

Lucrări practice

1. Generalități

Lucrările practice prezentate aici vin să completeze cunoștințele predate la curs, accentuând aspectele practice care implică prelucrarea propriu-zisă a unor date experimentale.

Toate datele utilizate au fost generate cu ajutorul unor programe de simulare, pornind de la situații reale prezentate în literatură. Lucrările practice au fost orientate spre utilizarea celui mai accesibil software pentru prelucrări statistice mai simple, și anume Microsoft Excel din pachetul Microsoft Office. Sunt trecute două variante de lucru: fie în Excel 2003, care este încă foarte răspândit, având meniul mai simplu – versiune pe care o utilizăm și în laboratorul disciplinei de informatică medicală, fie în Excel 2007, variantă solicitată de numeroși doctoranzi, care au la dispoziție acest soft. Fișierul de date este același și este accesibil pe internet la adresa:

www.medinfo.umft.ro/dim/biostatistica/tabele-date.xls, de unde poate fi copiat pe calculatorul personal. În prima parte sunt descrise tabelele cu urmează a fi folosite în continuare; unele date sunt folosite în mai multe lucrări practice.

În cadrul lucrărilor de laborator se distribuie fișierul și într-o versiune care conține rezultatele tuturor prelucrărilor expuse mai jos.

Lucrările sunt descrise sumar, comentariile fiind reduse la strictul necesar. S-a preferat stilul de redactare cu enumerarea pașilor de parcurs.

Denumirile unor comenzi în limba engleză au fost trecute în *italice*.

Nu s-au reluat noțiunile fundamentale din Excel. Este recomandabil – pentru cei care nu au lucrat anterior cu acest soft, să-și sacrifice câteva ore pentru o inițiere în elementele fundamentale privind calculul tabelar. În general, pentru parcurgerea lucrărilor practice s-au considerat cunoscute următoarele operațiuni și noțiuni:

- referirea unei celule (click cu mouse-ul pe celulă, sau referirea coordonatelor – coloană și linie, de ex.: B3),
- copierea datelor și copierea formulelor, pe linii și pe coloane, inclusiv înțelegerea modificărilor automate ale referințelor din formule,
- coordonate “înghețate” (utilizarea \$ pentru păstrarea coordonatelor din formule),
- selectare/deselectare celulă, linie, coloană, arie,
- referirea unui interval în meniuri (pe o linie, pe o coloană, pe mai multe linii și coloane),
- modificarea dimensiunilor coloanelor (cu mouse sau referire prin *Page Layout*),
- modificarea dimensiunilor unui *Chart*, completarea datelor din meniu, etc.

Sunt utile și alte noțiuni – filtrarea, folosirea funcțiilor etc., dar pot fi parcurse în paralel cu lucrările practice.

2. Descrierea sumară a tabelelor de date

Fișierul conține 7 secțiuni (WS - worksheet), selectabile din bara inferioară, prezentate în continuare. În Anexa 1 sunt trecute aceste tabele; coloanele foarte lungi au fost convertite "pe mai multe coloane" în forma tipărită.

WS1: grafic

Descriere worksheet

Date din studiul distribuția grupelor sanguine. În tabel-1 sunt centralizate datele obținute pe un lot de 400 persoane, 200 bărbați și 200 femei. În tabel-2 sunt datele pe un lot de 8 ori mai mic – 50 persoane, cu aceleași proporții ale grupelor sanguine. Sunt datele din exemplul 4.3 (cap. 4.2.3.1).

Structura:

- liniile 1, 2 – *labels* (titlul și capul de tabel),
- linia 3 – date bărbați (M),
- linia 4 – date femei (F),
- linia 5 – total [necalculat – va fi calculat în lucrările practice],
- coloana 1 – *labels*
- coloanele B – E: date pe grupele 0, A, B, AB
- coloana F – total – [necalculat – va fi calculat în lucrările practice].

WS2: par-st

Descriere worksheet

Date din studiul dezvoltarea somatică a copiilor în vârstă de 10 ani, pe un lot de 60 de copii.

Structura:

- liniile 1, 2 – *labels* (titlul și capul de tabel),
- liniile 3:62 – date,
- coloana A – nr. crt. (1- 60),
- coloana B – inițiale (2 litere) – necompletat,
- coloana C – sex (M sau F),
- coloana D – înălțimea în cm (cu 1 zecimală),
- coloana E – greutatea în kg (cu 1 zecimală),
- coloana F – indice de greutate corporală [necalculat – va fi calculat în lucrările practice].

WS3: teste

Descriere worksheet

Date din studiul privind efectul unui tratament cu un antihipertensiv asupra tensiunii arteriale, efectuat pe un lot de 60 subiecți;

Structura

- liniile 1, 2 – *labels* (titlul și capul de tabel),
- liniile 3:62 – date,
- coloana A: sexul,
- coloana B: tensiunea sistolică înainte de tratament,
- coloana C: tensiune sistolică după tratament,
- coloana D: tensiune diastolică înainte de tratament,
- coloana E: tensiunea diastolică după tratament.

WS4: val-dg

Descriere worksheet

Date din studiul: evaluarea unui test de sarcină efectuat pe un lot de 120 femei.

Structura

- liniile 1, 2 – *labels* (titlul și capul de tabel),
- liniile 3:122 – date,
- coloana A – nr. crt.,
- coloana B – rezultatul testului, not. cu „+” și „-”,
- coloana C – situația reală privind confirmarea sarcinii, notată la fel.

WS5: risc

Descriere worksheet

Date din studiul evaluarea riscului produs de fumat asupra tensiunii arteriale (risc de hipertensiune). Date dintr-un studiu transversal pe 120 persoane.

Structura

- liniile 1, 2 – *labels* (titlul și capul de tabel),
- liniile 3:162 – date,
- coloana A: tensiunea arterială sistolică,
- coloana B: fumător DA / NU, notat cu 1/0 (1 = DA, 0 = NU).

WS6: chi-sq

Descriere worksheet

Date din studiul distribuția grupelor sanguine. În exp.1 sunt datele pe un lot de 240 persoane, iar în exp.2 pe un lot de 10 ori mai mic, cu exact aceleași proporții ale grupelor sanguine.

Structura

- liniile 1, 2, 7 – *labels* (titlul și capul de tabel),
- linia 3 și 8 – date observate,
- linia 4 și 9 – date așteptate (teoretice), [necalculate – vor fi calculate în lucrările practice],
- coloana A și H – *labels*,
- coloanele B – E: date pe grupele 0, A, B, AB,

- coloana F – total – [necalculat – va fi calculat în lucrările practice],
- coloana I – valoarea lui p din testul chi-square.

WS7: regr

Descriere worksheet

Date din studiul dezvoltarea somatică a copiilor în vârstă de 10 ani, pe un lot de 40 copii.

Structura

- liniile 1, 2 – *labels* (titlul și capul de tabel),
- liniile 3:42 – date,
- coloana A – înălțimea în cm, fără zecimale,
- coloana B – greutatea în kg, cu 1 zecimală.

3. Descrierea lucrărilor practice pentru Excel 2007

Tema 1: Reprezentări grafice simple

Obiectiv: reprezentarea grafică a datelor din tabele

Worksheet utilizat: grafic

Modul de lucru:

Tema are 5 părți practice independente:

Subtema 1a: histograma 1 variabilă

- **Procedeu**
- se verifică dacă suntem în tab-ul *Home* al barei standard; dacă nu, se face clic pe *Home*
- se selectează din **tabel-2** porțiunea b8:e9 (2 linii x 4 coloane)
- din bara standard a meniului se alege tab-ul *Insert*, iar din grupul *Charts* se alege icon-ul *Column*; apare o fereastră tip "pop-up window", cu icon-uri în mai multe secțiuni;
- alegem varianta dorită – în exemplul nostru, din secțiunea *2-D Column* alegem primul icon (pentru coloane separate pentru fiecare grupă sanguină); apare o fereastră a chart-ului în mijlocul ecranului și banda "*ribbon*" din meniu se modifică;
- din grupul *Chart Style* alegem culorile din chart; de ex. *style 2* (culori diferite pe linii diferite – în cazul nostru esra o singură linie, deci toate coloanele vor apărea în aceeași culoare);
- din grupul *Chart Layout* alegem o variantă de reprezentare; de ex. *layout 1* (conține histograma, cu axa numerelor la stânga, precizarea claselor – grupele sanguine – sub coloane, cu titlu și legendă);
- pe chart, cu clic pe *Chart Title* se introduce noul titlu; de ex. "Distributia grupelor sanguine";
- din grupul *Data* selectăm icon-ul *Select Data*; apare o fereastră intitulată *Select Data Source*; (acum am putea schimba datele utilizate pentru reprezentare, dar lăsăm nemodificată căsuța *Chart data range*, care conține "=grafic!\$B\$8:\$E\$9", adică porțiunea selectată de noi); noi dorim acum să modificăm textul din legendă, așa că dăm

- clic pe "*Series1*" din cadrul deschis, apoi activăm *Edit*; apare o nouă fereastră în care, la *Series name* putem introduce textul dorit, de ex. "Lotul 2", terminând cu clic pe butonul *OK*; se închide fereastra mică și mai dăm *OK* și la fereastra *Select Data Source*;
- chart-ul este gata și poate fi mutat oriunde pe același worksheet, de ex. cu colțul sângea sus în H3 (sau se poate crea un alt worksheet – un clic dreapta pe mouse va deschide un pop-ul meniu din care selectăm *Move Chart*, iar în noua fereastră alegem butonul *New sheet* și îi dăm numele dorit).
 - **Interpretare:** fiecare coloană reprezintă nr persoanelor din grupa sanguină.

Subtema 1b: reprezentare sectorială

- **Procedeu**
- se verifică dacă suntem în tab-ul *Home* al barei standard
- se selectează din **tabel-2** porțiunea b8:e9 (2 linii x 4 coloane)
- din bara standard a meniului se alege tab-ul *Insert*, iar din grupul *Charts* se alege icon-ul *Pie*; apare o fereastră tip "pop-up window", cu icon-uri în mai multe secțiuni;
- din secțiunea *2-D Pie* alegem primul icon; apare fereastra chart-ului;
- din grupul *Chart Style* alegem culorile din chart; de ex. *style 3* (o culoare cu diverse saturații – vom avea acum 4 nuanțe, pentru cele 4 grupe sanguine);
- din grupul *Chart Layout* alegem reprezentarea; de ex. *layout 6*;
- continuăm ca mai sus: pe chart, cu clic pe *Chart Title* se introduce noul titlu;
- chart-ul este gata și poate fi mutat oriunde.
- **Interpretare:** reprezentare sectorială cu grupele sanguine în %, având afișat și procentul fiecare clasă.

Subtema 1c: histograma 2 variabile – varianta 1

- **Procedeu**
- se selectează din **tabel-1** porțiunea a2:e4 (2 linii x 4 coloane, cu grupele sanguine pe sexe)
- din bara standard se alege tab *Insert*, grup *Charts*, icon *Column*; apare o fereastră tip "pop-up window", cu icon-uri în mai multe secțiuni;
- din secțiunea *2-D Column* alegem al doilea icon (coloane separate pe grupe sanguine, dar suprapuse pentru sexe diferite);
- din grupul *Chart Style* alegem acum *style 1* (nuanțe de gri, varianta potrivită pentru pregătirea în vederea tipăririi în cărți)
- din grupul *Chart Layout* alegem *layout 3*; completăm titlul; chartul este gata;
- **Interpretare:** fiecare coloană are numărul persoanelor din grupa sanguină; coloanele au 2 segmente: bărbați și femei.

Subtema 1d: histogramă 2 variabile – varianta 2

- **Procedeu**
- se selectează din **tabel-1** din nou porțiunea a2:e4
- din bara standard se alege tab *Insert*, grup *Charts*, icon *Column*;
- din secțiunea *3-D Column* alegem al treilea icon ("100% Stacked Column", coloane separate pe grupe sanguine, dar suprapuse pentru sexe diferite);
- din grupul *Data* selectăm *Switch Row/Column* – vom avea 2 coloane în loc de 4; acum coloanele sunt pe sexe, iar porțiunile din ele reprezintă procentul corespunzător fiecărei grupe sanguine;

- continuăm procedura standard: din grupul *Chart Style* alegem acum *style 8*, iar din grupul *Chart Layout* alegem *layout 4*;
- **Interpretare:** 2 coloane bărbați și femei, fiecare având în % grupele sanguine.

Subtema 1e: histogramă 2 variabile – varianta 3

- **Procedeu**
- se selectează din **tabel-1** din nou porțiunea a2:e4
- din bara standard se alege tab *Insert*, grup *Charts*, icon *Column*;
- din secțiunea *3-D Column* alegem al patrulea icon (“3-D Column”, coloane separate pe grupe sanguine și pe sexe);
- continuăm procedura standard: grup *Chart Style* - *style 2*, grup *Chart Layout* *layout 7*;
- **Interpretare:** 2 serii (bărbați și femei) a câte 4 coloane (grupele sanguine), pe fiecare coloană este reprezentat numărul de subiecți.

Tema 2: Calcule simple în Excel

Obiectiv: utilizarea operatorilor de calcul, introducerea formulelor

Worksheet utilizat: grafic și par-st

Modul de lucru:

Tema are două părți independente:

Subtema 2a: calcul totaluri și procente

- **Procedeu**
- deschidem Worksheet **grafic**,
- calcul totaluri coloane:
 - selectăm b5 (click),
 - din bara de comenzi selectăm tab *Formulas*, grup *AutoSum* (Σ); se deschide un pop-up meniu din care selectăm *Sum*; atât în linia de comenzi cât și în căsuța b5 apare „= SUM (B3:B4)”; acceptăm tastând Enter;
 - pentru copierea formulei de sumă în căsuțele din dreapta, selectăm b5 și Ctrl + C (Copy),
 - selectăm c5:e5 și Ctrl + V (Paste);
- totaluri linii
 - procedăm similar pentru total pe linii: select f3 / bara de comenzi tab *Formulas*, grup *AutoSum* (Σ) / din pop-up selectăm *Sum*; apare „= SUM (B3:E3)” / acceptăm (Enter); copiem apoi în căsuțele de dedesubt: f3 Copy / select f4:f5 / Paste)
- completăm tabelul cu calcul procente
 - select a6 – scriem „%”
 - introducem formula de calcul a procentelor fiecărei grupe sanguine (total grupă x 100 / total general, aflat în f5): select b6 / în linia de comenzi: “= b5*100 / \$f5 (la copiere în alte căsuțe, coloana f la numitor va rămâne nemodificată)
 - copy din b6 în c6:f6
- **Interpretare:** tabelele sunt acum completate cu totaluri și procente.

Subtema 2b: calculul unor variabile intermediare sau finale

Vom calcula indicele de masă corporală BMI (Body Mass Index) conform relației BMI = Greut (kg) / înălțime² (m)

- **Procedeu**
- deschidem worksheet **par-st**,
- introducem 2 noi coloane:
 - F va conține înălțimea în m (în F2 notăm $h(m)$)
 - G va conține BMI (în G2 notăm BMI)
- selectăm căsuța F3
- în bara de inserare comenzi introducem: „= d3/100”;
- copiem formula pe toată coloana F (de la f3:f62) (revedem comenzile Ctrl + C pentru Copy și Ctrl + V pentru Paste)
- selectăm G3
- introducem comanda: „= e3 / (f3 * f3)”
- copiem formula de la g4:g62
- dorim afișarea BMI cu două zecimale: selectăm coloana G (cu mouse), apoi din bară, tab *Home*, grup *Cells*, apare pop-up și selectăm *Format Cells*; apare o fereastră, cu meniu cu 6 tab-uri; selectăm *Number*, iar din pop-up-ul care se deschide alegem *Number*, apoi la *decimal places*, din săgeți alegem "2" și *OK*,
- dorim evidențierea copiilor subponderali, normali, supraponderali și obezi prin culori diferite pe tabel. (Limitele între categoriile de mai sus sunt pentru copiii de 10 ani: 14.2, 19.4 și 22.1). Selectăm prima celulă din coloana BMI (adică g3), apoi, în bară, la tab *Home*, grupul *Styles*, icon-ul *Conditional Formatting* – apare un meniu pop-up; selectăm *Highlight Cell Rules*; lateral apare încă un meniu cu lista de condiții: Greater than >, Less than <, Between, etc; alegem întâi "*Greater than*"; se deschide o fereastră în care putem completa o valoare – vom pune limita superioară 22.1, peste care interpretăm obezitatea, apoi la condiție alegem "*Light Red Fill with Dark Red Text*" / *OK*; rămânem selectați pe g3 și reluăm *Conditional Formatting* / *Highlight Cell Rules*, acum luăm "*Between*" și vom introduce limitele 19.4 și 22.1, cu "*Light Yellow Fill with Dark Yellow Text*" / *OK*; vom mai alege "*Green ...*" pentru intervalul normal 14.2 – 19.4 și simplu "*Red text*" pentru subponderali, adică sub 14.2. Valoarea noastră va îndeplini una din condiții; când condiția este îndeplinită se va vedea culoarea în celulă; după introducerea tuturor condițiilor, selectăm din nou celula g3 și cu Copy (sau Ctrl+C) vom copia condițiile pe toată coloana G în jos. Vom folosi copierea specială din *Home* / *Clipboard* / *Paste* / *Paste Special* / *Formats*, altfel copiază și valorile. Vom vedea culorile alese condiționat, în funcție de valori, în celulele corespunzătoare.
- **Interpretare:** apare coloana întreagă BMI cu 2 zecimale, colorată astfel încât ies în evidență cele 4 clase în funcție de valoarea BMI.

Tema 3: Parametrii statistici

Obiectiv: determinarea parametrilor statistici ai unui eșantion

Worksheet utilizat: par-st

Modul de lucru:

Tema are două părți asemănătoare:

Tema 3a: parametrii statistici pentru 1 variabilă, pe un eșantion întreg pentru înălțimea grupului de copii

- **Procedeu**
- din bara de comenzi selectăm tab *Data*, grupul *Analysis*; (dacă nu există acest grup, el va putea fi introdus cu *Add-Ins*;
- selectăm „*Data Analysis*”; se deschide o fereastră din care selectăm linia „*Descriptive Statistics*” / OK,
- în noua fereastră deschisă selectăm:
 - *input range*: d2:d62
 - *grouped by*: Columns
 - bifăm „*labels in first row*” (este vorba de d2)
 - decidem unde să se afișeze rezultatele – în secțiunea *Output Options* putem alege fie un nou worksheet, sau chiar alt fișier, sau putem plasa tabelul în worksheet-ul nostru de lucru (par-st), bifând *output range*;
 - în *output range* – alegem o căsuță în dreapta, de ex. i3 (aici va fi colțul din stânga sus al tabelului cu rezultatele de statistică descriptivă), sau putem plasa tabelul sub date, de ex. i64
 - bifăm „*summary statistics*”
 - (opțional putem bifa și *confidence level for mean / ok*).
- **Interpretare:** apare tabelul cu rezultatele statistice privind înălțimea copiilor din lot, care este comentat în curs, cap. 2.3.1 (tabelul 2.2).

Tema 3b: parametrii statistici pentru 1 variabilă, pe un subgrup.

Vom obține parametrii statistici pentru înălțimea fetelor din grupul de copii.

- **Procedeu**
- selectăm coloana C (sex)
- din bara de comenzi, tab „*Data*”, grupul *Sort & Filter*, icon-ul *Filter*;
- în căsuța c1 apare un buton de selecție; la tastare pe el se deschide un pop-up meniu, din care putem selecta F (numai fetele); dacă este necesar debifăm celelalte opțiuni; tabelul se comprimă, liniile corespunzătoare băieților nu mai sunt vizibile;
- vom crea un tabel separat care să conțină numai fetele; din coloana D selectăm înălțimile celor 24 fete, pe care le copiem într-un nou Worksheet (din tab *Home*, grup *Cells*, icon *Insert / Worksheet*), să zicem de la a1 în jos (clic pe a1, apoi CTRL + V)
- mai departe procedăm ca mai sus: tab *Data*, grup *Analysis*, icon *Data Analysis*; „*Descriptive Statistics*” etc. (atenție: datele sunt acum în a1:a24, fără *labels in first row*, etc);
- putem da un nume acestui nou worksheet: în bara de jos, cu mouse-ul pe Sheet1 care este acum activ, clic drapta – se deschide un meniu din care selectăm *Rename*, apoi introducem noul nume, de ex. "**par-st2**".
- **Interpretare:** tabelul obținut conține acum rezultatele statistice privind înălțimea fetelor din lot.

Tema 4: Histograme din fișiere de date brute

Obiectiv: crearea limitelor intervalelor pentru histograme și ridicarea histogramelor. Limitele claselor, ce poartă denumirea de "bin", vor fi trecute într-o coloană ce o vom crea

alături de tabelul de date. Este bine ca valorile pentru bin-uri să se aleagă încât să se intercaleze între valorile experimentale. Limitele se pot alege arbitrar – noi vom lucra în două variante, cu clase late de 1 cm, respectiv cu lățimea optimă (rotunjită), dată de relația (2.4), p.27, adică $h = (X_M - X_m) / (1 + 3.322 \lg N)$.

Worksheet utilizat: par-st

Modul de lucru

Tema are 2 părți similare:

Subtema 4a: crearea limitelor pentru histograme cu lățimea clasei de 1 cm.

- **Procedeu**
- avem pe coloanele I și J parametrii statistici calculați în Tema 3a.
- pe coloana L vom crea limitele pentru prima histogramă
- selectăm L2, introducem valoarea cu lățimea clasei („1” în cazul nostru)
- selectăm L3, introducem o valoare pentru limita din dreapta a primei clase; dacă dorim să avem clasele centrate pe valori întregi sau rotunde (în cm), vom lua pentru limite jumătăți de unitate – pentru noi, având Minimum 128.2 (din tabelul obținut în subtema 3a) vom lua ca primă valoare 128.5;
- dorim să incrementăm aceste limite bin cu lățimea clasei la fiecare pas, deci completăm coloana J în jos astfel:
 - $L4 = L3 + L\$2$ (am fixat pasul de incrementare)
 - copiem L4 în jos încât să acoperim valoarea maximă (149.5 în cazul nostru); putem să și depășim valoarea, însă ulterior vom folosi doar intervalul ce acoperă datele reale;
- din bară tab *Data* / grup *Analysis* / icon *Data Analysis* / *Histogram* / OK,
- *input range*: D3:D62 (unde sunt datele de reprezentat),
- *bin range*: L3:L5 (unde am creat limitele),
- putem vizualiza graficul chiar în acest worksheet; luăm *output range* "N2",
- bifăm *chart output*, apoi OK;
- apare tabelul de valori în clase și histograma; graficul poate fi "mărit", prin "drag" cu mouse-ul;
- pregătim în final histograma cu date suplimentare (titlu, axe etc).
- **Interpretare:** tabelul afișat în stânga histogramei conține numărul de copii din fiecare clasă de înălțime cu lățimea de 1 cm. Prima clasă are marginea superioară valoarea din coloana bin, adică 128.5, deci cuprinde toate valorile până la 128.5 inclusiv. Observăm tendința centrală, însă histograma are "numeroși dinți", datorită claselor prea înguste. Vom putea compara acest grafic cu următorul, pe care îl realizăm cu lățimea adecvată.

Subtema 4b: crearea histogramelor cu lățime optimă

- **Procedeu**
- calculul lățimii optime
- pregătim datele într-o zonă convenabilă, de ex sub tabelul creat în subtema 3a (în cazul în care am ales pentru *output range* căsuța I3; în acest caz avem valoarea maximă în J15, cea minimă în J14 și N în J17; dacă tabelul creat la *Descriptive Statistics* este în altă parte, trebuie schimbate corespunzător referințele din relațiile de mai jos:

$$J30 = \log_{10} (J17)$$

$$J31 = 1 + 3.322 * J30$$

$$J32 = (J15 - J14) / J31$$

- valoarea din J32 (3.083 în cazul nostru) se rotunjește (la 3 în cazul nostru) și se introduce în L30 să reprezinte lățimea clasei,
- pe coloana L vor apare noile valori bin pentru clase de câte 3 cm. Dacă dorim să avem mijlocul clasei o valoare rotundă, este cel mai bine să ne orientăm după o valoare centrală; în cazul nostru având media la 139,3 cm (rotund 139), vom avea centrale din 3 în 3 cm, în sus și în jos, deci 139; 136; 133; 130; 127 etc. Pentru centru în 127 vom avea limita superioară la 128,5 cm, deci vom pune această valoare în L31, apoi $L32 = L31 + L30$; obținem limitele dorite în coloana L31:L39 (pentru *Bin Range* în fereastra ce se deschide la *Histogram*);
- mai departe procedăm ca la histograma precedentă, alegând la output range N30 (pentru alinierea datelor).
- **Interpretare:** tabelul afișat în stânga histogramei conține numărul de copii din fiecare clasă de înălțime cu lățimea de 3 cm. Pe grafic se observă o încadrare mult mai bună și dispariția "dinților".

Tema 5: Teste statistice

Obiectiv: compararea a două serii numerice sau calitative.

Worksheet utilizat: teste și chi-sq

Modul de lucru

Tema are 3 părți distincte, asemănătoare, una pentru testul t pereche și una pentru testul t nepereche, ultima pentru testul chi-sq.

Subtema 5a: testul t pereche

Vom aplica testul t pereche pentru a compara diferențele între tensiunile sistolice înainte și după tratament.

- **Procedeu**
- deschidem worksheet-ul **teste**;
- din bară, tab *Data* / grup *Analysis* / icon *Data Analysis* / *t – Test: Paired Two Samples for Means*
- în fereastra deschisă alegem:
 - *variable 1 range:* b2:b62 (coloana B – tensiuni sistolice înainte de tratament, coloana C – tensiuni sistolice după tratament)
 - *variable 2 range:* c2:c62
 - *hypothesized mean / difference:* 0 (sau lăsăm necompletat)
 - bifăm *labels* (avem în b2 și c2)
 - lăsăm Alpha 0.05
 - selectăm *output range* (de ex g2)
- apare tabelul de forma celui din cap. 4.2.1.3, tabelul 4.2.b
- **Interpretare:** în tabelul afișat apar multe date de interes: mediile celor două serii, varianțele (varianța este pătratul deviației standard), cea mai de interes valoare fiind din căsuța "P(T<=t) two-tail", care, fiind 0.0356 (sub 0.05), deci indică o diferență semnificativă.

Subtema 5b: testul t nepereche

Vom aplica testul t nepereche pentru a compara tensiunea sistolică înainte de tratament la bărbați, comparativ cu femei.

- **Procedeu**
- deschidem worksheet-ul **teste**;
- selectăm coloana A (sex),
- cu *Data / Sort & Filter / Filter*, preluăm pe rând:
 - datele pentru bărbați: selectăm M (debifăm celelalte) și copiem (Ctrl+C) datele rămase vizibile din coloana Sys-1, datele selectate sunt acum în memoria "clipboard" a calculatorului;
- creăm un nou workheet (*Home / Cells / Insert / Insert Sheet*) în care vom pune datele bărbaților în coloana A și ale femeilor în B:
 - în A1 scriem M, în B1 scriem F
 - copiem valorile din clipboard din A2 în jos (click pe A2, apoi Ctrl+V)
 - revenim în worksheet "teste"
 - în butonul de filtru debifăm M și bifăm F, apoi copiem valorile coloanei Sys-1 în noul worksheet, din B2 în jos (coloanele pot avea lungimi diferite!)
- din bară, tab *Data / grup Analysis / icon Data Analysis / t – Test: Two Sample Assuming Unequal Variances* (ulterior, dacă vedem că varianțele sunt apropiate, putem relua pentru *Equal Variances*)
- coloanele sunt a1:a28 și b1:b34, cu *Labels*,
- luăm *output range* d2
- din *Tools / Data Analysis / t – test: Two – Sample Assuming Unequal Variances / OK*
- obținem un tabel similar cu tabelul 4.2.a din cap. 4.2.1.2,
- putem în final denumi noul worksheet "teste2".
- **Interpretare:** în tabelul afișat valoarea din căsuța "P(T<=t) two-tail", este acum 0.91, mult peste 0.05, deci diferențele sunt ne semnificative.

Subtema 5c:

- **Procedeu**
- Se deschide worksheet *chi-sq*
- Presupunem o distribuție uniformă a persoanelor pe grupe sanguine, atunci valorile așteptate (căsuțele b4 : e4) se pot calcula cu relația $\text{total-general} / 4 = \$F\$3 / 4$
- Pentru calculul valorilor așteptate în tabelul "**exp.1**", facem întâi totalul pentru linia "observed":
- clic pe f3, apoi *Formulas / Function Library / AutoSum / Sum / B3:E3 / Enter*
- clic pe b4, apoi " $= \$f\$3/4$ " și copiem formula din b4 în c4:e4.
- Pentru aplicarea testului
 - selectăm căsuța de rezultat (i3)
 - selectăm funcția f_x ; se deschide fereastra "*Insert Function*"
 - pe linia "*Select a category*" alegem "All" sau "Statistical", apoi din fereastra mai largă de dedesubt alegem CHITEST /OK,
 - apare fereastra „*Function Arguments*” în care luăm:
- *actual_range* (se referă la valori experimentale, adică b3:e3 // va fi b8:e8 în tabelul **exp.2**)

- *expected_range* (se referă la valori teoretice; atunci luăm b4:e4, respectiv b9:e9 pentru tabelul de jos)
- **Interpretare:** apare valoarea lui p (0,000828) < 0,001 deci vom avea diferențe extrem de semnificative între valorile experimentale și cele teoretice; deci vom respinge ipoteza de zero H_0 .
- Repetăm toate operațiunile pentru un lot mult mai mic (tabelul **exp.2**), dar păstrând proporțiile – observăm că, în cazul loturilor mici, diferențele nu mai sunt semnificative (acum avem $p = 0.644$), deși proporțiile claselor sunt aceleași.

Tema 6: Analiza corelației

Obiectiv: determinarea coeficientului de corelație și trasarea dreptei de regresie

Worksheet utilizat: regr

Modul de lucru

Tema are 2 părți complementare

Subtema 6a: comanda *correlation*

- **Procedeu**
- deschidem worksheet **regr**
- completăm coloana C cu: indicele BMI (v. Tema 2b)
- alternativ putem introduce în căsuța c3 o relație directă de calcul pentru BMI "= b3 * 10000 / (a3 * a3)"
- din *Data / Analysis / Data Analysis / Correlation* se completează în fereastra deschisă:
 - *input range* a2:c42 (Obs: variabilele de corelat trebuie să fie plasate în 2 coloane vecine în tabel); coloana A va fi pe Ox iar B pe Oy
 - *grouped by columns*
 - bifăm *labels* (sunt a2, b2 și c2)
 - selectăm ieșirea (de ex e2)
- **Interpretare:** apare un tabel 3 x 3 cu coeficienții de corelație separat pentru corelația înălțime-greutate (0,751 în cazul nostru, deci avem o corelație directă, destul de puternică), corelația înălțime - bmi (-0,690 = inversă, medie) și corelația greutate – bmi (-0,043 = foarte slabă, practic absentă)

Subtema 6b: comanda *regression*

- **Procedeu**
- din *Data / Analysis / Data Analysis / Regression* completăm:
 - *input Y range:* b2:b42 (greutatea),
 - *input X range:* a2:a42 (înălțimea),
 - bifăm *labels* (sunt a2 și b2),
 - nu bifăm constant is zero (dacă se bifează atunci se calculează dreapta de regresie care trece prin origine),
 - nu bifăm: Standardized residuals, Residual Plots, nici Normal Probability Plots,
 - bifăm *Line Fit Plot* – pentru graficul dreptei de regresie,
- alegem *output range* (de ex e7),
- apar rezultatele (*Summary Output*) în 4 tabele și un grafic;
- este posibil să fie necesară prelucrarea graficului (dublu clic pe grafic):

- îi mărim dimensiunea cu mouse-ul pe un colț,
- din bară: *Insert / Charts / Scatter / primul icon* (numai puncte),
- din bara nouă (Chart Tools), tab-ul *Design*, grupul *Chart Layouts*, selectăm *Layout 3* (al 3-lea icon), care are un grid, se văd punctele experimentale și cele calculate "predicted",
- cu clic dreapta pe axa verticală apare o fereastră și selectăm *Format Axis*, în care putem lua pentru *Minimum* valoarea 25 în loc de 0 (punctele noastre se găsesc toate în partea superioară)
- **Interpretare:** din primul tabel (*Regression Statistics*) urmărim Multiple R – coeficientul de corelație (0,751 în cazul nostru)
- din al doilea tabel (*ANOVA*) urmărim Significance F (2.32×10^{-8} ceea ce înseamnă că avem o probabilitate extrem de mică să obținem din întâmplare așa un coeficient de corelație (este echivalentul lui p)
- din al treilea tabel găsim:
 - ordonata la origine = Intercept (-2.89 ± 4.99)
 - panta (slope) – în căsuța sub valoare intercept ($0,255 \pm 0.036$)
 - pentru ambele sunt date limitele 95%
- al patrulea tabel arată valorile calculate pentru greutatea calculată (predicted) corespunzătoare înălțimilor, conform:

$$\text{greut. calc} = \text{intercept} + \text{panta} * \text{înălțime, sau}$$

$$\text{greut. calc} = -2,89 + 0,255 * \text{înălțimea}$$
- următoarea coloană conține diferențele:

$$\text{residual} = \text{greut (reală)} - \text{greut. calc}$$
- În grafic mai facem următoarele operații:
 - trasăm dreapta de regresie: cu mouse, click dreapta pe un punct experimental (bleumarin) – se deschide un menu mic și selectăm „Add Trendline” și selectăm trend Linear (prima căsuță)
 - putem elimina valorile prezise cu un click dreapta pe un punct teoretic (magenta), iar din fereastra mică selectăm „delete”.

Tema 7: Analiza riscului

Obiectiv: realizarea tabelului de contingență pornind de la date brute și calculul indicatorilor de risc.

Worksheet utilizat: risc

Modul de lucru

Tema are 2 etape distincte.

Subtema 7a: realizarea tabelului de contingență

Varianta 1

- **Procedura**
- (a) **Construim scheletul tabelului de contingență**, fie sub date (sub linia 162) sau pe un nou worksheet (alegem această variantă), numit **risc2**
- *Home / Cells / Insert / Insert Sheet //* click dreapta pe nume / *rename / risc2*
- selectăm regiunea a1:d4;

- din bară, tab *Home*, grup *Font*, icon de borders (pătrățel), alegem varianta cu linii pentru toate căsuțele; introducem notațiile clasice în A2:A4 (E+, E-, total), respectiv în B1:D1 (B+, B-, total).
- Completăm D4 cu valoarea numărului total (160 în cazul nostru).
- **(b) Transformarea variabilei de expunere la risc**
- revenim în woksheet **risc**
- Vom transforma variabila numerică TA – Sys în variabila calitativă Hipertens cu convenția:
 - dacă $TA - Sys \geq 140$ atunci Hipertens = 1 (= DA)
 - dacă $TA - Sys < 140$ atunci Hipertens = 0 (= NU)
- În C2 notăm *label* Hipertens
- Selectăm C3 și dăm click pe f_x ; se deschide fereastra „*Insert function*”, în care luăm:
 - *select category*: All
 - baleiem toate funcțiile până la IF, selectăm IF, se deschide fereastra „*Function Arguments*”, în care luăm:
 - = *logical test*: $a3 > = 140$
 - = *value if true*: 1
 - = *value if false*: 0 / ok
 - în căsuța c3 apare 0 sau 1 (în cazul nostru 0)
- copiem formula din c3 în jos (c4:c162) (revedeți funcțiile Ctrl + C și Ctrl + V)
- **(c) Aflăm totalurile E+ și B+ :**
- selectăm b163, apoi în bară: *Formulas / Function Library / pe AutoSum Σ și Sum* (apare propunerea SUM (B3:B162), o acceptăm (enter); apare totalul pe coloana B (fumat); copiem valoarea în sheet **risc2**, D2 (76 în cazul nostru) (Obs: se exersează Copy din Sheet **risc** cu Paste în Sheet **risc2** cu: Select D4, apoi =, trecere la Sheet risc și click pe B163),
- similar cu c163 pe care o tecem în B4 (36 în cazul nostru),
- **(d)** Selectăm coloana B și filtrăm numai fumătorii (din bară *Data / Sort & Filter / Filter*), apoi click pe butonul din coloana B și din micul meniu ce apare selectăm 1; dispar liniile ce aveau 0 în coloana B;
- selectăm o căsuță pe coloana C jos, sub ultima linie scrisă (va fi C166 în cazul nostru); introducem funcția SUBTOTAL (*Formulas / Function Library / icon Math & Trig / SUBTOTAL*) cu proprietățile:
 - *function_num*: luăm 9 (se va face totalul ignorând valorile ascunse),
 - *Ref 1*: luăm c3:c162 /OK (precizăm toată coloana, știind că valorile ascunse vor fi ignorate);
- apare totalul pentru care aveam 1 și în coloana B și în coloana C, adică sunt și E+ și B+; introducem această valoare (clic pe c166 și Ctrl+C) (32 în cazul nostru) în tabelul din risc2, căsuța b2; Obs: preferăm să folosim Paste Special / Value (prin clic dreapta), altfel se dorește copierea formulei de subtotal!
- Restul tabelului se completează prin diferențe:
- $c2 = d2 - b2$; apoi copy c2 în c4,
- $b3 = b4 - b2$; apoi copy b3 în c3 și d3.

Varianta 2 – se modifică doar punctul (d).

- **Procedura**
- **(d)** Introducem coloana D, numită „e+ / b+” (scriem în d2)

- Selectăm d3, apoi f_x și alegem funcția AND, (dacă nu se vede se selectează întâi la category "All") în care luăm:
 - *Logical 1*: b3 = 1
 - *Logical 2*: c3 = 1 / ok
- apare TRUE sau FALSE (FALSE în cazul nostru)
- copiem formula AND din d3 (adică b3 = 1, c3 = 1) în căsuțele d4:d:162; vor fi TRUE numai căsuțele pacienților care au 1 și 1 (fumează și au hipertensiune)
- selectăm d163 și introducem funcția COUNTIF; în fereastra ei introducem:
 - *Range*: d3:d:162
 - *Criteria*: TRUE / ok
- apare valoarea corespunzătoare (32 în cazul nostru)

Varianta 3

- Procedura

O metodă care poate fi folosită în toate situațiile în care dorim construirea unui tabel de contingență 2 x 2 pornind de la date dihotomice trecute în două coloane constă în transformarea reprezentării în cele două coloane astfel:

- o coloană să conțină numai valori 0 și 1 pentru NU (-) și DA (+)
- cealaltă coloană să conțină numai valori 0 și 2 pentru NU (-) și DA (+).

Astfel, prin suma pe linii obținem segregarea celor 4 variante:

NU & NU devine $0 + 0 = 0$
 DA & NU devine $1 + 0 = 1$
 NU & DA devine $0 + 2 = 2$
 DA & DA devine $1 + 2 = 3$.

Deci, dacă din coloana de sumă (o vom numi Sumă Ponderată) numărăm câți de 0, 1, 2 sau 3 avem, obținem exact cele 4 căsuțe din tabelul de contingență.

De fapt nici nu este nevoie să creăm separat coloana a doua cu 0 și 2, ci o lăsăm tot cu 0 și 1 iar în sumă să luăm valoarea din coloana a doua înmulțită cu 2, adică vom calcula:

$$\text{suma} = \text{col.1} + 2 * \text{col.2}$$

Vom proceda astfel:

- prima coloană 0/1 este B și corespunde E+/E-
- a doua coloană 0/1 este C, creată în punctul (b) din Varianta 1 și corespunde B+/B-
- creăm coloana E cu suma ponderată (SP) a valorilor din B și C
- astfel pentru E3 luăm $= B3 + 2 * C3$
- copiem E3 în jos, de la E4:E162

În **risc2** putem reconstrui tabelul de contingență în aria a6:d9, cu același notații.

Interpretare: Am obținut coloana E cu următoarele proprietăți:

- dacă are valoarea 0, pacientul este e- și b- (căsuța c3 sau c8 din risc 2)
 - dacă are valoarea 1, pacientul este e+ și b- (c2 / c7)
 - dacă are valoarea 2, pacientul este e- și b+ (b3 / b8)
 - dacă are valoarea 3, pacientul este e+ și b+ (b2 / b7).
- folosim funcția COUNTIF de 4 ori, de fiecare dată având Range risc!\$E\$3:\$E\$162, direct în risc 2, precizând că:
 - dacă punem criteria = 0 obținem valoarea din c3 (sau c8) etc.

- completăm tabelul cu totaluri.

Subtema 7b: calculul indicatorilor

- Folosim pachetul statistic **Epi Info**
- Din meniul principal selectăm *Utilities*,
- Selectăm *Statcalc*,
- Completăm tabelul (în cazul nostru avem 32 și 44 respectiv 4 și 80)
- **Interpretare:** Apare tabelul de rezultate, cu interpretare conform cap. 6.4.2, figura 6.3.

Tema 8: Validarea testelor diagnostice

Obiectiv: crearea tabelului de contingență și calcularea indicatorilor.

Worksheet utilizat: val-dg

Modul de lucru

Tema are 2 părți succesive: crearea tabelului, respectiv calculul indicatorilor.

Subtema 8a: crearea tabelului de contingență.

Folosim notațiile din curs cap. 8.1, tabel 8.1

În worksheet am notat cu + sau – testul pozitiv sau negativ, respectiv prezența sau absența bolii.

Construim scheletul tabelului de contingență (similar temei 7a) cu următoarele modificări:

- spațiul H1:K4 (nu creem acum un alt worksheet, ci alcătuim tabelul în dreapta datelor),
- notăm h2 și h3 cu B+ și B-, iar i1 și j1 cu T+ și T-; adăugăm "total" în h4 și k1.

Folosim o procedură similară cu cea din varianta 3 a temei 7a:

- Vom crea coloana D, cu valori 0 și 1 corespunzătoare lui T- și T+
- Coloana E, cu valori 0 și 1 corespunzătoare lui B- și B+
- Coloana F pentru suma ponderată
 - În D3 introducem prin f_x funcția IF cu:
= *logical test*: b3 = „,+”
= *Value if true*: 1, *value if false*: 0
 - Copiem D3 în E3
 - clic pe f3, apoi "=d3+2*e3"
 - selectăm d3:f3, apoi Ctrl+C
 - Copiem totul în D4:F122 (pe trei coloane deodată).
- În coloana F avem valori între 0 și 3, cu semnificația:
 - b+ / t+ valoarea 3 (va fi în căsuța i2)
 - b+ / t- valoarea 2 (j2)
 - b- / t+ valoarea 1 (i3)
 - b- / t- valoarea 0 (j3).

Cu COUNTIF vom completa căsuțele din tabel:

i2 = Countif (\$f\$3:\$f\$122,3)

j2 = Countif (\$f\$3:\$f\$122,2)

i3 = Countif (\$f\$3:\$f\$122,1)

$j_3 = \text{Countif}(\$f\$3:\$f\$122,0)$.

Completăm tabelul cu totaluri.

Subtema 8b: calculul indicatorilor.

- numele lor simbolic – în h8 : h16
- valorile în i8:i16
- vom folosi relațiile:
 - i8 FP (fals pozitivi) = i_3 / k_3
 - i9 FN (fals negativi) = j_2 / k_2
 - i10 SN (sensibilitate) = $[i_2 = \text{TP (real / „true” pozitivi)}] = i_2 / k_2$
 - i11 SP (specificitate) = $[j_3 = \text{TN (real / „true” negativi)}] = j_3 / k_3$
 - i12 VPP (valoarea predictivă pozitivă) = i_2 / i_4
 - i13 VPN (valoarea predictivă negativă) = j_3 / j_4
 - i14 AC (acuratețea) = $(i_2 + j_3) / k_4$
 - i15 RE (rata erorii de clasificare) = $(i_3 + j_2) / k_4$
 - i16 Y (indice Youden) = $i_{10} + i_{11} - 1$.

Interpretarea rezultatelor se face conform celor prezentate în cap. 8.2 și cap. 8.3.

Pentru Excel 2003

Tema 1: Reprezentări grafice simple

Obiectiv: reprezentarea grafică a datelor din tabele

Worksheet utilizat: grafic

Modul de lucru:

Tema are 5 părți practice independente:

Subtema 1a: histograma 1 variabilă

- **Procedeu**
- se selectează din **tabel-2** porțiunea b8: e9 (2 linii x 4 coloane)
- din bara standard se selectează “*Chart Wizzard*”;
- se va deschide o fereastră în care se lucrează 4 pași:
 - se alege *chart type* “column”, *sub-type1* ”clustered column” (prima căsuță) / Next
 - *data range:* b8:e9; *series in:* Rows [apare graficul] / Next
 - se completează *chart title* (de ex: Repartiție grupe sanguine), *category X axis* (Grupele), *Value Y axis* (număr) / Next
 - *chart location* – fie new sheet, fie object in [grafic] / Finish
- **Interpretare:** fiecare coloană reprezintă nr persoanelor din grupa sanguină.
-

Subtema 1b: reprezentare sectorială

- **Procedeu**
- folosim aceleași date din **tabel-2**
- selectăm “*Chart Wizzard*”
 - *chart type:* pie, subtype 1 “pie” / Next
 - *data range:* b8:e9 *series:* Rows / Next
 - titlul; *data labels:* “values” și “percentage” / Next
 - *location* / Finish
- **Interpretare:** reprezentare sectorială cu grupele sanguine în %, având afișat și procentul pentru fiecare clasă.
-

Subtema 1c: histograma 2 variabile – varianta 1

- **Procedeu**
- folosim aceleași date din **tabel-1**
- din bara standard se selectează din nou “*Chart Wizzard*”
 - *chart type* “column”, sub-type 5 “Stacked Column” / Next
 - *data range:* a2:e4, *series in* Rows / Next
 - se completează titlul și numele axelor / Next
 - se alege locația / Finish
- **Interpretare:** fiecare coloană are numărul persoanelor din grupa sanguină; coloanele au 2 segmente: bărbați și femei

Subtema 1d: histogramă 2 variabile – varianta 2

- **Procedeu**
- folosim aceleași date din **tabel-1**
- selectăm “*Chart Wizzard*”
 - *chart type* “column”, subtype 3 “100% Stacked Column with a 3-D visual effect”
 - *data range:* a2:e4, series in Columns / Next
 - se completează titlul și axele / Next
 - se alege locația / Finish
- **Interpretare:** 2 coloane - bărbați și femei, fiecare având în % grupele sanguine
-

Subtema 1e: histogramă 2 variabile – varianta 3

- **Procedeu**
- folosim aceleași date din **tabel-1**
- selectăm “*Chart Wizzard*”
 - *chart type* “column”, subtype 6 "3-D Column"
 - *data range:* a2:e4, series in Rows / Next
 - se completează titlul și axele / Next
 - se alege locația / Finish
- **Interpretare:** 2 serii (bărbați și femei) a câte 4 coloane (grupele sanguine), pe fiecare coloană este reprezentat numărul de subiecți.

Tema 2: Calcule simple în Excel

Obiectiv: utilizarea operatorilor de calcul, introducerea formulelor

Worksheet utilizat: grafic și par-st

Modul de lucru:

Tema are două părți independente.

Subtema 2a: calcul totaluri și procente

- **Procedeu**
- deschidem Worksheet: **grafic**
- calcul totaluri coloane
 - selectăm b5 (click)
 - selectăm linia de comenzi
 - selectăm Σ ; apare „= SUM (B3:B4)”; acceptăm tastând Enter
 - selectăm b5 și Ctrl + C (Copy)
 - selectăm c5:e5 și Ctrl + V (Paste)
- totaluri linii
 - procedăm similar pentru total pe linii (select f3 / linia comenzi / Σ / f3 Copy / f4:f5 Paste)
- completăm tabelul cu calcul procente
 - select a6 – scriem „%”
 - select b6 / = b5 / \$ f5 (explicația în Excel pentru utilizarea \$)
 - copy din b6 în c6:f6
- **Interpretare:** tabelele sunt acum completate cu totaluri și procente.

Subtema 2b: calculul unor variabile intermediare sau finale

Vom calcula indicele de masă corporală BMI (Body Mass Index) conform relației BMI = Greutate (kg) / înălțime² (m)

- **Procedeu**
- introducem 2 noi coloane:
 - F va conține înălțimea în m (în F2 notăm h(m))
 - G va conține BMI (În G2 notăm BMI)
- selectăm căsuța F3
- în bara de inserare comenzi introducem: „= d3/100”;
- copiem formula pe toată coloana F (de la f3:f62) (revedem comenzile Ctrl + C pentru Copy și Ctrl + V pentru Paste)
- selectăm G3
- introducem comanda: „= e3 / (f3 * f3)”
- copiem formula de la g4:g62

- dorim afișarea BMI cu două zecimale: selectăm coloana G (cu mouse), apoi din bară, Format / Conditional Formatting - apare o fereastră cu același nume, având 4 rubrici în secțiunea Condition 1:
 - prima – alegem / lăsăm *Cell Value Is*,
 - a doua – putem selecta condiții: Greater than, Less than, Between etc.; alegem întâi *greater than*;
 - completăm limita dorită, de ex. 22.1, peste care interpretăm obezitatea
 - dăm click pe butonul *Format*; se deschide o nouă fereastră – *Format Cells* în care putem allege o variantă de afișare dacă este îndeplinită condiția: font, culoarea textului, culoarea fondului etc.; de ex. alegem în Pattern o culoare roșie pentru fond, apoi OK, se revine în fereastra Conditional Formatting
 - dăm acum clic pe butonul *Add*; fereastra își adaugă încă o secțiune – Condition 2, în care repetăm pașii anteriori; se pot aplica maximum 3 condiții.
- să mai adăugăm condițiile:
 - fond galben pentru valori între 19.4 și 22.1 (supraponderali)
 - fond verde pentru valori între 14.2 și 19.4 (normali)
 - vor rămâne neschimbate celulele cu valori sub 14.2 (subponderali).
- valoarea din celula noastră va îndeplini una din condiții; când condiția este îndeplinită se va vedea culoarea în celulă; după introducerea tuturor condițiilor, selectăm din nou celula g3 și cu Copy (sau Ctrl+C) vom copia condițiile pe toată coloana G în jos, cu *Edit / Paste Special / Formats* (altfel copiază și valorile). Vom vedea culorile alese condiționat, în funcție de valori, în celulele corespunzătoare.
- **Interpretare:** apare coloana întreagă BMI cu 2 zecimale, colorată astfel încât ies în evidență cele 4 clase în funcție de valoarea BMI.

Tema 3: Parametrii statistici

Obiectiv: determinarea parametrilor statistici ai unui eșantion

Worksheet utilizat: par-st

Modul de lucru:

Tema are două părți asemănătoare:

Tema 3a: parametrii statistici pentru 1 variabilă, pe un eșantion întreg pentru înălțimea grupului de copii.

- **Procedeu**
- din bara de comenzi selectăm „Tools”; apare un pop-up meniu,
- din meniu selectăm „Data Analysis” (dacă nu există această comandă, ea putea fi introdusă cu *Add-Ins*),
- apare un nou meniu – selectăm „Descriptive Statistics”
- în noua fereastră deschisă selectăm:
 - *input range*: d2:d62
 - *grouped by*: Columns
 - bifăm „*labels in first row*” (este vorba de d2)
 - *output range* – alegem o căsuță în dreapta, de ex g2
 - bifăm „*summary statistics*”
 - (opțional putem bifa și *confidence level for mean*) / ok.
- **Interpretare:** apare tabelul cu rezultate care este comentat în curs, cap. 2.3.1 (tabelul 2.2)

Tema 3b: parametrii statistici pentru 1 variabilă, pe un subgrup.

Vom obține parametrii statistici pentru înălțimea fetelor din grupul de copii.

- **Procedeu**
- selectăm coloana C (sex)
- din bara de comenzi selectăm „Data”; din submenu selectăm “2 Filter”, apoi „Auto filter”
- apare un buton cu care putem selecta F (numai fetele)
- din coloana D selectăm înălțimile celor 24 fete, pe care le copiem într-un nou Worksheet (din Insert / Worksheet), să zicem de la a1 în jos (click pe a1, apoi CTRL + V)
- mai departe procedăm ca mai sus: *Tools / Data Analysis „Descriptive Statistics”* etc. (atenție: datele sunt acum în a1:a24, fără *labels in first row*, etc);
- putem da un nume acestui nou worksheet: în bara de jos, cu mouse-ul pe Sheet1 care este acum activ, clic dreapta – se deschide un meniu din care selectăm *Rename*, apoi introducem noul nume, de ex. "**par-st2**".
- **Interpretare:** tabelul obținut conține acum rezultatele statistice privind înălțimea fetelor din lot.

Tema 4: Histograme din fișiere de date brute

Obiectiv: crearea limitelor intervalelor pentru histograme și ridicarea histogramelor. Limitele claselor, ce poartă denumirea de "bin", vor fi trecute într-o coloană ce o vom crea alături de tabelul de date. Este bine ca valorile pentru bin-uri să se aleagă încât să se intercaleze între valorile experimentale. Limitele se pot alege arbitrar – noi vom lucra în două variante, cu clase late de 1 cm, respectiv cu lățimea optimă (rotunjită), dată de relația (2.4), p.27, adică $h = (X_M - X_m) / (1 + 3.322 \lg N)$.

Worksheet utilizat: par-st

Modul de lucru

Tema are 2 părți similare:

Subtema 4a: crearea limitelor pentru histograme cu lățimea clasei de 1 cm.

- **Procedeu**
- avem pe coloanele G și H parametrii statistici calculați în Tema 3a.
- pe coloana L vom crea limitele pentru prima histogramă
- selectăm L2, introducem valoarea cu lățimea clasei („1” în cazul nostru)
- selectăm L3, introducem o valoare pentru limita din dreapta a primei clase; dacă dorim să avem clasele centrate pe valori întregi sau rotunde (în cm), vom lua pentru limite jumătăți de unitate – pentru noi, având Minimum 128.2 (din tabelul obținut în subtema 3a) vom lua ca primă valoare 128.5;
- dorim să incrementăm aceste limite bin cu lățimea clasei la fiecare pas, deci completăm coloana J în jos astfel:
 - $L4 = L3 + L\$2$ (am fixat pasul de incrementare)
 - copiem L4 în jos încât să acoperim valoarea maximă (149.5 în cazul nostru); putem să și depășim valoarea, însă ulterior vom folosi doar intervalul ce acoperă datele reale;
- din *Tools / Data Analysis / Histogram*
- *input range:* d3:d62
- *bin range:* j3:j25
- *chart output*
- apare tabelul de valori în clase și histograma; graficul poate fi "mărit", prin "drag" cu mouse-ul;
- pregătim în final histograma cu date suplimentare (titlu, axe etc).
- **Interpretare:** tabelul afișat în stânga histogramei conține numărul de copii din fiecare clasă de înălțime cu lățimea de 1 cm. Prima clasă are marginea superioară valoarea din coloana bin, adică 128.5, deci cuprinde toate valorile până la 128.5 inclusiv. Observăm tendința centrală, însă histograma are "numeroși dinți", datorită claselor prea înguste. Vom putea compara acest grafic cu următorul, pe care îl realizăm cu lățimea adecvată.

Subtema 4b: crearea histogramelor cu lățime optimă

Se procedează la fel ca în Excel 2007.

Tema 5: Teste statistice

Obiectiv: compararea a două serii numerice sau calitative

Worksheet utilizat: teste și chi-sq

Modul de lucru

Tema are 3 părți distincte, asemănătoare, una pentru testul t pereche și una pentru testul t nepereche, ultima pentru testul chi-sq.

Subtema 5a: testul t pereche

Vom aplica testul t pereche pentru a compara diferențele între tensiunile sistolice înainte și după tratament.

- **Procedeu**
- din *Tools / Data Analysis / t – Test: Paired Two Samples for Means*
- în fereastra deschisă alegem:
 - *variable 1 range:* b2:b62
 - *variable 2 range:* c2:c62
 - *hypothesized mean difference:* 0 (sau lăsăm necompletat)
 - bifăm *labels* (avem în b2 și c2)
 - lăsăm Alpha 0.05
 - selectăm *output* (de ex g2)
- apare tabelul de forma celui din cap. 4.2.1.3, tabelul 4.2.b
- **Interpretare:** în tabelul afișat apar multe date de interes: mediile celor două serii, varianțele (varianța este pătratul deviației standard), cea mai de interes valoare fiind din căsuța "P(T<=t) two-tail", care, fiind 0.0356 (sub 0.05), deci indică o diferență semnificativă.

Subtema 5b: testul t nepereche

Vom aplica testul t nepereche pentru a compara tensiunea sistolică înainte de tratament la bărbați, comparativ cu femeii.

- **Procedeu**
- deschidem worksheet **teste**
- selectăm coloana A (sex)
- cu *Data / Filter / Autofilter*, preluăm pe rând:
 - datele pentru bărbați, selectând M și copiind din coloana S7S-1 datele într-un nou worksheet în coloana A de la A2 în jos; în A1 scriem "M"
 - datele pentru femei, similar în coloana B de la B2 în jos; în B1 scriem "F".
- alegem *Tools / Data Analysis / t – test: Two – Sample Assuming Unequal Variances*
- obținem un tabel similar cu tabelul 4.2.a din cap. 4.2.1.2
- putem în final denumi noul worksheet "teste2".
- **Interpretare:** în tabelul afișat valoarea din căsuța "P(T<=t) two-tail", este acum 0.91, mult peste 0.05, deci diferențele sunt nesemnificative.

Subtema 5c:

- **Procedeu**
- Se deschide worksheet **chi-sq**
- Presupunem o distribuție uniformă a persoanelor pe grupe sanguine, atunci valorile așteptate (căsuțele b4:e4) se pot calcula cu relația = total – general / 4 = \$ / F / \$ / 3 / 4.
- Pentru calcul:
 - selectăm căsuța de rezultat (i3)
 - selectăm funcția f_x și alegem CHITEST; se deschide fereastra „Function Arguments”
- *actual_range* (se referă la valori experimentale, adică b3:e3 // va fi b8:e8 pentru **tabelul exp. 2**)
- *expected_range* (se referă la valori teoretice; atunci luăm b4 : e4 sau b9:c9)
- **Interpretare:** apare valoarea lui p (0,000828) < 0,001 deci vom avea diferențe extrem de semnificative în valorile experimentale și cele teoretice; vom respinge H_0 .
- Repetăm toate operațiunile pentru un lot mult mai mic (**tabelul exp. 2**), dar păstrând proporțiile – observăm că, în cazul loturilor mici, diferențele nu mai sunt semnificative (p = 0.644), deși proporțiile claselor sunt aceleași.

Tema 6: Analiza corelației

Obiectiv: determinarea coeficientului de corelație și trasarea dreptei de regresie

Worksheet utilizat: regr

Modul de lucru

Tema are 2 părți complementare

Subtema 6a: comanda *correlation*

- **Procedeu**
- completăm coloana C cu: indicele BMI (v. Tema 2b)
- din *Tools / Data Analysis / Correlation* se completează în fereastra deschisă:
 - *input range* a2:c42 (Obs: variabilele de corelat trebuie să fie plasate în 2 coloane vecine în tabel); coloana A va fi pe Ox iar B pe Oy
 - *grouped by columns*
 - bifăm *labels* (sunt a2, b2 și c2)
 - selectăm ieșirea (de ex e2)
- **Interpretare:** apare un tabel 3 x 3 cu coeficienții de corelație separat pentru corelația înălțime-greutate (0,75 în cazul nostru, deci avem o corelație directă, destul de puternică), corelația înălțime - bmi (-0,69 = inversă, medie) și corelația greutate – bmi (-0,04 = foarte slabă)

Subtema 6b: comanda *regression*

- **Procedeu**
- din *Tools / Data Analysis / Regression* completăm:
 - *input Y range:* b2:b42 (greutatea)
 - *input X range:* a2:a42 (înălțimea)
 - bifăm *labels* (sunt a2 și b2)
 - nu bifăm *constant is zero* (dacă se bifează atunci se calculează dreapta de regresie care trece prin origine)
 - nu bifăm: *Standardized residuals, Residual Plots, nici Normal Probability Plots*
 - bifăm *Line Fit Plot* – pentru graficul dreptei de regresie
- alegem *output range* (de ex e7)
- **Interpretare:** apar rezultatele (*Summary Output*) în 4 tabele și un grafic
- din primul tabel (*Regression Statistics*) urmărim Multiple R – coeficientul de corelație (0,75 în cazul nostru)
- din al doilea tabel (*ANOVA*) urmărim Significance F (2.32×10^{-8} ceea ce înseamnă că avem o probabilitate extrem de mică să obținem din întâmplare așa un coeficient de corelație (este echivalentul lui p))
- din al treilea tabel găsim:
 - ordonata la origine = Intercept (-2.89 ± 4.99)
 - panta (slope) – în căsuța sub valoare intercept ($0,255 \pm 0.036$)
 - pentru ambele sunt date limitele 95%

- al patrulea tabel arată valorile calculate pentru greutatea calculată (predicted) corespunzătoare înălțimilor, conform:
 $greut. calc = intercept + panta * înălțime, sau$
 $greut. calc = -2,89 + 0,255 * înălțimea$
- următoarea coloană conține diferențele:
 $residual = greut (reală) - greut. calc$
- In grafic facem următoarele operații:
 - mărim graficul („tragem” cu mouse-ul de colțuri)
 - schimbăm axa Oy: facem (dublu) clic pe axă; se deschide fereastra „Format Axis”; selectăm Scale și pentru Minimum introducem altă valoare (de ex 27)
 - trasăm dreapta de regresie: cu mouse, click dreapta pe un punct experimental (bleumarin) – se deschide un menu mic și selectăm „Add Trendline” și selectăm trend Linear (prima căsuță)
 - putem elimina valorile prezise cu un click dreapta pe un punct teoretic (magenta), iar din fereastra mică selectăm „clear”.

Tema 7: Analiza riscului

Obiectiv: realizarea tabelului de contingență pornind de la date brute și calculul indicatorilor de risc.

Worksheet utilizat: risc

Modul de lucru

Tema are 2 etape distincte

Subtema 7a: realizarea tabelului de contingență

Varianta 1

- **(a) Construim scheletul tabelului de contingență**, fie sub date (sub linia 162) sau pe un nou worksheet (alegem această variantă), numit **risc2**
- *Insert / Worksheet* // click dreapta pe nume / *rename / risc2*
- selectăm regiunea a1 : d4; din bara de formatare alegem borders cu linii pentru toate căsuțele; introducem notațiile clasice în A2 : A4 (E+, E-, total), respectiv în B1 : D1 (B+, B-, total).
- Completăm D4 cu valoarea numărului total (160 în cazul nostru)

- **(b) Transformarea variabilei de expunere**
- revenim în worksheet **risc**
- Vom transforma variabila numerică TA – Sys în variabila calitativă Hipertens cu convenția:
 - dacă $TA - Sys \geq 140$ atunci Hipertens = 1 (= DA)
 - dacă $TA - Sys < 140$ atunci Hipertens = 0 (= NU)
- În C2 notăm *label* Hipertens
- Selectăm C3 și dăm click pe f_x ; se deschide fereastra „*Insert function*”, în care luăm:
 - *select category:* All
 - baleiem toate funcțiile până la IF, selectăm IF, se deschide fereastra „*Function Arguments*”, în care luăm:
 - = *logical test:* $a3 > = 140$
 - = *value if true:* 1
 - = *value if false:* 0 / ok
 - în căsuța c3 apare 0 sau 1 (în cazul nostru 0)
- copiem formula din c3 în jos (c4:c162) (revedeți funcțiile Ctrl + C și Ctrl + V)

- **(c) Aflăm totalurile E+ și B+:**
- selectăm b163, apoi pe Σ și *Sum* (apare propunerea SUM (B3:B162), o acceptăm (enter); apare totalul pe coloana B (fumat); copiem valoarea în sheet **risc2**, D2 (76 în cazul nostru) (Obs: se exersează Copy din Sheet **risc** cu Paste în Sheet **risc2** cu: Select D4, apoi =, trecere la Sheet risc și click pe B163)
- similar cu c163 pe care o trecem în B4 (36 în cazul nostru)

- **(d)** Selectăm coloanele B și C (ar fi suficient B) și filtrăm numai fumătorii (*Data / Filter / Autofilter*), apoi click pe butonul din coloana B și din micul menu ce apare selectăm 1; dispar liniile ce aveau 0 în coloana B
- selectăm o căsuță pe coloana C jos, sub ultima linie scrisă (va fi C164 în cazul nostru); introducem funcția SUBTOTAL cu proprietățile:
 - *function_num*: luăm 9 (se va face totalul ignorând valorile ascunse),
 - *Ref 1*: luăm c3:c162 /OK (precizăm toată coloana, știind că valorile ascunse vor fi ignorate);
- apare totalul pentru care aveam 1 și în coloana B și în coloana C, adică sunt și E+ și B+; introducem această valoare (clic pe c166 și Ctrl+C) (32 în cazul nostru) în tabelul din risc2, căsuța b2; Obs: preferăm să folosim Paste Special / Value (prin clic dreapta), altfel se dorește copierea formulei de subtotal!
- Restul tabelului se completează prin diferențe:
- $c2 = d2 - b2$; apoi copy c2 în c4,
- $b3 = b4 - b2$; apoi copy b3 în c3 și d3.

Varianta 2 – se modifică doar punctul (d).

- **Procedura**
- **(d)** Introducem coloana D, numită „e+ / b+” (scriem în d2)
- Selectăm d3, apoi **f_x** și alegem funcția AND, (dacă nu se vede se selectează întâi la category "All") în care luăm:
 - *Logical 1*: $b3 = 1$
 - *Logical 2*: $c3 = 1 / ok$
- apare TRUE sau FALSE (FALSE în cazul nostru)
- copiem formula AND din d3 (adică $b3 = 1$, $c3 = 1$) în căsuțele d4:d:162; vor fi TRUE numai căsuțele pacienților care au 1 și 1 (fumează și au hipertensiune)
- selectăm d163 și introducem funcția COUNTIF; în fereastra ei introducem:
 - *Range*: d3:d:162
 - *Criteria*: TRUE / ok
- apare valoarea corespunzătoare (32 în cazul nostru)

Varianta 3

- **Procedura**

O metodă care poate fi folosită în toate situațiile în care dorim construirea unui tabel de contingență 2 x 2 pornind de la date dihotomice trecute în două coloane constă în transformarea reprezentării în cele două coloane astfel:

- o coloană să conțină numai valori 0 și 1 pentru NU (-) și DA (+)
- cealaltă coloană să conțină numai valori 0 și 2 pentru NU (-) și DA (+).

Astfel, prin suma pe linii obținem segregarea celor 4 variante:

NU & NU devine $0 + 0 = 0$
 DA & NU devine $1 + 0 = 1$
 NU & DA devine $0 + 2 = 2$
 DA & DA devine $1 + 2 = 3$.

Deci, dacă din coloana de sumă (o vom numi Sumă Ponderată) numărăm câți de 0, 1, 2 sau 3 avem, obținem exact cele 4 căsuțe din tabelul de contingență.

De fapt nici nu este nevoie să creăm separat coloana a doua cu 0 și 2, ci o lășăm tot cu 0 și 1 iar în sumă să luăm valoarea din coloana a doua înmulțită cu 2, adică vom calcula:

$$\text{suma} = \text{col.1} + 2 * \text{col.2}$$

Vom proceda astfel:

- prima coloană 0/1 este B și corespunde E+/E-
- a doua coloană 0/1 este C, creată în punctul (b) din Varianta 1 și corespunde B+/B-
- creăm coloana E cu suma ponderată (SP) a valorilor din B și C
- astfel pentru E3 luăm $= B3 + 2 * C3$
- copiem E3 în jos, de la E4:E162

În **risc2** putem reconstrui tabelul de contingență în aria a6:d9, cu același notații.

Interpretare: Am obținut coloana E cu următoarele proprietăți:

- dacă are valoarea 0, pacientul este e- și b- (căsuța c3 sau c8 din risc 2)
- dacă are valoarea 1, pacientul este e+ și b- (c2 / c7)
- dacă are valoarea 2, pacientul este e- și b+ (b3 / b8)
- dacă are valoarea 3, pacientul este e+ și b+ (b2 / b7).
- folosim funcția COUNTIF de 4 ori, de fiecare dată având Range risc!\$E\$3:\$E\$162, direct în risc 2, precizând că:
 - dacă punem criteria = 0 obținem valoarea din c3 (sau c8) etc.
- completăm tabelul cu totaluri.

Subtema 7b: calculul indicatorilor

- Folosim pachetul statistic **Epi Info**
- Din meniul principal selectăm *Utilities*,
- Selectăm *Statcalc*,
- Completăm tabelul (în cazul nostru avem 32 și 44 respectiv 4 și 80)
- **Interpretare:** Apare tabelul de rezultate, cu interpretare conform cap. 6.4.2, figura 6.3.

Tema 8: Validarea testelor diagnostice

Obiectiv: crearea tabelului de contingență și calcularea indicatorilor.

Worksheet utilizat: val-dg

Modul de lucru

Tema are 2 părți succesive: crearea tabelului, respectiv calculul indicatorilor.

Subtema 8a: crearea tabelului de contingență.

Folosim notațiile din curs cap. 8.1, tabel 8.1

În worksheet am notat cu + sau – testul pozitiv sau negativ, respectiv prezența sau absența bolii.

Construim scheletul tabelului de contingență (similar temei 7a) cu următoarele modificări:

- spațiul H1:K4 (nu creem acum un alt worksheet, ci alcătuim tabelul în dreapta datelor),
- notăm h2 și h3 cu B+ și B-, iar i1 și j1 cu T+ și T-; adăugăm "total" în h4 și k1.

Folosim o procedură similară cu cea din varianta 3 a temei 7a:

- Vom crea coloana D, cu valori 0 și 1 corespunzătoare lui T- și T+
- Coloana E, cu valori 0 și 1 corespunzătoare lui B- și B+
- Coloana F pentru suma ponderată
 - În D3 introducem prin f_x funcția IF cu:
= *logical test*: b3 = „+”
= *Value if true*: 1, *value if false*: 0
 - Copiem D3 în E3
 - clic pe f3, apoi "=d3+2*e3"
 - selectăm d3:f3, apoi Ctrl+C
 - Copiem totul în D4:F122 (pe trei coloane deodată).
- În coloana F avem valori între 0 și 3, cu semnificația:
 - b+ / t+ valoarea 3 (va fi în căsuța i2)
 - b+ / t- valoarea 2 (j2)
 - b- / t+ valoarea 1 (i3)
 - b- / t- valoarea 0 (j3).

Cu COUNTIF vom completa căsuțele din tabel:

i2 = Countif (\$f\$3:\$f\$122,3)

j2 = Countif (\$f\$3:\$f\$122,2)

i3 = Countif (\$f\$3:\$f\$122,1)

j3 = Countif (\$f\$3:\$f\$122,0).

Completăm tabelul cu totaluri.

Subtema 8b: calculul indicatorilor.

- numele lor simbolic – în h8 : h16
- valorile în i8:i16
- vom folosi relațiile:
 - i8 FP (fals pozitivi) = $i3 / k3$
 - i9 FN (fals negativi) = $j2 / k2$
 - i10 SN (sensibilitate) = $[i2=TP \text{ (real / „true” pozitivi)}] = i2 / k2$
 - i11 SP (specificitate) = $[j3=TN \text{ (real / „true” negativi)}] = j3 / k3$
 - i12 VPP (valoarea predictivă pozitivă) = $i2 / i4$
 - i13 VPN (valoarea predictivă negativă) = $j3 / j4$
 - i14 AC (acuratețea) = $(i2 + j3) / k4$
 - i15 RE (rata erorii de clasificare) = $(i3 + j2) / k4$
 - i16 Y (indice Youden) = $i10 + i11 - 1$.

Interpretarea rezultatelor se face conform celor prezentate în cap. 8.2 și cap. 8.3.