățile de prelucrare statistică cu facilitățile oferite de programele de calcul tabelar (Excel, Lotus, Quattro Pro) pentru condensarea datelor în tabele și reprezentarea lor grafică.

Programul este un produs al firmei SPSS Inc., care s-a impus în domeniul realizării de software pentru prelucrarea statistică a datelor, în principal prin SPSS și SYSTAT.

SPSS a fost creat la Universitatea din Stanford, în anul 1968, pentru a asigura gestiunea și analiza datelor statistice în domeniul științelor sociale și al psihologiei. Ulterior, utilizarea programului s-a extins spre economie, medicină etc. În același timp, evoluțiile din domeniul calculatoarelor au marcat și dezvoltarea SPSS, prin apariția imediată a unor noi versiuni.

Începând cu versiunea 7, realizată în 1995, SPSS a devenit un produs pentru Windows, ajungându-se astăzi la versiunea 12.

La ora actuală, SPSS este realizat sub formă modulară, fiecare utilizator putându-și achiziționa doar acele componente care îi sunt necesare. Cele mai „comercializate” module sunt: Base module, Professional Statistics, Advanced Statistics, Tables, Exact Tests, CHIAD și Categories.

Modulul de bază - **Base module** - permite gestionarea datelor și fișierelor, transformarea datelor, precum și prelucrarea statistică a acestora prin: calculul frecvențelor, al indicatorilor tendinței centrale, dispersiei și formeii unei distribuții; calculul măsurilor de asociere și testarea independenței probabilistice; pentru date incluse în tabelele de contingență; compararea mediilor, proporțiilor și dispersiilor eșantioanelor; analiza unifactorială a variantei; calculul coeficienților de corelație Pearson, Kendall, Spearman; analiza regresiei liniare; teste neparametrice.

Modulul **Professional Statistics** include proceduri pentru cercetarea relațiilor dintre variabile, folosind ca metode: analiza de discriminant; analiza factorială; analiza de clusteri; scalarea multidimensională; regresia ponderată; analiza fidelității.

Modulul **Advanced Statistics** permite efectuarea unor prelucrări statistice complicate, apelând la următoarele metode: analiza de regresie logistică; diverse extinderi ale analizei unifactoriale a variantei ANOVA; analiza variantei multifactorială MANOVA; analiza logliniară; analiza de regresie neliniară; analiza probit și logit; analiza duratei de viață; analiza de supraviețuire Kaplan-Meier; modelul liniar general.

Modulul **Tables** permite condensarea datelor în tabele cu una, două sau trei dimensiuni, fiecare dimensiune fiind definită printr-o variabilă sau printr-un grup de variabile. Pe lângă valorile variabilelor, tabelele pot conține frecvențe și valori ale unor indicatori statistici (medie, abatere standard etc.).

Modulul **Exact Tests** determină nivelul de semnificație (valorile p sau *Sig.*) pentru: teste neparametrice aplicate pe un eșantion, pe două eșantioane, independente sau perechi, și pe k eșantioane dependente sau independente; teste aplicate tabelor de contingență 2×2 și $r \times c$; teste de semnificație pentru coeficienții de corelație Pearson și Spearman; teste referitoare la relațiile dintre variabile măsurate pe scală nominală sau pe scală ordinală.

Modulul **CHIAD** (Chi-squared Automatic Interaction Detector) aplică algoritmi de segmentare pentru împărțirea unei populații în grupe disjuncte, care diferă între ele în funcție de un criteriu precizat. La fiecare pas al algoritmului, grupele constituite sunt vizualizate sub forma dendrogramelor.

Modulul **TRENDS** asigură analiza și reprezentarea grafică a seriilor de timp. Este posibilă estimarea coeficienților modelului de trend prin următoarele tehnici: procedee de ajustare; metode de regresie; analiza Box-Jenkins (ARIMA); procedee de descompunere sezonieră, pentru determinarea factorilor aditivi și multiplicativi, în cazul seriilor de timp cu caracter sezonier; analiza componentei aleatorii.

Caracteristicile principale ale SPSS pot fi sintetizate astfel:

• **Soluții pentru probleme complexe.** Având la dispoziție instrumente specifice metodelor statistice avansate, SPSS permite rezolvarea problemelor oricât de complexe, din diverse domenii, oferind soluții care să asigure o cunoaștere mai bună a fenomenelor cercetate și, implicit, să sprijine procesul de fundamentare a deciziilor.

• **Prezentarea sugestivă a rezultatelor.** Utilizatorul are control deplin asupra tuturor variabilelor prelucrate, stabilind modul de afișare a valorilor din listele de ieșire (lungime, număr de zecimale) și ce text să fie scris în locul denumirilor variabilelor (atunci când acestea nu sunt destul de sugestive) sau în locul valorilor variabilelor (dacă în fișierul de date s-au introdus coduri).

Listele de rezultate, tabelele și graficele realizate de SPSS pot fi incluse în rapoarte, așa cum se prezintă pe ecran, sau într-o formă modificată prin: editarea de texte; stabilirea caracteristicilor fonturilor/caracterelor (tip, stil, mărime, culoare); modificarea desenelor prin deplasarea și/sau rotirea axelor ori schimbarea tipului de grafic; ascunderea unor variabile din tabele; reorganizarea informațiilor în tabele (de exemplu, un tabel de frecvențe care conține pe linii răspunsurile la un chestionar, iar pe coloane localitatea de domiciliu și, în cadrul fiecărei localități, sexul clienților, poate fi transformat într-un tabel cu numai două coloane, corespunzător sexului, și cu grupe de linii, câte o grupă pentru fiecare localitate).

Toate aceste operații sunt ușor de executat, datorită existenței a trei editoare: de text, de tabele și de grafice. Rezultatele prelucrărilor statistice se pot vizualiza prin tabele de diverse formate și prin multiple tipuri de reprezentări grafice: histograme, diagrame „coloane” - izolate sau grupate -, diagrame de structură, nor de puncte - în care punctele corespunzătoare unor grupe distincte sunt colorate diferit -, diagrame „bare” - care indică în același timp media, valorile extreme și repartiția valorilor unei variabile pentru valori diferite ale altei variabile.

• **Suplețe în stabilirea condițiilor de prelucrare a datelor.** Domeniile diferite în care SPSS își găsește aplicare oferă o mare diversitate a condițiilor de prelucrare. SPSS permite realizarea oricărei variante de prelucrare, ori de câte ori este nevoie, la nivelul întregii baze de date sau la nivelul unui subansamblu de date selectat.

Dacă un grup de prelucrări se efectuează periodic, întreaga succesiune de căutări prin meniuri și de alegeri de opțiuni nu se repetă de fiecare dată, SPSS putând înregistra într-un fișier de comenzi toate aceste operații. Ulterior, fișierul va fi rulat ori de câte ori este necesar. În plus, fișierul poate fi actualizat, în sensul că i se pot adăuga sau șterge comenzi.

Utilizatorul poate alege cazurile care să fie luate în considerare la efectuarea prelucrărilor, formulând condiții asupra uneia sau mai multor variabile. De asemenea, utilizatorul poate decide modul în care să fie tratate de SPSS cazurile în care valoarea unei variabile nu este cunoscută sau nu prezintă interes pentru cercetare.

Înainte de efectuarea prelucrărilor statistice, SPSS poate modifica automat datele pe baza unor algoritmi indicați de utilizator pentru recodificarea valorilor sau prin aplicarea unor funcții matematice. De exemplu, într-un fișier în care cazurile sunt reprezentate de diverse mărfuri, iar variabilele de însușirile acestora, prețurile mărfurilor pot fi schimbate prin adăugarea TVA sau toate mărfurile produse înainte de 2000 pot primi aceeași valoare a variabilei „data de fabricație”, care să semnifice „înainte de 2000”.

• **Simplitate în exploatare.** SPSS este un program care poate fi exploatat și de persoane mai puțin inițiate în statistică, asigurând simplitate în manevrarea datelor de intrare. Introducerea și modificarea datelor este o operație simplă datorită existenței unui editor de tabele de tip spreadsheet. Pe ecran este afișat un tabel cu linii și coloane. Liniile corespund cazurilor (subiecți care răspund unui chestionar sau obiecte observate), iar coloanele conțin variabilele (răspunsuri date de subiecți sau rezultatele unor măsurători ori observații). Utilizatorul poate „naviga” prin acest tabel, după dorință, analizând valorile existente, schimbând unele date, adăugând sau ștergând cazuri și variabile. SPSS adaptează automat dimensiunile tabelului, astfel încât să nu se piardă nici o valoare introdusă.

SPSS asigură prelucrarea datelor preluate din registrele de lucru Excel, Lotus 1-2-3, bazele de date dBase sau fișierele de text ASCII. În același timp, fișierele create în SPSS pot fi exportate în Excel, Lotus 1-2-3 sau fișiere text.

Rezumând cele de mai sus, se poate aprecia că SPSS este un produs orientat spre utilizator și permite:

- analiza datelor sub multiple aspecte;
- extinderea datelor cercetărilor realizate pe un eșantion, la nivel național;
- construirea tabelelor de ieșire în diverse forme, inclusiv cu totaluri și structuri pe orizontală și verticală;
- construirea diagramelor sub diferite forme (linii, bare, sectoare etc);
- crearea prezentărilor și a rapoartelor;
- utilizarea datelor în regim interactiv și construirea seturilor de funcții ale sistemului pentru folosirea lor repetată (automatizarea analizei datelor);
- exploatarea facilităților oferite de Internet.

Microsoft Excel este un pachet din produsul Microsoft Office al firmei Microsoft Inc. Acesta furnizează o serie de funcții statistice, financiare și inginerești pentru analiza datelor din foile de lucru. Unele dintre funcțiile statistice sunt predefinite iar altele devin disponibile după instalarea pachetului Data Analysis. În acest scop MS Excel pune la dispoziție două module importante:

• **Modulul privind reprezentarea grafică a datelor.** Diagramele sunt apelate vizual și permit utilizatorilor să observe cu ușurință comparații, modele și tendințe în date. De exemplu, în loc să fie necesar să se analizeze câteva coloane de numere ale foii de lucru, se poate observa imediat dacă vânzările scad sau cresc pe perioade trimestriale sau cum se compară vânzările efective cu vânzările estimate.

Pentru aceasta, modulul pune la dispoziție metode facile de creare, editare și actualizare a graficelor, editarea interactivă a atributelor și recrearea graficelor, grafice în dimensiunea 3D, plasarea mai multor grafice pe același worksheet, setarea implicită a unor attribute conform preferințelor utilizatorului, salvarea graficelor în diferite formate.

• **Modulul Data Analysis.** Microsoft Excel furnizează un set de instrumente de analiză a datelor - denumit Data Analysis (Pachet de instrumente de analiză) - care se utilizează pentru a economisi etape atunci când se dezvoltă analize complexe, statistice sau de inginerie. Se furnizează datele și parametrii pentru fiecare dintre analize; instrumentul de analiză utilizează macro-funcțiile statistice sau de inginerie corespunzătoare și afișează rezultatele într-un tabel de ieșire. Unele instrumente auxiliare generează și diagrame auxiliare tabelelor de ieșire.

Principalele funcții de analiză statistică sunt: analiza varianței ANOVA (Single factor, Two-Factor With Replication, Two-Factor Without Replication); regresia; corelația; covarianța; statistici descriptive; netezirea exponențială; test F; analiză Fourier; histograme; media mobilă; rang și percentilă; generator de numere aleatoare; test t; test z; eșantionare.

Pentru a completa funcțiile modului Data Analysis MS Excel a lansat de curând pe piață modulul Analyse-it care conține peste 30 de proceduri pentru statistici descriptive, testarea normalității, compararea grupurilor, corelația și regresia precum și facilități de analiză a trendului.